



LUIZ ANTONIO YANES BERNARDO JÚNIOR

**NEW INSIGHTS ON GENOTYPES BY ENVIRONMENTS
INTERACTION USING GEOGRAPHICAL COORDINATES**

LAVRAS – MG

2019

LUIZ ANTONIO YANES BERNARDO JÚNIOR

**NEW INSIGHTS ON GENOTYPES BY ENVIRONMENTS INTERACTION USING
GEOGRAPHICAL COORDINATES**

Tese apresentada à Universidade Federal de Lavras,
como parte das exigências do Programa de Pós-
Graduação em Genética e Melhoramento de Plantas,
área de concentração em Genética e Melhoramento de
Plantas, para a obtenção do título de Doutor.

Orientador

Dr. Renzo Garcia Von Pinho

Coorientador

Dr. Márcio Balestre

LAVRAS – MG

2019

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Bernardo Júnior, Luiz Antonio Yanes.

New insights on genotypes by environments interaction using
geographical coordinates / Luiz Antonio Yanes Bernardo Júnior. -
2019.

108 p. : il.

Orientador(a): Renzo Garcia Von Pinho.

Coorientador(a): Márcio Balestre.

Tese (doutorado) - Universidade Federal de Lavras, 2019.

Bibliografia.

1. Interação genótipos . I. Von Pinho, Renzo Garcia. II.
Balestre, Márcio. III. Título.

LUIZ ANTONIO YANES BERNARDO JÚNIOR

**NEW INSIGHTS ON GENOTYPES BY ENVIRONMENTS INTERACTION USING
GEOGRAPHICAL COORDINATES**

**NOVAS IDEIAS EM INTERAÇÃO GENÓTIPOS POR AMBIENTES USANDO
COORDENADAS GEOGRÁFICAS**

Tese apresentada à Universidade Federal de Lavras,
como parte das exigências do Programa de Pós-
Graduação em Genética e Melhoramento de Plantas,
área de concentração em Genética e Melhoramento de
Plantas, para a obtenção do título de Doutor.

APROVADA em 27 de setembro de 2019.

Dr. José Airton Rodrigues Nunes	UFLA
Dr. José Maria Villela Pádua	SOUZA CRUZ
Dr. Narjara Fonseca Cantelmo	BAYER

Orientador

Dr. Renzo Garcia Von Pinho

Coorientador

Dr. Márcio Balestre

LAVRAS – MG

2019

*À Deus, por me capacitar, inspirar e guiar
Dedico*

AGRADECIMENTOS

Primeiramente a Deus, por possibilitar a realização do sonho de obter o doutorado. À minha namorada, Aline, por estar sempre ao meu lado em todos os momentos e ser a base de meu sucesso, motivando-me e apoiando-me em qualquer situação. Aos meus pais, tios, primos e avós, por sempre acreditarem em mim e colocarem meu nome em suas orações. Aos meus irmãos de república, por tornarem engraçados, até mesmo, os momentos mais difíceis.

Agradeço aos professores, Renzo Garcia Von Pinho e Márcio Balestre, pela orientação, amizade e dedicação, durante todos os anos em que trabalhamos juntos. Ao professor Daniel Furtado Ferreira e sua esposa Lidiane por toda atenção e dedicação em me ajudar quando mais precisei, tanto na elaboração da tese quanto nos momentos de descontração. Aos amigos, Carlos Pereira e Luciano Oliveira, por todo suporte dado durante a condução dos trabalhos desta tese.

Agradeço à Universidade Federal de Lavras, Programa de Pós-Graduação em Genética e Melhoramento de Plantas, Núcleo de Estudos em Genética e Melhoramento de Plantas, Departamento de Biologia, Departamento de Agricultura, Departamento de Estatística e todos os professores e funcionários que enriqueceram meus anos de estudo e foram fundamentais no meu crescimento profissional e pessoal.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, pela concessão de bolsa de estudos.

RESUMO GERAL

Ensaio multambientais estão entre os experimentos mais comumente conduzidos na área de ciências agrárias. Devido à aplicabilidade do método AMMI (Additive Main Effects and Multiplicative Interaction) e a forma como tem sido estudado e aplicado, o presente trabalho explorou novas alternativas de estudo da interação genótipos x ambientes (GEI) por meio deste modelo. O primeiro trabalho foi conduzido com o objetivo de propor um modelo de predição das capacidades geral e específica de combinação, e suas interações com ambientes, associado ao uso de regiões de credibilidade em biplots obtidas por meio do modelo AMMI-Bayesiano. Em geral, para a análise de dados simulados, as predições obtidas tiveram alta correlação com os valores reais. Para os efeitos das capacidades geral e específica de combinação (GCA e SCA, respectivamente), as predições mantiveram o padrão de sinais e ranqueamento. Além disso, o modelo foi eficiente em fornecer intervalos de credibilidade que cobriam os valores simulados. Para a análise de dados reais, as estimativas de GCA e SCA para todos os genótipos avaliados não diferiram de zero. Os biplots para as interações GCA x Ambientes e SCA x Ambientes permitiram determinar quais genótipos apresentam efeitos de GCA e SCA estáveis de forma mais precisa. As elipses nos biplots demonstraram a incerteza em torno das estimativas da interação. O modelo mostra-se como ferramenta promissora para auxiliar na tomada de decisão do melhorista na seleção e recomendação de genótipos. O segundo trabalho foi conduzido com o objetivo de estudar GEI por meio do uso de variáveis ambientais no modelo AMMI sob abordagem funcional com diferentes volumes de dados e níveis de desbalanceamento. Para isso, visou-se avaliar o comportamento da análise de componentes principais de dados funcionais (FPCA) em cenários de desbalanceamento de dados e integrá-la ao método EM-AMMI (Expectation-Maximization Additive Main Effects and Multiplicative Interaction) para realizar a predição genotípica sob uma perspectiva funcional. Por meio da análise de dados funcionais, é possível captar informações que vão além de dados discretos que representam genótipos e ambientes, o que permite captar o padrão de interação e o comportamento dos genótipos ao longo dos ambientes. A integração do método smooth SVD (Singular Value Decomposition) ou SSVD ao método EM-AMMI para realizar a imputação de dados e lidar com o desbalanceamento pode alterar a perspectiva de estudo de GEI. A identificação de padrões de efeitos da interação por meio do método EM-AMMI Funcional representa uma nova forma de entender a interação, e abre diversas possibilidades para aplicação no melhoramento de plantas.

Palavras-chave: Interação genótipos x ambientes; estabilidade; adaptabilidade.

GENERAL ABSTRACT

Multi-environmental trials are among the most commonly conducted experiments in the field of agricultural science. Due to the applicability of the AMMI method (Additive Main Effects and Multiplicative Interaction) and the way it has been studied and applied, this work explored new alternatives for the study of genotype x environment interaction (GEI) through this model. The first work was conducted with the objective of proposing a model of prediction of general and specific combining abilities, and their interactions with environments, associated with the use of credible regions in biplots obtained by the AMMI-Bayesian model. In general, for the analysis of simulated data, the predictions obtained had a high correlation with the real values. For the general and specific combining capabilities effects (GCA and SCA, respectively), the predictions maintained the signal pattern and ranking. In addition, the model was efficient in providing credible intervals that covered the simulated values. For real data analysis, GCA and SCA estimates for all evaluated genotypes did not differ from zero. The biplots for the GCA x Environments and SCA x Environments interactions allowed us to determine which genotypes have stable GCA and SCA effects in a more accurate way. Ellipses in the biplots demonstrated the uncertainty around the interaction estimates. The model is a promising tool to help the breeder's decision making in selecting and recommending genotypes. The second work was conducted with the objective of studying GEI through the use of environmental variables in the AMMI model under functional approach with different data volumes and unbalance levels. To this end, it was aimed to evaluate the behavior of the principal component analysis of functional data (FPCA) in data imbalance scenarios and integrate it with the EM-AMMI (Expectation-Maximization Additive Main Effects and Multiplicative Interaction) method to perform the genotypic prediction from a functional perspective. Through functional data analysis, it is possible to capture information that goes beyond discrete data representing genotypes and environments, which allows to capture the pattern of interaction and behavior of genotypes across environments. Integrating the smooth SVD (Singular Value Decomposition) or SSVD method with the EM-AMMI method to perform data imputation and deal with imbalance can change the perspective of GEI study. The identification of interaction effect patterns through the EM-AMMI Functional method represents a new way of understanding interaction, and opens up several possibilities for application in plant breeding.

Keywords: Genotype x environment interaction; stability; adaptability.

SUMÁRIO

PRIMEIRA PARTE.....	10
1. INTRODUÇÃO GERAL.....	10
REFERÊNCIAS	13
SEGUNDA PARTE.....	14
CAPÍTULO 1 - AMMI-BAYESIAN MODEL AND USE OF CREDIBLE REGIONS IN THE STUDY OF COMBINING ABILITY IN MAIZE.....	14
ABSTRACT	15
INTRODUCTION.....	16
MATERIAL AND METHODS	19
<i>Bayesian Analysis of experiment groups using diallel model.....</i>	<i>19</i>
<i>Priori distributions for parameters of model and function of likelihood.....</i>	<i>21</i>
<i>Gibbs Algorithm</i>	<i>22</i>
<i>Convergence and credible intervals in biplot</i>	<i>25</i>
<i>EM-REML.....</i>	<i>26</i>
RESULTS.....	27
<i>Analysis of Simulated Data.....</i>	<i>27</i>
<i>Analysis of real data.....</i>	<i>37</i>
DISCUSSION	44
ACKNOWLEDGEMENT	47
REFERENCES.....	47
SUPPLEMENTAL MATERIAL	51
CAPÍTULO 2 – USO DE VARIÁVEIS AMBIENTAIS NO ESTUDO DA INTERAÇÃO GENÓTIPOS x AMBIENTES	61
RESUMO	62
INTRODUÇÃO	63
MATERIAL E MÉTODOS.....	66
<i>Simulação dos dados</i>	<i>66</i>
<i>Análise de dados funcionais (FDA).....</i>	<i>67</i>
Predição dos dados faltantes	72
<i>Capacidade preditiva de ajuste de modelo.....</i>	<i>72</i>
Análise com dados completos	72
<i>Capacidade preditiva com dados faltantes.....</i>	<i>73</i>

Desbalanceamento de 1 ambiente	73
Desbalanceamento Sistemático	74
Desbalanceamento Aleatório	74
<i>Capacidade preditiva de GEI</i>	75
EM-AMMI Funcional	75
RESULTADOS	76
<i>Capacidade preditiva de ajuste de modelo</i>	76
Análise com dados completos	76
<i>Capacidade preditiva com dados faltantes</i>	84
Desbalanceamento de 1 ambiente	84
Desbalanceamento Sistemático	85
Desbalanceamento Aleatório.....	87
<i>Capacidade preditiva de GEI</i>	88
EM-AMMI Funcional	88
DISCUSSÃO	98
AGRADECIMENTOS	104
REFERÊNCIAS	104
MATERIAL SUPLEMENTAR	107

PRIMEIRA PARTE

1. INTRODUÇÃO GERAL

No melhoramento de plantas, os ganhos obtidos ao longo dos ciclos de seleção e recombinação vão se tornando cada vez menores. Isso exige o emprego de técnicas cada vez mais sofisticadas, as quais permitem captar diferenças cada vez menores entre os genótipos. O aumento do volume de informações fenotípicas, genotípicas e ambientais disponíveis estimulam o desenvolvimento de novas técnicas estatísticas que possibilitem uma descrição e predição mais precisa da interação genótipos x ambientes (GEI) e uma maior eficiência nos programas de melhoramento (van Eeuwijk et al., 2016). Em um país como o Brasil, que possui grande extensão territorial e, conseqüentemente, uma grande diversidade climática, o uso de tais métodos para a avaliação da adaptabilidade e estabilidade de genótipos se torna fundamental para a obtenção de sucesso em um programa de melhoramento.

Ensaio multiambientais estão entre os experimentos mais comumente conduzidos na área de ciências agrárias. Análises estatísticas efetivas dos dados obtidos nesses ensaios podem auxiliar agrônomos, melhoristas e outros pesquisadores a terem um maior progresso. Entre essas técnicas, o método AMMI (Additive Main Effects and Multiplicative Interaction) tem sido um dos mais empregados nos últimos anos (Gauch, 2006; Gauch et al., 2008; van Eeuwijk et al., 2016; Bernardo Júnior et al., 2018). Esse método utiliza a análise de variância para particionar a variação em efeitos de genótipos, ambientes e GEI. Em seguida, realiza-se a análise de componentes principais para avaliar a interação (Zobel et al., 1988; Gauch, 2013).

Devido à aplicabilidade do método AMMI e a forma como tem sido estudado e aplicado, o presente trabalho explorou novas alternativas de estudo da interação genótipos por ambientes por meio desse modelo. Os artigos constantes nesta tese apresentam novas abordagens que podem solucionar problemas comumente enfrentados por melhoristas. O capítulo 1 trata da análise AMMI integrada à análise dialélica sob uma abordagem bayesiana. Já o capítulo 2 traz proposta de uso do método AMMI associado ao algoritmo Expectation-Maximization (EM) (Gauch e Zobel, 1990) sob o enfoque da análise de dados funcionais.

Atualmente, um dos desafios em programas de melhoramento é encontrar genótipos que sejam adaptados e ao mesmo tempo estáveis nas diversas regiões de cultivo. No caso do milho, a diversidade de ambientes é especialmente elevada devido à segunda safra. Porém, apesar de extensivamente utilizada pelos melhoristas na seleção de linhagens, a capacidade de combinação não garante que os híbridos mais produtivos sejam os mais estáveis. A avaliação

de um grande número de combinações híbridas em ensaios multiambientais pode facilitar a identificação de híbridos de alto desempenho e expressão da heterose (Abera et al., 2016).

O estudo da interação genótipos por ambientes por meio de ensaios multiambientais pode apresentar algumas dificuldades como, por exemplo, prever a performance híbrida em ambientes os quais não foi avaliado; heterogeneidade de variâncias; diferente número de repetições entre experimentos, o que leva a uma não ortogonalidade da análise; e populações segregantes de diferentes tamanhos, que leva a contribuições desbalanceadas das linhagens iniciais para os híbridos (Rojas & Sprague, 1952; Kadam et al., 2016; Acosta-Pech et al., 2017; Giraud et al., 2017). Essas dificuldades podem ser solucionadas por meio de análises sob abordagem bayesiana. Nesse contexto, o estudo apresentado no capítulo 1 deste trabalho foi conduzido com o objetivo de propor um modelo de predição das capacidades geral e específica de combinação, e suas interações com ambientes, associado ao uso de regiões de credibilidade em biplots obtidas por meio do modelo AMMI-Bayesiano.

Uma das variações do método AMMI é o EM-AMMI, proposto por Gauch e Zobel (1990). Esse método implementa o algoritmo EM para realizar a imputação de dados e permitir o ajustamento do modelo AMMI em cenários de desbalanceamento de dados. Esse método permite uma exata e completa decomposição de toda informação retirada de um ensaio em informações diretas e informações indiretas. Para um determinado tratamento, as informações diretas representam suas repetições, enquanto as informações indiretas representam todos os outros dados do experimento.

A análise de dados funcionais (FDA) é uma metodologia a qual considera as funções dos dados observados como entidades únicas, e não como uma simples sequência de observações. O termo funcional refere-se à estrutura intrínseca dos dados. Na prática, dados funcionais são observados discretamente como n pares (t_j, y_j) , em que y_j representa um instante da função no tempo t_j . Essa função t_j comumente representa tempo, mas outras medidas contínuas podem ser consideradas, como posição espacial, frequência, peso, entre outras (Ramsay e Silverman, 2005, p. 38). Dados funcionais são intrinsecamente infinitos, e trazem desafios tanto para a teoria quanto para a computação. A alta ou infinita dimensionalidade dos dados constitui uma fonte rica de informações, e pode criar inúmeras oportunidades para pesquisa e análise de dados (Wang et al., 2016). A aplicabilidade e utilidade da análise de componentes principais de dados funcionais (FPCA) já foi demonstrada em estudos da interação genótipos por ambientes (Krzyśko et al., 2013). Nesse sentido, o estudo apresentado no capítulo 2 deste trabalho foi conduzido com o objetivo de estudar GEI por meio do uso de variáveis ambientais no modelo AMMI sob abordagem funcional com diferentes volumes de dados e níveis de

desbalanceamento. Para isso, visou-se avaliar o comportamento da FPCA em cenários de desbalanceamento de dados e integrá-la ao método EM-AMMI para realizar a predição genotípica sob uma perspectiva funcional.

REFERÊNCIAS

- ABERA, W. et al. Heterosis and combining ability of elite maize inbred lines under northern corn leaf blight disease prone environments of the mid-altitude tropics. **Euphytica**, v. 208, n. 2, p. 391-400, 2016.
- ACOSTA-PECH, R. et al. Genomic models with genotype \times environment interaction for predicting hybrid performance: an application in maize hybrids. **Theoretical and applied genetics**, v. 130, n. 7, p. 1431-1440, 2017.
- BERNARDO JÚNIOR, L. A. Y. et al. AMMI bayesian models to study stability and adaptability in maize. **Agronomy Journal**, v. 110, n. 5, p. 1765-1776, 2018.
- GAUCH, H. G.; ZOBEL, R. W. Imputing missing yield trial data. **Theoretical and Applied Genetics**, v. 79, n. 6, p. 753-761, 1990.
- GAUCH, H. G. Statistical analysis of yield trials by AMMI and GGE. **Crop science**, v. 46, n. 4, p. 1488-1500, 2006.
- GAUCH, H. G.; PIEPHO, H. P.; ANNICCHIARICO, P. Statistical analysis of yield trials by AMMI and GGE: Further considerations. **Crop science**, v. 48, n. 3, p. 866-889, 2008.
- GAUCH, H. G. A simple protocol for AMMI analysis of yield trials. **Crop Science**, v. 53, n. 5, p. 1860-1869, 2013.
- GIRAUD, H. et al. Reciprocal genetics: Identifying QTL for general and specific combining abilities in hybrids between multiparental populations from two maize (*Zea mays* L.) heterotic groups. **Genetics**, v. 207, n. 3, p. 1167-1180, 2017.
- KADAM, D. C. et al. Genomic prediction of single crosses in the early stages of a maize hybrid breeding pipeline. **G3: Genes, Genomes, Genetics**, v. 6, n. 11, p. 3443-3453, 2016.
- KRZYŚKO, M. et al. Principal component analysis for functional data on grain yield of winter wheat cultivars. **Biometrical Letters**, v. 50, n. 2, p. 81-94, 2013.
- Ramsay, J. O.; Silverman, B. W. From functional data to smooth functions. In: **Functional Data Analysis**, Springer, New York, NY, p. 37-58, 2005.
- ROJAS, B. A.; SPRAGUE, G. F. A comparison of variance components in corn yield trials: III. General and specific combining ability and their interaction with locations and years. **Agronomy Journal**, v. 44, n. 9, p. 462-466, 1952.
- VAN EEUWIJK, F. A.; BUSTOS-KORTS, D. V.; MALOSETTI, M. What should students in plant breeding know about the statistical aspects of genotype \times environment interactions? **Crop Science**, v. 56, n. 5, p. 2119-2140, 2016.
- WANG, J. L.; CHIOU, J. M.; MÜLLER, H. G. Functional data analysis. **Annual Review of Statistics and Its Application**, v. 3, p. 257-295, 2016.
- ZOBEL, R. W.; WRIGHT, M. J.; GAUCH, H. G. Statistical analysis of a yield trial. **Agronomy journal**, v. 80, n. 3, p. 388-393, 1988.

SEGUNDA PARTE**CAPÍTULO 1 - AMMI-BAYESIAN MODEL AND USE OF CREDIBLE REGIONS IN
THE STUDY OF COMBINING ABILITY IN MAIZE**

Preliminary version of the article written according to the standard of the scientific journal Agronomy Journal. The editorial board of the journal can recommend changes in order to adjust it to its own style.

Core ideas

- The model proposed allows predict GCA and SCA effects and the interaction of these effects with environments with high accuracy in multi-environmental trials.
- The ellipses constructed in the biplot permit to infer about stability of GCA and SCA effects, and the influence of genotype x environment interaction.

AMMI-BAYESIAN MODEL AND USE OF CREDIBLE REGIONS IN THE STUDY OF COMBINING ABILITY IN MAIZE

Luiz Antonio Yanes Bernardo Júnior, Carlos Pereira da Silva, Indalécio Cunha Vieira Júnior,
Luciano Antonio de Oliveira, Marcio Balestre, Renzo Garcia Von Pinho

L.A.Y. Bernardo Júnior, I. C. Vieira Júnior, Federal University of Lavras, Department of Biology, Lavras, Minas Gerais, Brazil; C. P. da Silva, M. Balestre, Federal University of Lavras, Department of Statistics, Lavras, Minas Gerais, Brazil; L.A. de Oliveira, College of Exact Sciences and Technologies, Federal University of Grande Dourados, Dourados, Mato Grosso do Sul, Brazil; R.G. Von Pinho, Federal University of Lavras, Department of Agriculture, Lavras, Minas Gerais, Brazil.

Abbreviations: AMMI, Additive Main Effects and Multiplicative Interaction; BOA, Bayesian output analysis; SCA, specific combining ability; GCA, general combining ability; GEI, genotype x environment interaction; PCA, principal component analysis; SVD, singular value decomposition.

ABSTRACT

The development of lines with high performance and stability for synthesis of superior hybrids is the most expensive and time-consuming phase in the maize hybrid breeding. Several times, due to available resources, only a part of possible hybrid combinations is tested. Therefore, the breeder needs methods that allow the evaluation of genotypes untested in the field. This work was carried out with the objective of proposing a prediction model of GCA and SCA, and interactions with environments, associated with the use of credible regions in biplots obtained through AMMI-Bayesian model. Two analyses were done, in which the first one was conducted with simulated data, and the second one with real data. Credible ellipses were constructed in biplot in order to evaluate the stability of interaction effects for GCA and SCA. In general, for the analysis of simulated data, the predictions obtained had high correlation with the real values.

For the effects of GCA and SCA, the predictions kept the standard of signals and rank. Moreover, the model was efficient to provide credible intervals which covered the simulated values. For the analysis of real data, estimates of GCA and SCA for all genotypes evaluated do not differ from zero. The biplots for GCA x environment and SCA x environment interactions allowed evaluate the genotype stability in a more accurate way and the uncertainty about interaction estimates. The model is shown as a promising tool for helping the breeder take a decision about selecting and recommending genotypes.

INTRODUCTION

In plant breeding, multi-environmental trials play an important role in studying genotype x environment interaction (GEI), and selecting individuals of high genotypic expression and elevated stability (Acosta-Pech et al., 2017). The development and identification of promising lines for synthesis of superior hybrids is the most expensive and time-consuming phase in maize hybrid breeding (Abera et al., 2016). However, depending on the available resources, only a small fraction of possible hybrid combinations is tested. In this case, the prediction of hybrids performance using data of field trials obtained from related crossings becomes an important tool for identifying promising intergroup hybrids without testing them in the field (Schrag et al., 2009).

The heterosis is an important predictor of genetic value in hybrid breeding programs, and crossing lines with suitable genetic divergence plays a crucial role in expressing heterosis (Abera et al., 2016). The crossing of lines belonging to different heterotic groups allows the identification of promising crossings. The selection is performed in a group according to the performance in crossing with lines from another group, which allows maximizing the performance of hybrids obtained (Giraud et al., 2017). The accurate evaluation of lines in hybrid combinations is essential for getting success in developing commercial hybrids. Furthermore, the classification of lines in heterotic groups is essential to maximize the potential for developing hybrids and synthetics of high performance or identifying new heterotic groups (Amegbor, Badu-Apraku & Annor, 2017).

The selection can be done through diallel crosses among lines being studied and obtaining the combining ability among them. Likewise, the breeder may emphasize the average performance of lines through crosses done by the general combining ability (GCA) or only select determined hybrid combinations with higher performance through the specific combining ability (SCA). Understanding genetic basis of hybrid performance between two heterotic

groups is essential for conducting an efficient breeding program, and the effects of GCA and SCA are important indicators of line and hybrid potentials which will aid the breeder to find out ideal strategies for obtaining superior hybrids (Abera et al., 2016; Acosta-Pech et al., 2017; Giraud et al., 2017). The extensive evaluation of hybrids let the breeder perform the selection of parents in a more accurate way for recurrent selection programs or line development (Jumbo & Carena, 2008).

Jumbo and Carena (2008), evaluating populations and single maize hybrids, verified the predominance of additive effects for maize productivity. Kulka et al. (2018), studying hybrids obtained from crossings between tropical and tempered elite lines, also confirmed a higher contribution of GCA effects, which shows the importance of additive gene action and high frequency of favorable alleles. Therefore, the value of a test-cross hybrid of a line is the combination of GCA and SCA with a certain tester, and effects of SCA for maize show values that justify the use of several testers and a great number of resources (Giraud et al., 2017). Kadam et al. (2016) showed that the inclusion of SCA effects increases the predictive accuracy for productivity of single maize hybrids, and it can be more beneficial for hybrids which have untested parents than for hybrids with one or two tested parents.

The environmental effect and the interaction of genetic effects with environments are factors that have to be taken into consideration. Currently, one of the challenges in breeding programs is to find genotypes that are adapted, and at the same time stable in several cultivation regions. In the case of maize, the diversity of environments is especially high due to the second season. The combining ability does not guarantee the most productive hybrids be the most stable ones even though being extensively used by breeders in line selection. Likewise, several researchers have used methods that associate the combining ability and GEI (Rojas & Sprague, 1952; Matzinger, Sprague, & Cockerham, 1959; Beck, Vasal & Crossa, 1990; Nass et al., 2000; Balestre et al., 2008; Derera et al., 2008; Jumbo & Carena, 2008; Acosta-Pech et al., 2017; Adebayo et al., 2017; Amegbor, Badu-Apraku & Annor, 2017; Camargos et al., 2017; Kulka et al., 2018).

These kinds of studies demand extensive genotype evaluation in several environments. The evaluation of a large number of hybrid combinations in multi-environmental trials can facilitate the identification of hybrids with high performance and heterosis expression (Abera et al., 2016). Besides, the lines which show high performance in specific environments can be recombined to develop populations for extracting new sets of lines adjusted to a certain local, while lines with general high combining ability can be recombined for formation of a more robust recurrent population for developing productive lines with broad adaptability (Adebayo

et al., 2017). However, this kind of study can present some difficulties, such as predict hybrid performance in environments that was not evaluated (Acosta-Pech et al., 2017).

Acosta-Pech et al. (2017) demonstrated that the inclusion of interaction increases the predictive capacity of the model in predicting hybrid performance based only on parents' information. The authors observed a great reduction in residual variance when including GCA x environment and SCA x environment interactions, which indicates that it is responsible for a huge amount of total variance of genotypes through environments. Results that also demonstrate improvement in the predictive accuracy by the inclusion of GEI in the model were obtained by Burgueño et al. (2012), Jarquín et al. (2014) and Lopez-Cruz et al. (2015).

Other difficulties presented in this kind of study are heterogeneity of variance; different number of repetitions among experiments, which takes to a non-orthogonality of the analysis; and segregant populations of different sizes, which take to unbalanced contributions of initial lines for hybrids (Rojas & Sprague, 1952; Kadam et al., 2016; Giraud et al., 2017). These difficulties can be resolved through analyses under bayesian approach.

The AMMI analysis allows to obtain biplot graphics, which are broadly used by breeders in selection and recommendation of genotypes for high adaptability and stability. Yang et al. (2009) question the credibility provided by visual analysis of biplot, once genotypic and environmental scores presented in the graphic are point estimates without a statistical test. That way, recent works propose the construction of credible ellipses for scores of biplot using the AMMI Bayesian model (Crossa et al., 2011; de Oliveira et al., 2015; Bernardo Júnior et al., 2018). The ellipses obtained allowed performing a graphic analysis with a level of credibility, reducing the subjectivity of decision-making simply visual. Moreover, Bernardo Júnior et al. (2018) demonstrated the efficiency of AMMI Bayesian model in predicting GEI effect in a representative scenario of a breeding program, where the inclusion and removal of genotypes occurs throughout the selection cycles (de Figueiredo et al., 2015).

This work was carried out with the objective of proposing a prediction model of GCA and SCA, and their interactions with environments, associated with the use of credible regions in biplots obtained from AMMI Bayesian model.

MATERIAL AND METHODS

Bayesian Analysis of experiment groups using diallel model

In this study, diallel analysis and graphical analysis were associated through biplots for obtaining credible regions. In order to verify the applicability of proposed model, it was simulated a data set of 21 possible hybrids of 7 lines, without reciprocal, in 5 environments (Table 1). Also, it was simulated the singular values and singular vectors of GCA x environment and SCA x environment effects (Tables 2 and S3 to S10 of Supplemental Material).

The simulation was performed based on a randomized complete block design with three replications. For the SCA x environment matrix, two groups of five hybrids was simulated from a Gaussian distribution with large variance. It resulted in one group with positive marginal effect and one with negative marginal effect, which represent the unstable hybrids. Another group with 11 hybrids was also simulated from a Gaussian distribution with variances 1. It resulted in a group of hybrids with zero marginal effect, which represents the stable hybrids. The simulation of the GCA x environment interaction was performed in the same way. Two groups with two lines each was simulated from a Gaussian distribution with large variance. One group has positive marginal effect and the other one has negative marginal effect. The two groups represent the stable lines. Another group with three lines was simulated from a Gaussian distribution with variances 1. It resulted in a group of lines with marginal effect equal to zero, that is, stable. All Gaussian distributions had a mean of zero. The simulation was performed in R statistical software, version 2.15.2 (R Development Core Team, 2018).

Table 1. Simulated effects of GCA for seven lines (P1 to P7) and SCA for 21 hybrids F₁ obtained from two by two crossing, simulating a complete diallel without reciprocal.

Line	Effect	Hybrid	Effect	Hybrid	Effect	Hybrid	Effect
P1	2.988983	1x2	5.884364	2x4	0.701597	3x7	-4.13248
P2	1.028147	1x3	5.115463	2x5	-3.04453	4x5	-0.2488
P3	0.397292	1x4	4.48952	2x6	-0.01438	4x6	0.770553
P4	0.381383	1x5	4.050951	2x7	-2.98362	4x7	-2.78207
P5	0.228032	1x6	-1.57809	3x4	-0.17185	5x6	-2.73492
P6	-0.52903	1x7	-4.70707	3x5	0.388865	5x7	-1.33139
P7	-3.27055	2x3	-4.00874	3x6	-1.72551	6x7	0.025389

Table 2. Singular values simulated for each principal component of GCA x environment and SCA x environment for the analysis of 7 lines and 21 simulated hybrids in 5 environments.

	GCA x environment	SCA x environment
λ_1	10.2525	12.9457
λ_2	5.4597	7.2077
λ_3	2.7822	6.7538
λ_4	1.3118	4.7214

The applicability of the method in a set of real data was evaluated. For this, data of 73 hybrids obtained from 67 lines of data set used in the study of Pires (2017) were used. The experiments were carried out in 14 locals distributed in South and Southeast regions of Brazil (Table 3). These hybrids were evaluated by incomplete block design with variable number of repetitions. The hybrids in more advanced phases of assessment were evaluated with higher number of repetitions. The yield and Coefficient of Variation (CV) were obtained through individual analysis for each local using a mixed model with random genotype effects and fixed replicate effects.

Table 3. Environments utilized to conduct the experiments and evaluate the 73 hybrids with its respective average yield and Coefficient of Variation.

Environment	Local	Yield (t ha ⁻¹)	CV (%)
E1	Guarapuava-PR	10.163	16.451
E2	Vacaria-RS	5.289	20.867
E3	Ipiranga-PR	6.374	19.728
E4	Sananduva-RS	8.071	15.290
E5	Faxinal-PR	10.416	15.749
E6	Presidente Olegário-MG	8.985	13.507
E7	Uberaba-MG	9.509	13.895
E8	Capinópolis-MG	7.652	23.451
E9	Araguari-MG	8.116	14.496
E10	Madre de Deus-MG	9.485	17.875
E11	Nazareno-MG	7.627	17.638
E12	Boa Esperança-MG	6.349	24.871
E13	Lavras-MG	7.253	13.410
E14	Itapeva-SP	7.892	16.992

Adapted from Pires (2017).

The diallel scheme used in this study was method IV of Griffing (1956), in which just one set of F₁'s is included, without lines and reciprocals. The whole analysis was performed

under Bayesian perspective. The diallel model for groups of experiments can be represented in the following way:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \mathbf{W}\mathbf{p} + \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k) \mathbf{X}_2 \boldsymbol{\gamma}_k + \sum_{i=1}^j \tau_i \text{diag}(\mathbf{W}\boldsymbol{\rho}_i) \mathbf{X}_2 \boldsymbol{\delta}_i + \boldsymbol{\varepsilon}$$

in which \mathbf{y} is the vector of phenotypic answers composed of $n = lr$ observations, in which r is the number of hybrids and $l = bc$ is the number of repetitions obtained by confounding of blocks and environments effects, in other words, blocks are nested with environments; b and c represent the number of blocks and environments, respectively; $\boldsymbol{\beta}$ is the effect of block and environment vector nested; \mathbf{g} is the vector of GCA main effects; \mathbf{p} is the vector of SCA main effects; $\boldsymbol{\alpha}_k$ and $\boldsymbol{\gamma}_k$ are singular vectors of SCA and environments; λ_k is the singular value of k vector with $k = 1, \dots, t$, in which $t = \min(r, c)$ is the matrix rank of SCA x environment interaction; $\boldsymbol{\rho}_i$ and $\boldsymbol{\delta}_i$ are singular vectors of GCA and environments; τ_i is the singular value for j -th vector with $i = 1, \dots, j$, in which $j = \min(d, c)$ is the matrix rank of GCA x environment interaction and d is the number of lines; \mathbf{X}_1 , \mathbf{X}_2 , \mathbf{Z} and \mathbf{W} are incidence matrices; $\boldsymbol{\varepsilon}_{n \times 1}$ is the vector of random errors, in which $\boldsymbol{\varepsilon} \sim \mathbf{N}_n(0, \sigma_e^2 \mathbf{I}_n)$, with $\mathbf{0}$ representing the vector of zeros; σ_e^2 is the error variance; \mathbf{I}_n is the identity matrix.

Priori distributions for parameters of model and function of likelihood

The priori distributions for all estimated parameters were attributed based on the work presented by de Oliveira et al. (2015). Priori multivariate distributions to vectors $\boldsymbol{\beta}$, \mathbf{g} and \mathbf{p} are attributed. For vector $\boldsymbol{\beta}$, a high variance was considered, indicated by constant. The random SCA effect was obtained attributing a priori hierarchic for \mathbf{g} . Then σ_g^2 has a proportional distribution at $(\sigma_g^2)^{-1}$, result of assuming zero value for the hiperparameters (v_g, S_g^2) for density of chi-square inverse scaled distributions. The uncertainty related to estimate of \mathbf{g} is basically determined by likelihood function, in other words, through data in analysis, due to the fact of a priori assumed for σ_g^2 is not informative. The random GCA effect was also obtained attributing a priori hierarchic for \mathbf{p} . Thus, σ_p^2 has a proportional distribution at $(\sigma_p^2)^{-1}$, result of assuming zero value for the hiperparameters (v_p, S_p^2) for density of chi-square inverse scaled distributions.

Similarly to \mathbf{g} , the uncertainty related to estimate of \mathbf{p} is basically determined by likelihood function, in other words, through data in analysis because the priori assumed for σ_p^2 is not informative either.

For singular vectors (\mathbf{a}_k and γ_k) and (ρ_i and δ_i) uniform spherical distributions are attributed in the correct space (Viele and Srinivasan, 2000). These vectors are distributed in a restricted space at \mathbb{R}^p and \mathbb{R}^q , being orthogonal at $t-1$ and $j-1$ vectors in the spaces of dimensions p ($p = r$ or $p = c$) and q ($q = r$ or $q = d$), respectively. The uniform spherical distribution is a non-informative distribution. For singular values, priori normal truncated distributions are attributed because these variables only assume positive values, in other words, $\lambda_k \geq \lambda_{k+1} \geq 0$ and $\tau_i \geq \tau_{i+1} \geq 0$ in both interaction matrices.

These priors can be summarized by $p(\boldsymbol{\beta}) = \frac{1}{\sigma_\beta^2} = c$, $g | \mu_g, \sigma_g^2 \sim N(\mathbf{0}, \mathbf{I}\sigma_g^2)$, $p | \mu_p, \sigma_p^2 \sim N(\mathbf{0}, \mathbf{I}\sigma_p^2)$, $p(\lambda_k) \sim \frac{1}{\sigma_{\lambda_k}} = K$, $p(\lambda_{\tau_i}) \sim \frac{1}{\sigma_{\tau_i}} = C$, $p(\sigma_g^2) \propto \frac{1}{\sigma_g^2}$, $p(\sigma_p^2) \propto \frac{1}{\sigma_p^2}$ and $p(\sigma_e^2) \propto \frac{1}{\sigma_e^2}$.

The likelihood used in this study is the same one used by de Oliveira et al. (2015) and da Silva et al. (2015), in which the data are modeled by

$$y \sim N\left(\mathbf{X}_1\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \mathbf{W}\mathbf{p} + \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\mathbf{a}_k)\mathbf{X}_2\boldsymbol{\gamma}_k + \sum_{i=1}^j \tau_i \text{diag}(\mathbf{W}\rho_i)\mathbf{X}_2\boldsymbol{\delta}_i, \mathbf{I}\sigma_e^2\right).$$

Gibbs Algorithm

Combined likelihood information with the priori distributions, a joint posteriori is obtained for the parameters from where a posteriori complete distributions will be obtained.

$$P(\Phi | y) \propto L(\theta | y)P(\boldsymbol{\beta} | \mu_\beta, \sigma_\beta^2)P(g | \mu_g, \sigma_g^2)P(p | \mu_p, \sigma_p^2)P(\sigma_g^2 | v_g, S_g^2)P(\sigma_p^2 | v_p, S_p^2) \times \\ \times P(\sigma_e^2 | v_e, S_e^2) \prod_{k=1}^t P(\lambda_k | \mu_{\lambda_k}, \sigma_{\lambda_k}^2)P(\mathbf{a}_k)P(\boldsymbol{\gamma}_k) \prod_{i=1}^j P(\tau_i | \mu_{\tau_i}, \sigma_{\tau_i}^2)P(\rho_i)P(\boldsymbol{\delta}_i)$$

in which $\Phi = (\boldsymbol{\beta}, \mathbf{g}, \mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\rho}, \boldsymbol{\delta}, \sigma_g^2, \sigma_p^2, \sigma_e^2)$.

It were simulated Markov Chains through 50000 iterations for each parameter of the model. For each chain, it were realized burn-in of 8000 first iterations and thinning at each 8 observations. The convergence criteria used were according to Raftery and Lewis (1992) and

Heidelberger and Welch (1983). The method of Monte Carlo via Markov Chains (MCMC) based on Gibbs algorithm can be described as:

Step 1: To be sampled $\boldsymbol{\beta}$ of a normal distribution with average and variance given by

$$\boldsymbol{\beta} | \dots \sim N \left[\left(\mathbf{X}_1^\top \mathbf{X}_1 \right)^{-1} \mathbf{X}_1^\top (\mathbf{y} - \mathbf{Z}\mathbf{g} - \mathbf{W}\mathbf{p} - \mathbf{A}_1 - \mathbf{A}_2), \left(\mathbf{X}_1^\top \mathbf{X}_1 \right)^{-1} \sigma_e^2 \right]$$

$$\text{in which } \mathbf{A}_1 = \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\mathbf{a}_k) \mathbf{X}_2 \boldsymbol{\gamma}_k \text{ and } \mathbf{A}_2 = \sum_{i=1}^j \tau_i \text{diag}(\mathbf{W}\rho_i) \mathbf{X}_2 \delta_i.$$

Step 2: To be sampled \mathbf{g} of a normal distribution with average and variance given by

$$\mathbf{g} | \dots \sim N \left[\left(\mathbf{Z}^\top \mathbf{Z} + \mathbf{I} \frac{\sigma_e^2}{\sigma_g^2} \right)^{-1} \mathbf{Z}^\top (\mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta} - \mathbf{W}\mathbf{p} - \mathbf{A}_1 - \mathbf{A}_2), \left(\mathbf{Z}^\top \mathbf{Z} + \mathbf{I} \frac{\sigma_e^2}{\sigma_g^2} \right)^{-1} \sigma_e^2 \right]$$

Step 3: To be sampled \mathbf{p} of a normal distribution with average and variance given by

$$\mathbf{p} | \dots \sim N \left[\left(\mathbf{W}^\top \mathbf{W} + \mathbf{I} \frac{\sigma_e^2}{\sigma_p^2} \right)^{-1} \mathbf{W}^\top (\mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta} - \mathbf{Z}\mathbf{g} - \boldsymbol{\Theta}), \left(\mathbf{W}^\top \mathbf{W} + \mathbf{I} \frac{\sigma_e^2}{\sigma_p^2} \right)^{-1} \sigma_e^2 \right]$$

Step 4: To be sampled singular value λ_k of a normal truncated distribution with average

$$\text{and variance given by } \lambda_k | \dots \sim N^+ \left[\left(\mathbf{A}_{3k}^\top \mathbf{A} \right)^{-1} \mathbf{A}_{3k}^\top \Delta_{1k}, \left(\mathbf{A}_{3k}^\top \mathbf{A}_{3k} \right)^{-1} \sigma_e^2 \right]$$

$$\text{in which } \Delta_{1k} = \mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta} - \mathbf{Z}\mathbf{g} - \mathbf{W}\mathbf{p} - \sum_{k \neq k'}^t \lambda_k \text{diag}(\mathbf{Z}\mathbf{a}_{k'}) \mathbf{X}_2 \boldsymbol{\gamma}_{k'} - \mathbf{A}_2,$$

$$\mathbf{A}_{3k} = \text{diag}(\mathbf{Z}\mathbf{a}_k) \mathbf{X}_2 \boldsymbol{\gamma}_k \text{ and } \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_t \geq 0.$$

Step 5: To be sampled the k -th singular vector \mathbf{a}_k referent to SCA x environment interaction of a von Mises-Fisher distribution with directional vector and parameters of

$$\text{concentration given by } \mathbf{a}_k^* \sim \text{VMF} \left(r - s, \frac{c_k \lambda_k}{\sigma_e^2}, \tilde{\boldsymbol{\mu}}_{\alpha_k} \right), \quad \tilde{\boldsymbol{\mu}}_{\alpha_k} = c_k^{-1} \mathbf{H}_k^\top \boldsymbol{\mu}_{\alpha_k},$$

$$c_k = \sqrt{\boldsymbol{\mu}_{\alpha k}^\top \mathbf{H}_k \mathbf{H}_k^\top \boldsymbol{\mu}_{\alpha k}}; \mathbf{H}_k \text{ is a matrix in which } H_k \text{ columns form a orthonormal set of}$$

$$\boldsymbol{\alpha} \text{ vectors, and } \boldsymbol{\mu}_{\alpha k} = \mathbf{A}_5^\top (\mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta} - \mathbf{Z}\mathbf{g} - \mathbf{W}\mathbf{p} - \mathbf{A}_2), \text{ in which } \mathbf{A}_5 = \text{diag}(\mathbf{X}_2 \boldsymbol{\gamma}_k) \mathbf{Z}.$$

Therefore, \mathbf{a}_k^* can be sampled in S_{r-s} , and to obtain a vector in the right subspace, in \mathbb{R}^p , it is simply necessary to execute inverse transformation given by $\mathbf{a}_k = \mathbf{H}_k \mathbf{a}_k^*$.

Step 6: To be sampled the k -th singular vector γ_k referent to SCA x environment interaction of a von Mises-Fisher distribution with directional vector and parameter of

concentration given by $\gamma_k^* \sim VMF\left(r-s, \frac{d_k \lambda_k}{\sigma_e^2}, \tilde{\boldsymbol{\mu}}_{\gamma_k}\right)$, $\tilde{\boldsymbol{\mu}}_{\gamma_k} = d_k^{-1} \mathbf{R}_k^\top \boldsymbol{\mu}_{\gamma_k}$,

$d_k = \sqrt{\boldsymbol{\mu}_{\gamma_k}^\top \mathbf{R}_k \mathbf{R}_k^\top \boldsymbol{\mu}_{\gamma_k}}$; \mathbf{R}_k is a matrix in which R_k columns form an orthonormal set of

γ vectors, and $\boldsymbol{\mu}_{\gamma_k} = \mathbf{A}_6^\top (\mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta} - \mathbf{Z} \mathbf{g} - \mathbf{W} \mathbf{p} - \mathbf{A}_2)$, in which $\mathbf{A}_6 = \text{diag}(\mathbf{Z} \mathbf{a}_k) \mathbf{X}_2$.

Therefore, γ_k^* can be sampled in S_{r-s} , and to obtain in the right subspace, in \mathbb{R}^p , it is simply necessary to execute inverse transformation given by $\gamma_k = \mathbf{R}_k \gamma_k^*$.

Step 7: To be sampled the singular value τ_i of a normal truncated distribution with

average and variance given by $\tau_i | \dots \sim N^+ \left[\left(\mathbf{A}_{4i}^\top \mathbf{A}_{4i} \right)^{-1} \mathbf{A}_{4i}^\top \Delta_{2i}, \left(\mathbf{A}_{4i}^\top \mathbf{A}_{4i} \right)^{-1} \sigma_e^2 \right]$

in which $\Delta_{2i} = \mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta} - \mathbf{Z} \mathbf{g} - \mathbf{W} \mathbf{p} - \sum_{i \neq i'}^j \tau_{i'} \text{diag}(\mathbf{W} \boldsymbol{\rho}_{i'}) \mathbf{X}_2 \boldsymbol{\delta}_{i'} - \mathbf{A}_1$, $\mathbf{A}_{4i} = \text{diag}(\mathbf{W} \boldsymbol{\rho}_i) \mathbf{X}_2 \boldsymbol{\delta}_i$

and $\tau_1 \geq \tau_2 \geq \dots > \tau_j \geq 0$.

Step 8: To be sampled the i -th singular vector ρ_i referring to GCA x environment interaction of a von Mises-Fisher distribution with directional vector and parameter of

concentration given by $\rho_i^* \sim VMF\left(r-s, \frac{c_i \tau_i}{\sigma_e^2}, \tilde{\boldsymbol{\mu}}_{\rho_i}\right)$, $\tilde{\boldsymbol{\mu}}_{\rho_i} = c_i^{-1} \mathbf{Q}_i^\top \boldsymbol{\mu}_{\rho_i}$, $c_i = \sqrt{\boldsymbol{\mu}_{\rho_i}^\top \mathbf{Q}_i \mathbf{Q}_i^\top \boldsymbol{\mu}_{\rho_i}}$;

\mathbf{Q}_i is a matrix which Q_k columns form an orthonormal set of ρ vectors, and

$\boldsymbol{\mu}_{\rho_i} = \mathbf{A}_7^\top (\mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta} - \mathbf{Z} \mathbf{g} - \mathbf{W} \mathbf{p} - \mathbf{A}_1)$ in which $\mathbf{A}_7 = \text{diag}(\mathbf{X}_2 \boldsymbol{\delta}_i) \mathbf{W}$. Therefore, ρ_i^* can be

sampled in S_{r-s} , and to obtain vector in the right subspace, in \mathbb{R}^q , it is simply necessary

to execute inverse transformation given by $\rho_i = \mathbf{Q}_i \rho_i^*$.

Step 9: To be sample the k -th singular vector δ_i referring to GCA x environment interaction of von Mises-Fisher distribution with directional vector and parameter of concentration given by $\delta_i^* \sim VMF\left(r-s, \frac{d_i \tau_i}{\sigma_e^2}, \tilde{\boldsymbol{\mu}}_{\delta_i}\right)$, $\tilde{\boldsymbol{\mu}}_{\delta_i} = d_i^{-1} \mathbf{B}_i^\top \boldsymbol{\mu}_{\delta_i}$, $d_k = \sqrt{\boldsymbol{\mu}_{\delta_i}^\top \mathbf{B}_i \mathbf{B}_i^\top \boldsymbol{\mu}_{\delta_i}}$; \mathbf{B}_i is a matrix which B_i columns form an orthonormal set o δ vectors, and $\boldsymbol{\mu}_{\delta_i} = \mathbf{A}_8^\top (\mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta} - \mathbf{Z} \mathbf{g} - \mathbf{W} \mathbf{p} - \mathbf{A}_2)$, in which $\mathbf{A}_8 = \text{diag}(\mathbf{W} \delta_i) \mathbf{X}_2$. Therefore, δ_i^* can be sampled in S_{r-s} , and to obtain vector in the right subspace, in \mathbb{R}^q , it is simply necessary to execute inverse transformation given by $\delta_i = \mathbf{B}_i \delta_i^*$.

Step 10: To be sampled σ_g^2 of a chi-square inverse scaled distribution $\sigma_g^2 | \dots \sim \text{inv-}\chi^2[n_g, \mathbf{g}^\top \mathbf{g}]$ in which n_g is the number of hybrids.

Step 11: To be sampled σ_p^2 of a chi-square inverse scaled distribution $\sigma_p^2 | \dots \sim \text{inv-}\chi^2[n_p, \mathbf{p}^\top \mathbf{p}]$ in which n_p is the number of lines.

Step 12: To be sampled σ_e^2 of a chi-square inverse scaled distribution given by $\sigma_e^2 | \dots \sim \text{inv-}\chi^2[n, (\mathbf{y} - \boldsymbol{\theta})^\top (\mathbf{y} - \boldsymbol{\theta})]$ in which n is the length of \mathbf{y} and $\boldsymbol{\theta} = \mathbf{X}_1 \boldsymbol{\beta} + \mathbf{Z} \mathbf{g} + \mathbf{W} \mathbf{p} + \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z} \boldsymbol{\alpha}_k) \mathbf{X}_2 \boldsymbol{\gamma}_k + \sum_{i=1}^j \tau_i \text{diag}(\mathbf{W} \boldsymbol{\rho}_i) \mathbf{X}_2 \boldsymbol{\delta}_i$.

Convergence and credible intervals in biplot

The diagnostic of convergence was realized according to the method of Raftery and Lewis (1992), and criteria of Heidelberger and Welch (1983), using the library Bayesian output analysis (BOA) (Smith, 2007) available in R statistical software, version 2.15.2 (R Development Core Team, 2018).

The credible intervals HPD (Highest Probability Density) for parameters were constructed according to the method proposed by Chen and Shao (1999) also available in the BOA library.

The credible intervals at 95% in biplot were constructed for the two first principal components according to biplot scores $(\lambda_1^{1/2}\mathbf{a}_1, \lambda_2^{1/2}\mathbf{a}_2)$ and $(\lambda_1^{1/2}\boldsymbol{\gamma}_1, \lambda_2^{1/2}\boldsymbol{\gamma}_j)$ referents to SCA x environment interaction, and $(\tau_1^{1/2}\boldsymbol{\rho}_1, \tau_2^{1/2}\boldsymbol{\rho}_2)$ and $(\tau_1^{1/2}\boldsymbol{\delta}_1, \tau_2^{1/2}\boldsymbol{\delta}_2)$ referents to GCA x environment interaction using the method described in de Oliveira et al. (2015).

EM-REML

To evaluate predictions obtained with Bayesian analysis, EM-REML analysis was done (Expectation Maximization Restricted Maximum Likelihood) for comparing results. The model used was adapted from the model proposed by Guedes (2012), and can be described as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\boldsymbol{\alpha} + \mathbf{Z}_2\boldsymbol{\omega} + \mathbf{Z}_3\boldsymbol{\theta} + \mathbf{Z}_4\boldsymbol{\tau} + \mathbf{e}$$

in which \mathbf{y} is the average value of plot; \mathbf{X} is the incidence matrix of fixed effects; \mathbf{Z}_1 is the incidence matrix of GCA effects; \mathbf{Z}_2 is the incidence matrix of SCA effects; \mathbf{Z}_3 is the incidence matrix of GCA x environment interaction; \mathbf{Z}_4 is the incidence matrix of SCA x environment interaction; $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, $\boldsymbol{\omega}$, $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ are vectors of effects \mathbf{X} , \mathbf{Z}_1 , \mathbf{Z}_2 , \mathbf{Z}_3 and \mathbf{Z}_4 , respectively; \mathbf{e} is the error.

The heritability estimates were obtained (h^2) for unbalanced experiments with different number of repetitions adapted from Holland, Nyquist, and Cervantes-Martínez (2003) and Schmidt et al. (2019):

$$h^2 = \frac{\sigma_{\text{GCA}}^2}{\sigma_{\text{P}}^2} ; h^2 = \frac{\sigma_{\text{SCA}}^2}{\sigma_{\text{P}}^2}$$

$$\sigma_{\text{P}}^2 = \sigma_{\text{GCA}}^2 + \sigma_{\text{SCA}}^2 + \frac{\sigma_{\text{GCA} \times \text{environment}}^2 + \sigma_{\text{SCA} \times \text{environment}}^2}{c} + \frac{\sigma_{\text{e}}^2}{c\bar{r}}$$

in which σ_{GCA}^2 is the component of variance for GCA; σ_{SCA}^2 is the component of variance for SCA; σ_{P}^2 is the component of phenotypic variance; $\sigma_{\text{GCA} \times \text{environment}}^2$ and $\sigma_{\text{SCA} \times \text{environment}}^2$ are variance components of GCA x environment and SCA x environment interactions, respectively; c is the number of environments; and \bar{r} is the harmonic average of maximum number of repetitions of each local.

The biplots were built in the conventional way, according to Zobel, Wright, and Gauch (1988). The singular value decomposition (SVD) was performed in the GCA x environment and SCA x environment matrices to obtain the scores for genotypes and environments. The genotype and environment scores are expressed as a unit vector times the square root of the eigenvalue (Gabriel, 1971; Bradu and Gabriel, 1978):

$$\lambda\zeta_g\eta_e = (\lambda^{0.5}\zeta_g)(\lambda^{0.5}\eta_e)$$

in which $\lambda^{0.5}\zeta_g$ is the genotype score and $\lambda^{0.5}\eta_e$ is the environment score. The estimated interaction is obtained through the multiplication of a genotype score by an environment score.

All analyses were done using R software, version 2.15.2 (R Development Core Team, 2018) and SAS software, version 9 (SAS Institute Inc., 2002).

RESULTS

Analysis of Simulated Data

The mean estimated effects, standard deviations and credible intervals at 95% of GCA for the seven lines are indicated in Table S1 of the Supplemental Material. The intervals of effects are represented in Figure 1 in ascending order from left to right. The credible intervals showed intervals with similar ranges. This is expected for balanced data analyses (Bernardo Júnior et al., 2018).

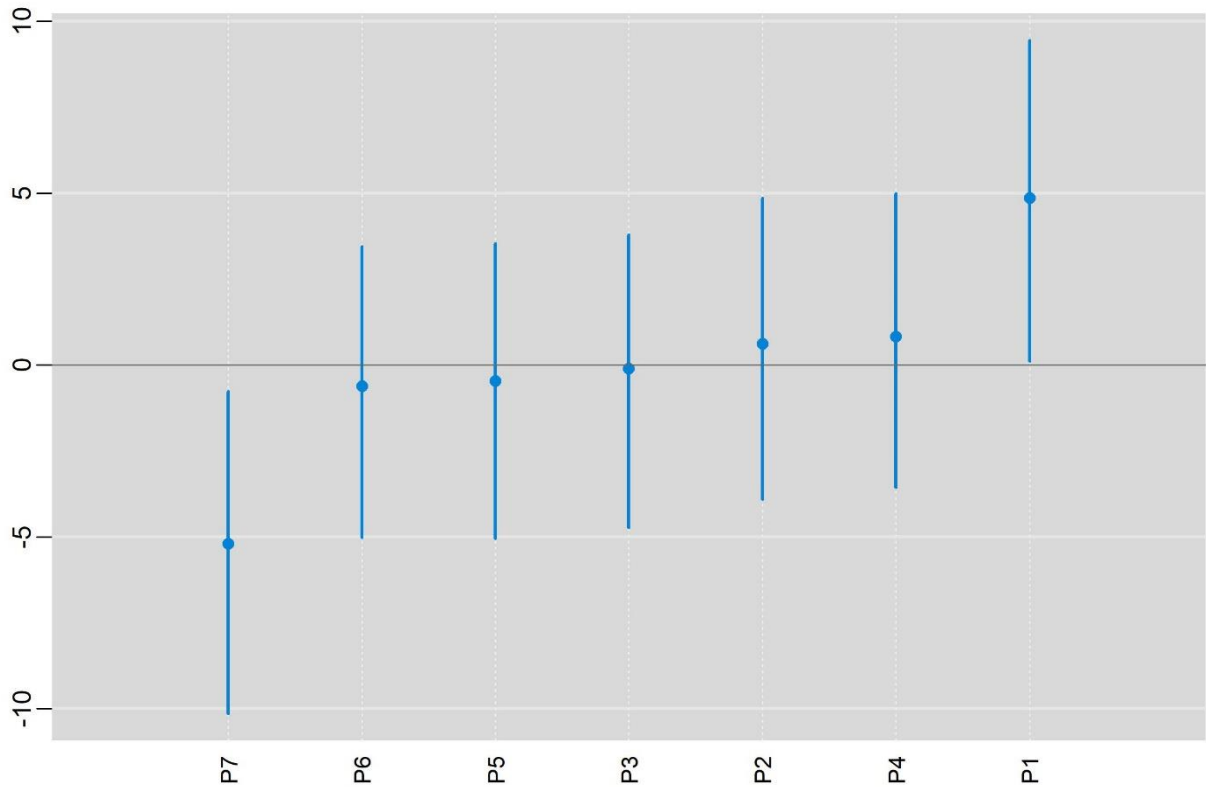


Figure 1. Point estimates and its respective maximum density posteriori intervals at level of 95% of credibility for the GCA simulated effects of 7 lines in 5 environments.

From credible intervals obtained by the model, it is possible to infer that P1 line was the only one to present GCA effects significantly higher than zero. The lines P6, P5, P3, P2 and P4 showed credible intervals for GCA that encompassed zero, in other words, the means did not differ significantly from the general mean. P7 line showed GCA effects significantly different from zero and lower than the general mean.

The average effects, standard deviations and credible intervals at 95% for SCA are represented in Table S2 of Supplemental Material. The credible intervals are represented in Figure 2. As well as for intervals of GCA effects of lines, SCA intervals are represented in ascending order from left to right. Again, the credible intervals showed intervals with similar ranges.

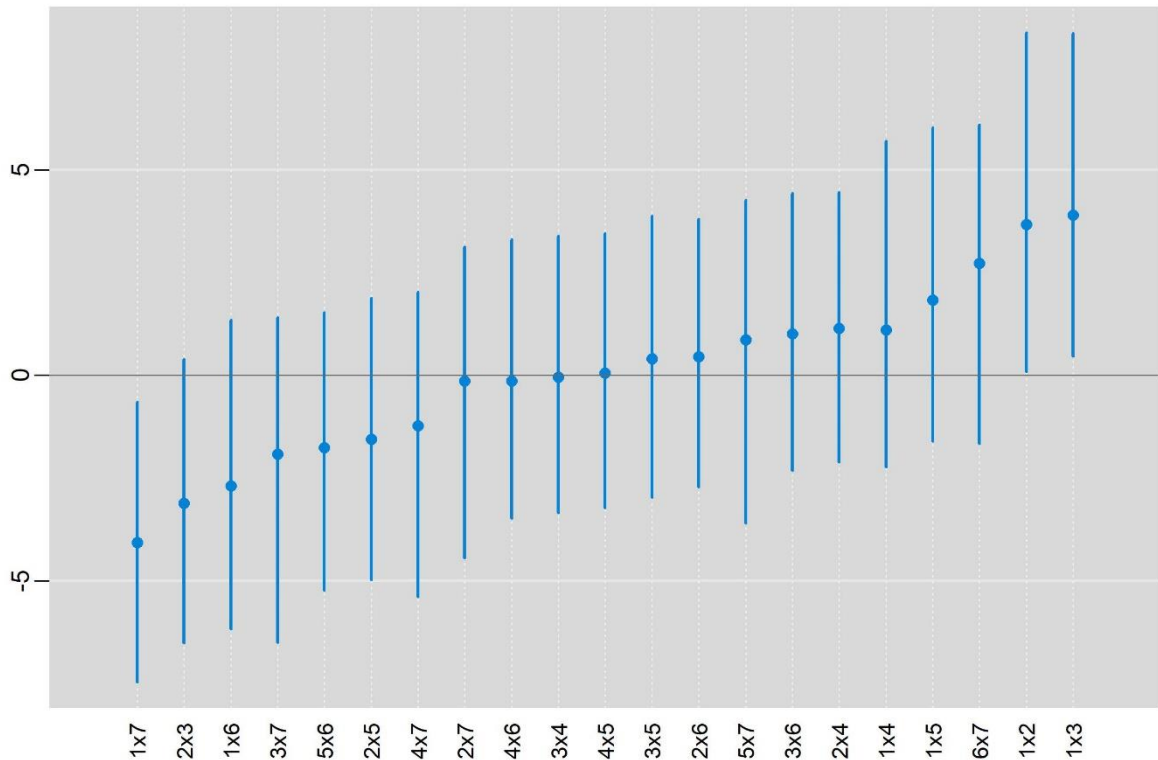


Figure 2. Point estimates and its respective maximum density posteriori intervals at level of 95% of credibility for the SCA simulated effects of 21 hybrids in 5 environments.

The hybrids 1x2 and 1x3 showed SCA effects significantly different from zero and over the general mean, and therefore, would be genotypes of higher expression of dominance and epistatic effects. The hybrids 2x3, 1x6, 3x7, 5x6, 2x5, 4x7, 2x7, 4x6, 3x4, 4x5, 3x5, 2x6, 3x6, 2x4, 1x4, 1x5 and 6x7 showed credible intervals which included the zero value. Therefore, the averages did not significantly differ from general mean. The hybrid 1x7 showed SCA effects significantly lower than zero, in other words, the effects were significantly lower than the general mean.

According to Cruz, Regazzi, and Carneiro (2012), the hybrid combinations with higher interest are those that present higher SCA estimates, and which were obtained from at least one parent with high GCA. In this study, the hybrids 1x2 and 1x3 were the only ones to show SCA effects significantly higher than zero. These hybrids were obtained from crossings that involved P1 line, which was the only one with GCA effects significantly over general mean. Therefore, the hybrids 1x2 and 1x3 would be the best ones in performance due to the higher expression of non-additives effects, and thus, higher expression of heterosis.

Table 4 presents estimates of variance components for GCA, SCA and error, standard deviations and credible intervals. The credible intervals for

σ_{GCA}^2 indicate that GCA effects did not differ significantly from zero, and therefore, additive effects were not significant. This shows a higher importance of dominance effects, represented by σ_{SCA}^2 .

Table 4. Estimates of variance components, standard deviation and credible intervals obtained by model for GCA, SCA and error.

	Mean	σ	Lowe Limit	Upper Limit
σ_{GCA}^2	14.6669	24.5392	0.0000	61.4696
σ_{SCA}^2	7.1760	4.8797	2.5111	17.8612
σ_e^2	7.9212	10.4835	5.0866	17.9220

The greatest part of values predicted for GCA and SCA did not differ significantly from zero based on posteriori values and credible intervals. In general, the point predictions were close to simulated values (Tables S1 and S2). The Pearson correlation between the simulated and predicted values of GCA and SCA effects was of 0.9830 and 0.8564, respectively. Among the seven predicted values for GCA, five of them presented the same signal for simulated effects. For SCA, 16 of the 21 predictions kept the signals. Likewise, the positive and negative simulated effects were kept in the predictions. Furthermore, all credible intervals of predicted effects for GCA and SCA had the real simulated value. For GCA effects, there was little change in the ranking between simulated and predicted values of lines. For SCA effects, the change was greater. However, among the eight simulated hybrids with SCA positive effects, six of them were kept in better rankings. This shows the efficiency of the method to predict GCA and SCA and determine the significance of predictions in multi-environmental trials. Moreover, except for P1 and P7 lines, the shrinkage effect was highlighted, typical of predictions via Bayesian inference using non-informative prioris (Viele and Srinivasan, 2000; Crossa et al. 2011; de Oliveira et al., 2015), which allow the reduction of variance for predicted effects.

The model also had high accuracy in prediction of singular values for each principal component. The Pearson correlations between simulated and predicted values of GCA x environment and SCA x environment interactions were of 0.9943 and 0.9702, respectively. Except for the singular value of the fourth principal component referred to SCA x environment interaction, all real values were included in the credible intervals of (Tables 6 and 7). This result shows a great advantage of Bayesian inference due to the fact of being possible to obtain credible intervals that have only positive values, which would be more difficult in the classical approach considering asymptotic distribution. The analysis under Bayesian optic allows the exact calculation of a posteriori distribution, and substitutes the uncertainty of asymptotic

approximation by uncertainty of determining priori weight on posteriori (Viele and Srinivasan, 2000).

The simulated and predicted values of singular vectors for the two first principal components of GCA x environment and SCA x environment interactions are indicated in Supplemental Material (Tables S3 to S10). For singular vectors of genotype and environmental effects for GCA x environment interaction, the correlations between real values and mean and median were high, with values that vary from 0.7839 to 0.9963. Besides, all credible intervals covered simulated value. On the other hand, for singular vectors of SCA x environment interaction the correlation were low, with values that vary from -0.3189 to 0.3764. This can have occurred because of great difficulty of estimating parameters related to SCA. However, almost all credible intervals obtained had the simulated value.

The decomposition through singular values of GCA x environment and SCA x environment interactions allowed capturing 98.3713% and 95.1034% of interaction with the two first principal components, respectively (Tables 5 and 6). Biplots represented in Figures 3 to 6 describe interaction standards for GCA x environment and SCA x environment effects, respectively.

Table 5. Singular value and proportion of GCA x environment interaction explained for each principal component for analysis of 7 simulated lines in 5 environments.

	Simulated Value	Mean	σ	Lower Limit	Upper Limit	% Explained	% Accumulated
λ_1	10.2525	10.3429	3.5820	0.7594	13.6173	84.6928	84.6928
λ_2	5.4597	4.1566	2.0802	0.0001	6.2249	13.6785	98.3713
λ_3	2.7822	1.3490	1.1391	0.0001	3.0105	1.4407	99.8120
λ_4	1.3118	0.4873	0.5190	0.0000	1.5644	0.1880	100

Table 6. Singular values and proportion of SCA x environment interaction explained for each principal component for analysis of 21 simulated hybrids in 5 environments.

	Simulated Value	Mean	σ	Lower Limit	Upper Limit	% Explained	% Accumulated
λ_1	12.9457	10.8274	7.7651	0.2116	28.5655	75.2582	75.2582
λ_2	7.2077	5.5600	4.1011	0.0004	12.0876	19.8452	95.1034
λ_3	6.7538	2.4984	2.7826	0.0002	8.9172	4.0071	99.1105
λ_4	4.7214	1.1771	1.5265	0.0000	4.2685	0.8895	100

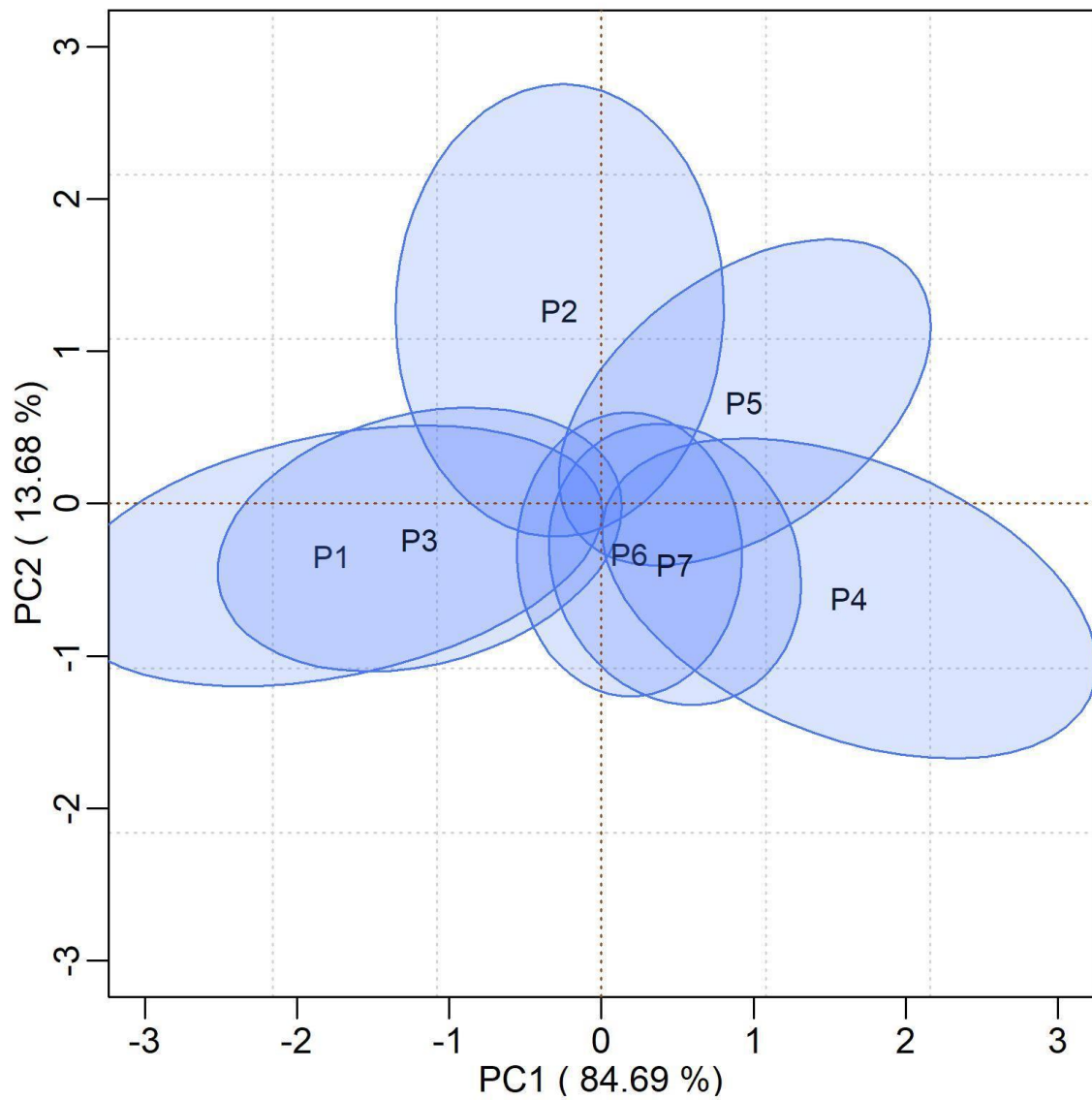


Figure 3. Maximum density posteriori intervals at level of 95% of credibility for the two first principal components related to effects of GCA x environment interaction. In this biplot only ellipses of lines are represented.

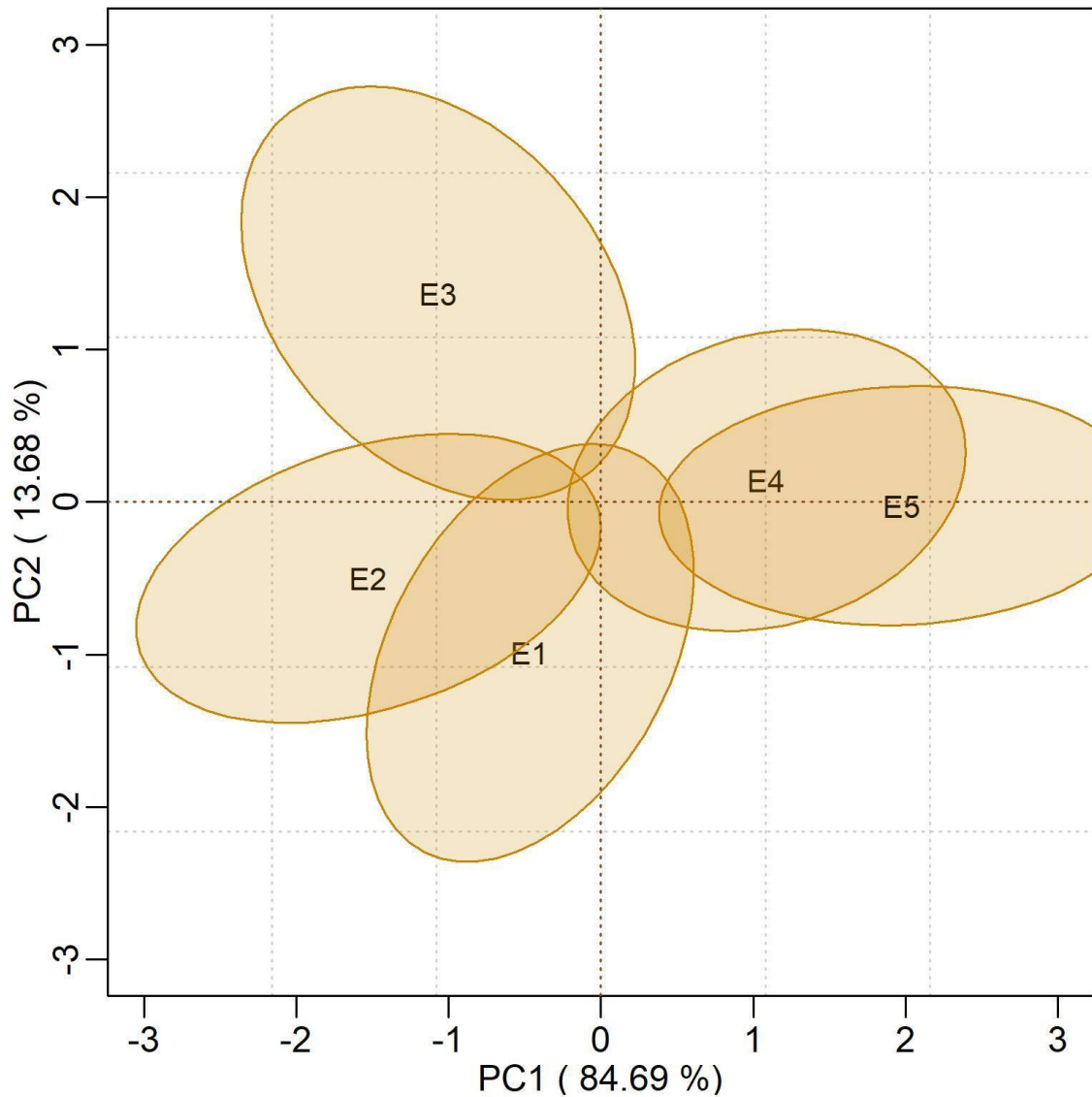


Figure 4. Maximum density posteriori intervals at level of 95% of credibility for the two first principal components related to the effects of GCA x environment interaction. In this biplot only environment ellipses are represented.

Figure 3 indicates high uncertainty around predictions obtained for all lines under study indicated by the huge size of the ellipses. Moreover, except for P4 line ellipse, all credible ellipses have the origin of biplot [0,0], in other words, encompass zero in the two first principal components. Thus, these lines do not suffer significant influence of GCA x environment interaction effects and can not be considered as unstable GCA effects. Although P4 line does not have the origin of biplot, it encompasses zero in the second principal component. In addition, there is an overlapping with P2, P3, P5, P6 and P7 ellipses. These facts indicate that P4 line has significant differences in interaction pattern just from P1 line in the first component.

The same inferences can be drawn for the environments from Figure 4. The ellipses also show the high uncertainty observed in Figure 3. The environments E1 and E4 presented ellipses containing the origin. Thus, these environments showed low contribution for GCA x environment interaction due to the posterior ones not allows to reject biplot origin. E2 and E5 environments not encompass zero value in the first principal component, and E3 does not include zero in the second principal component. Therefore, they are inconclusive in terms of their contribution to GEI.

As for GCA x environment interaction, there was great uncertainty for SCA x environment interaction (Figures 5 and 6). Both genotypic ellipses and environmental ones have origin [0,0], which demonstrates huge uncertainty existent around the estimates. In this case, the hybrids do not suffer significant influence of interaction as well as environments do not contribute for its manifestation.

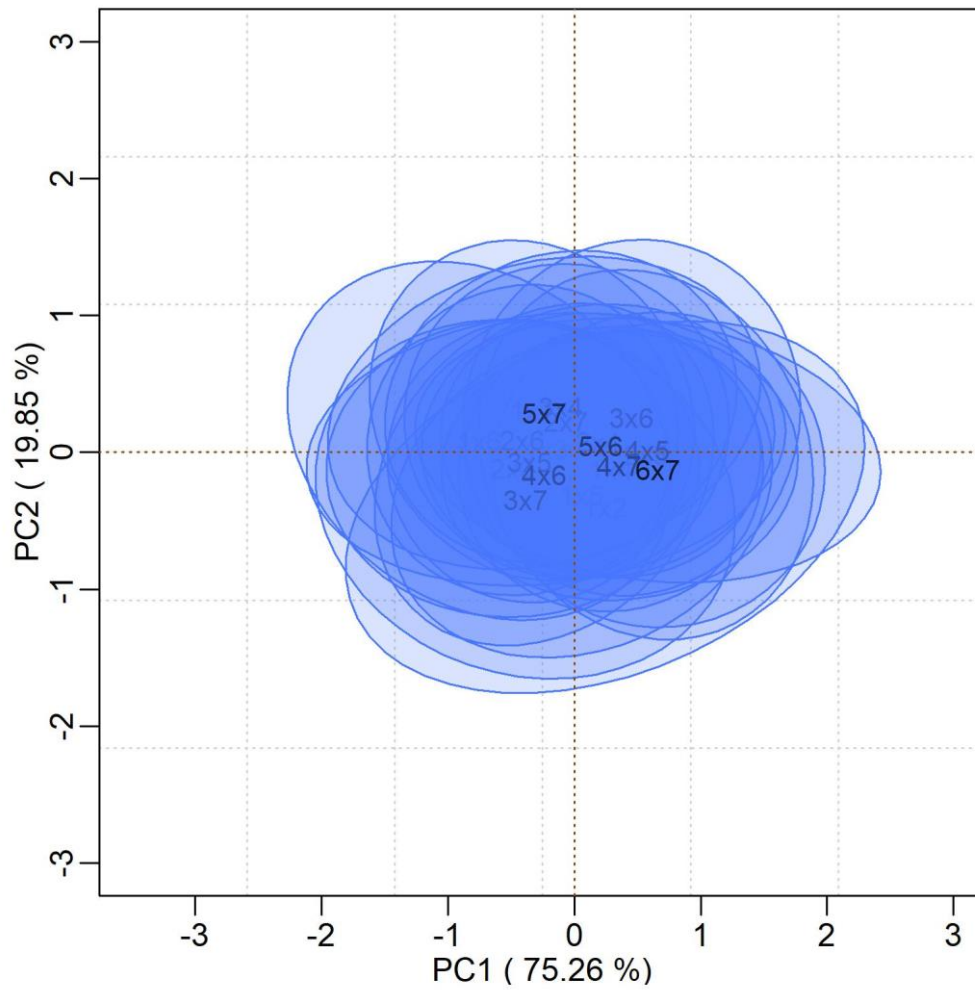


Figure 5. Maximum density posteriori intervals at level of 95% credibility for the two first principal components related to effects of SCA x environment interaction. In this biplot are only represented ellipses of hybrids.

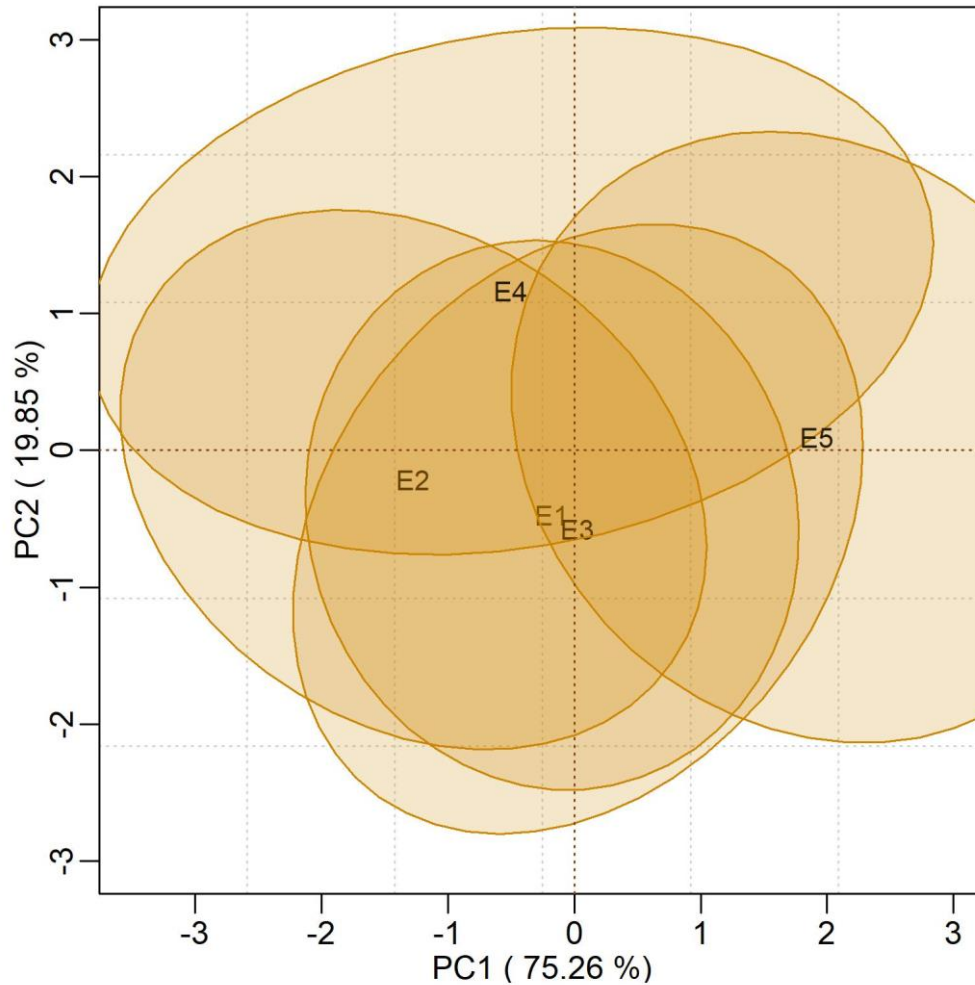


Figure 6. Maximum density posteriori intervals at level of 95% credibility for the two first principal components related to effects of SCA x environment interaction. In this biplot are only represented ellipses of environments.

Inferences about stability of genotypes and environments can only be realized when ellipses are completely inside of a unique biplot quadrant, which indicates a unique standard of interaction in all possible cases. All genotypes and environments, because of presenting zero in at least one principal component, are inconclusive in relation to the contributions for GCA x environment and SCA x environment interactions. In these cases, there can be a change of signals in coordinates and, consequently, change of interaction standard for genotypes or environments. Therefore, there is a possibility of occurring opposite interaction standards in some situations (Bernardo Júnior et al., 2018). These results reinforce the importance of being careful to take a decision in selecting and recommending genotypes. The proposed model in this study showed to be an important tool for the breeder select genotypes of high stability for GCA and SCA effects for providing higher reliability to take a decision.

Analysis of real data

As it was realized for the analysis of simulated data, the mean effects, standard deviations and credible intervals at 95% were obtained for GCA effects from 67 lines and SCA effects of 73 hybrids (Tables S11 and S12 of Supplemental Material). Figures 7 and 8 indicate intervals obtained, represented in ascending order from left to right. The credible intervals obtained indicate that all estimated GCA and SCA effects do not differ from zero, in other words, estimated means did not differ significantly from general mean. Likewise, it is not possible to select a line with greater contribution for character expression for presenting a greater gene expression of additive effects. Similarly, it is not possible to select a hybrid combination obtained from at least one parent with high GCA that presents higher SCA estimates.

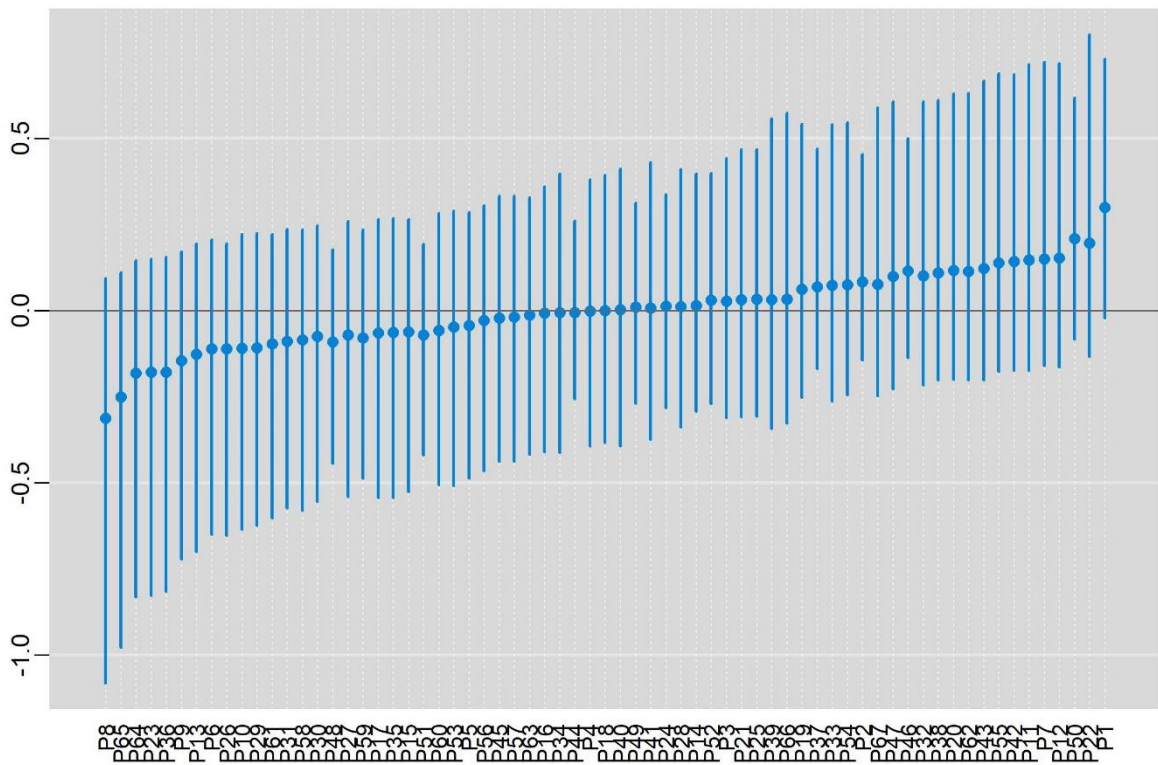


Figure 7. Maximum density posteriori intervals at level of 95% for GCA effects from 67 lines in 14 environments.

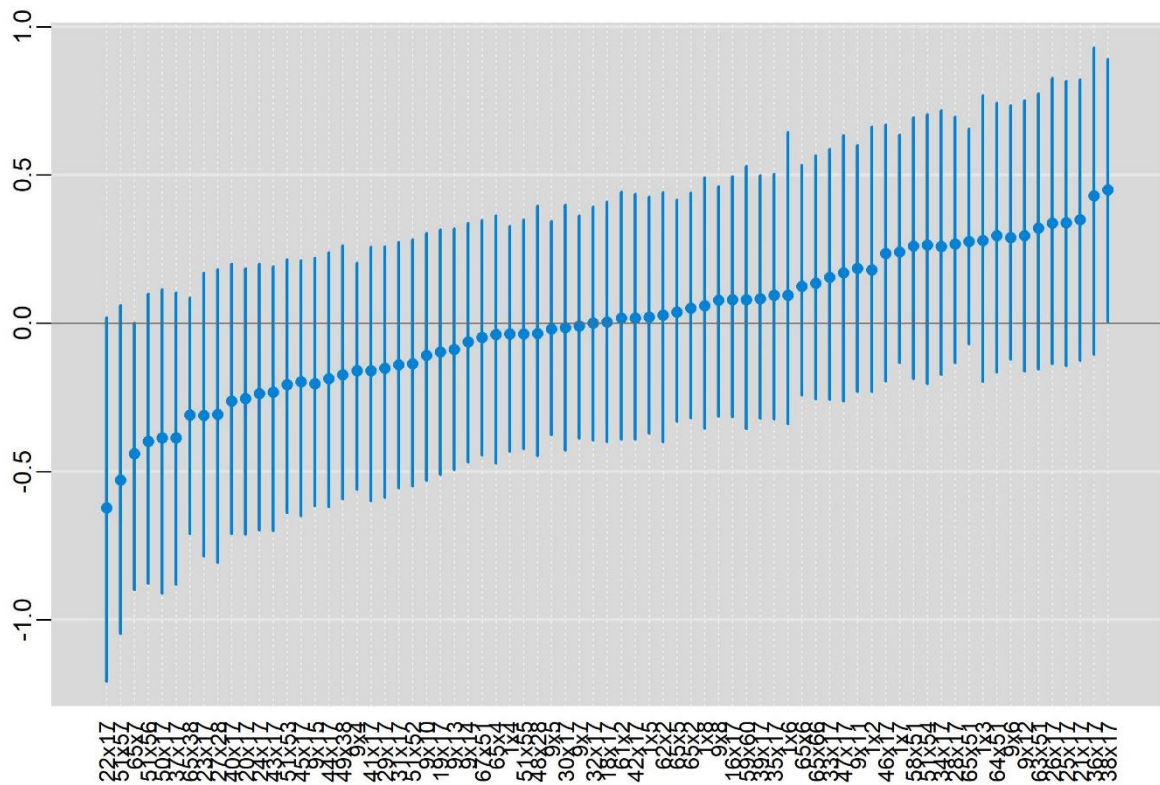


Figure 8. Maximum density posteriori intervals at level of 95% credibility for SCA effects of 73 hybrids evaluated in 14 environments.

For comparative effect, the same effects by REML methodology were estimated (Tables S13 and S14 of Supplemental Material) and obtained Pearson correlations between estimated values from both analyses. The values obtained for SCA were very close to those estimated by bayesian analysis with correlation of 0.9982. However, for GCA effects, the correlation with estimates obtained from bayesian analysis was of -0.0059. Another important aspect is in relation to the ranking of genotypes in both analyses. For SCA effects, all 15 genotypes with greater point estimates were coincident for both analyses, having few position changes between them. On the other hand, for GCA effects, only three genotypes were coincident among the 15 best for each methodology.

The variance components for GCA, SCA and estimated error by the two models showed similar values (Tables 7 and 8). Through REML methodology, variance components of GCA x environment and SCA x environment interactions were obtained. GCA heritability was of only 0.1192, while that of SCA was of 0.4102. Thus, the low Pearson correlation obtained between GCA effects estimated by REML and bayesian analyses and the great change in ranking can be justified by the low GCA heritability. Moreover, results of bayesian analysis indicate non significance for variance components of GCA and SCA.

Table 7. Estimates of variance components, standard deviation and credible interval obtained through Bayesian model for GCA, SCA and error of real data analysis

	Mean	σ	Lower Limit	Superior Limit
σ_{GCA}^2	0.0649	0.0537	0	0.1651
σ_{SCA}^2	0.1086	0.0546	0	0.1994
σ_e^2	1.0094	0.0399	0.9335	1.087

Table 8. Estimates of variance components obtained from REML model for real data analysis.

σ_{GCA}^2	0.0215
σ_{SCA}^2	0.0740
$\sigma_{GCA \times environments}^2$	0.1033
$\sigma_{SCA \times environments}^2$	0.9389
σ_e^2	0.8098

The decomposition through singular values of GCA x environment interaction permitted to capture 96.65% and 31.67% of interaction with the two first principal components for bayesian and REML analyses, respectively. From the biplot obtained through bayesian analysis (Figure 9), it is clear there is a great uncertainty around the estimates of GCA effects. All ellipses encompassed the biplot origin, and therefore, the effects of evaluated lines can not be considered unstable. It can be said that the lines do not suffer significant effects of GEI. The biplot obtained through REML method assures these results (Figure 10) because there is a big concentration of genotypic scores around origin. The ellipses of environmental scores demonstrate the same uncertainty. All environments presented ellipses containing zero in the two first principal components and do not contribute significantly to GEI, which makes the recommendation of genotypes difficult. Once more, the biplot obtained through REML analysis indicates a great concentration of environmental scores around origin.

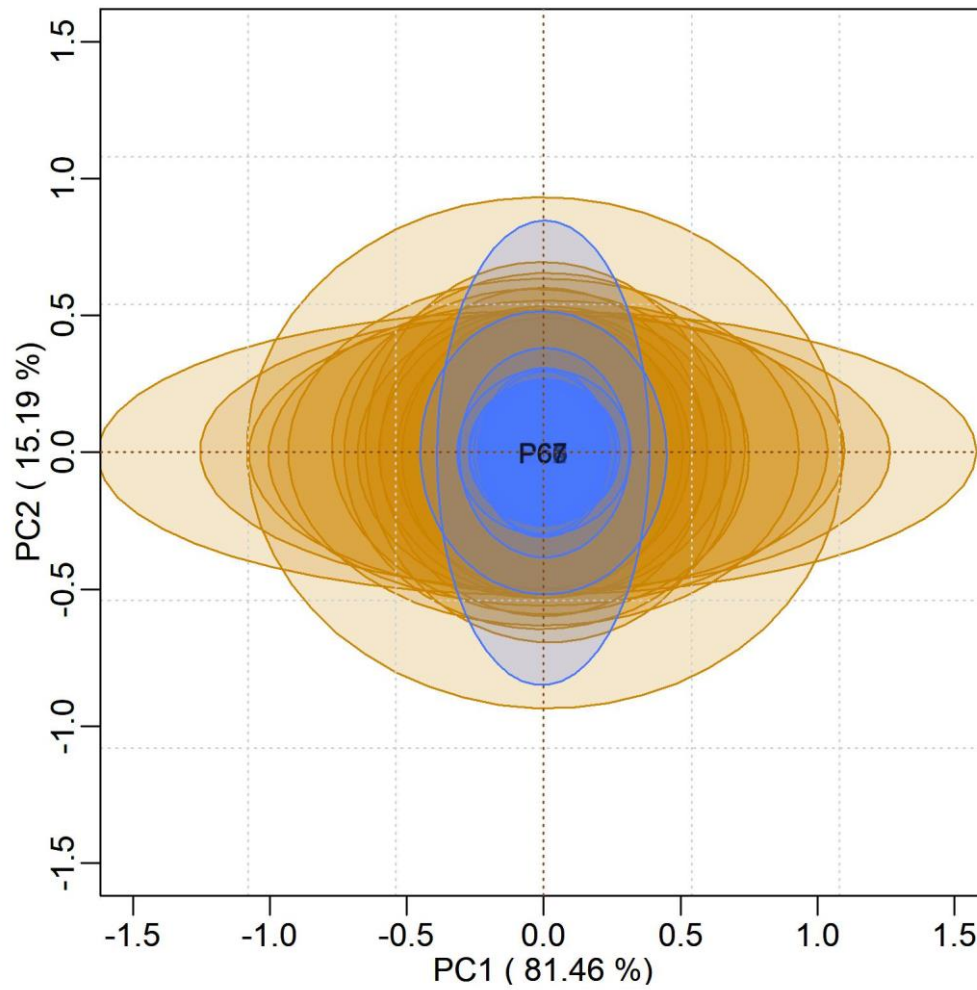


Figure 9. Maximum density posteriori intervals at 95% level of credibility for the two first principal components related to effects of GCA \times environment interaction for Bayesian analysis of real data. The blue ellipses represent the genotypes and the orange ellipses represent the environments.

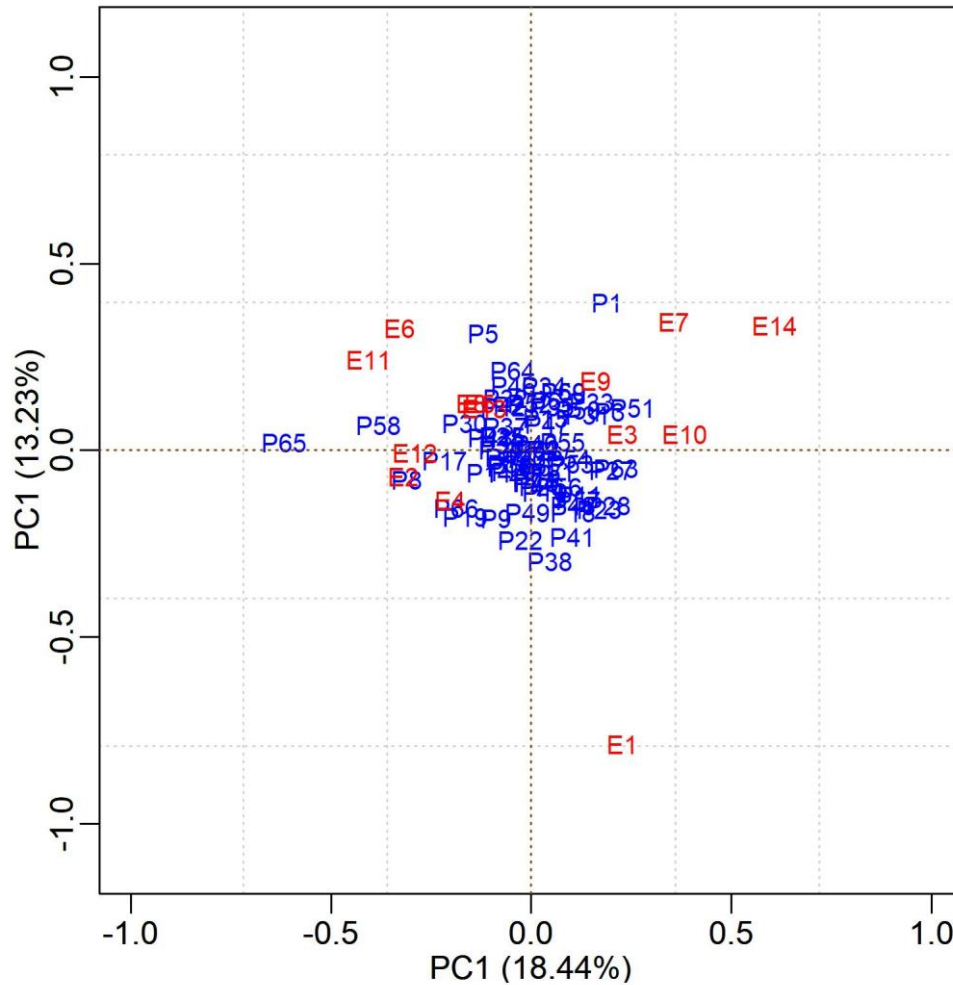


Figure 10. Scores for the two first principal components related to effects of GCA x environment interaction for REML analysis of real data. The blue scores represent the genotypes and the red scores represent the environments.

The biplots for SCA x environment interaction for both analyses are represented in Figures 11 and 12. The decomposition by singular values permitted to capture 35.14% and 30.44% of interaction with the two first principal components for bayesian and REML analyses, respectively.

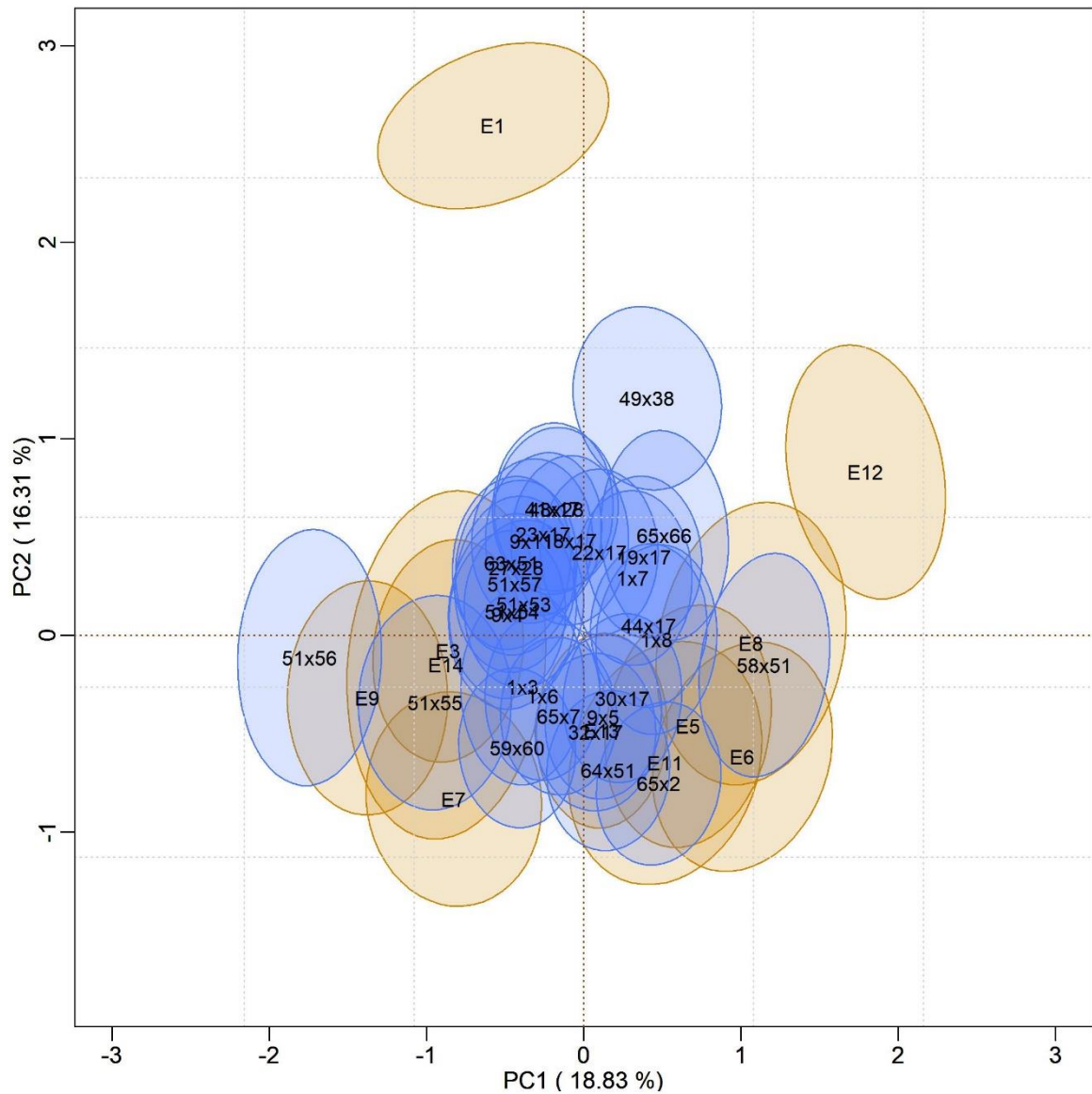


Figure 11. Maximum density posteriori intervals at level 95% of credibility for the two first principal components related to the effects of SCA x environment interaction for bayesian analysis of real data. The blue ellipses represent the genotypes and the orange ellipses represent the environments.

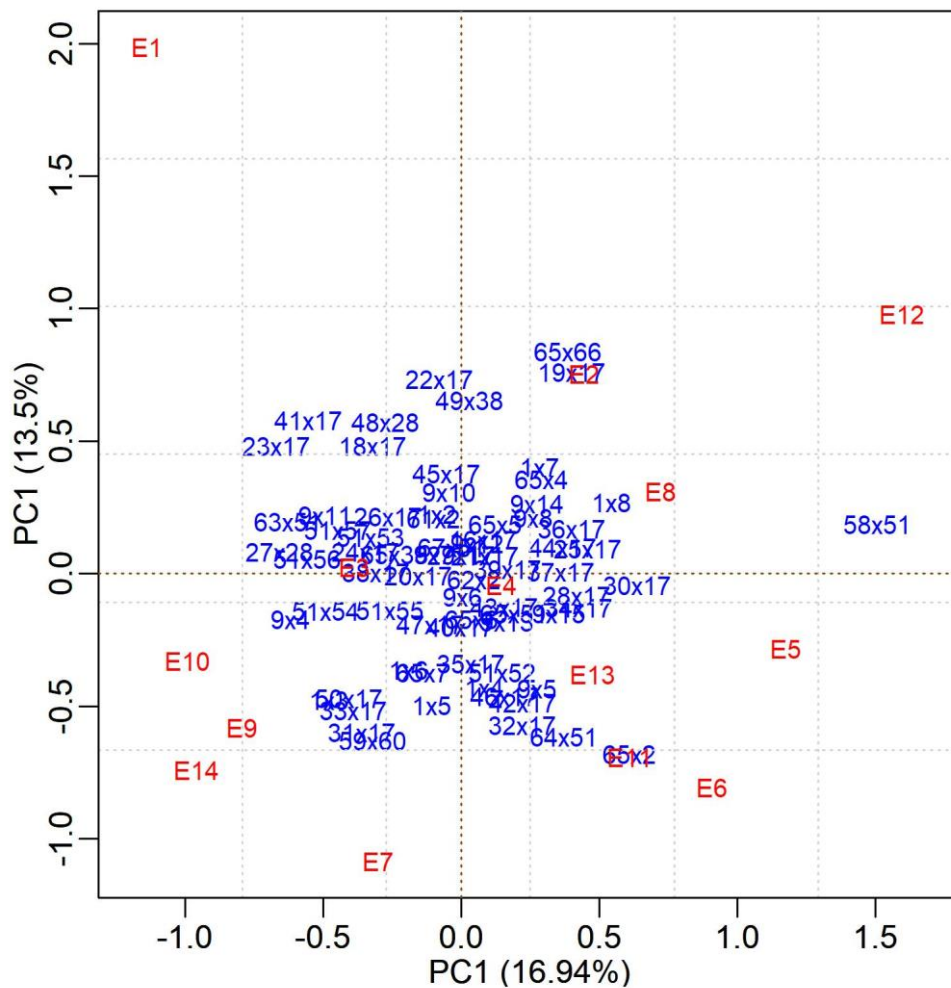


Figure 12. Scores for the two first principal components related to effects of SCA x environment interaction for REML analysis of real data. The blue scores represent the genotypes and the red scores represent the environments.

Among the 73 hybrids evaluated, 43 were not represented in the biplot due to ellipses have the origin (Figure 11). These hybrids can not be considered as unstable for not suffer significant influence of SCA x environment interaction. Likewise, the environments E2, E4 and E10 also presented the ellipses containing origin and can be considered environments with low contribution for interaction.

Figure 11 also reveals specific adaptedness of hybrids to particular environments. The hybrid 59x60 and the environment E7 show ellipses completely inside the same quadrant, and therefore, interact positively in all possible situations. The same thing can be said about hybrid 65x2 and environment E6. For the other hybrids and environments, the ellipses are present in more than a quadrant. Thus, there can be a signal change of the coordinates and, consequently,

change of interaction, as mentioned previously. Once more, these results show the importance of being careful when visually analyzing a biplot.

The biplot obtained through REML methodology (Figure 12) proves the results obtained through bayesian analysis. Almost all genotypes scores concentrate around the origin, which indicates high stability of hybrids evaluated. Moreover, it gets clear the specific adaptedness of hybrids 59x60 and 65x2 to environments E7 and E6, respectively, as indicated in Figure 11. However, the biplot constructed through bayesian approach shows all uncertainty existent around the estimates of effects for SCA x environment interaction. The use of credible ellipses is essential to decision-making be taken precisely and recommendation be right.

DISCUSSION

The statistical treatment for GEI had a great evolution over time due to the development of new statistical methods and the type of available information, such as the increase of information obtained through molecular markers and more accurate estimates about environmental effects (Lopez-Cruz et al., 2015). Acosta-Pech et al. (2017) cite several models proposed for GEI prediction (Burgueño et al., 2012; Massman et al., 2013; Heslot et al., 2014; Jarquín et al., 2014). However, none of them associates the proposed model to credible intervals and ellipses in biplots obtained through Bayesian inference. The use of biplots in multi-environmental trials is just the first step for understanding variability of interaction, and demands complementary predictive models for identifying factors related to causes of this variability (Yang et al., 2009). Therefore, the model proposed in this work is constituted of a complete method by association of diallel analysis to graphical analysis through biplots using ellipses that provide statistical inference to the breeder's decision-making. The predicted values obtained by model proposed in this study were, in general, close to the simulated values. It is clear when analyzing the signal changes and the ranking of lines and hybrids, the high correlation obtained between simulated and predicted values, and the presence of the true value in the credible intervals.

The importance of GCA, SCA effects and GCA x environment interaction vary according to the character in study; however, all these parameters represent a great part of total variance, and the inclusion of interaction effects in the model improves the predictive accuracy of non-tested hybrids (Acosta-Pech et al., 2017). Therefore, it is essential to obtain accurated predictions of each one of these parameters. In this study, the proposed model presented high predictive accuracy of combining abilities effects, singular values and singular vectors. The

Pearson correlations for simulated and predicted values had a high impact, except for singular vectors of SCA x environment interaction.

The bayesian analysis of real data highlighted the importance of credible intervals in the selection and recommendation of the best hybrids. As all GCA and SCA estimates did not differ from zero, it was not possible to select hybrid combinations of higher interest according to Cruz, Regazzi, and Carneiro (2012). This also occurred for almost all GCA and SCA effects in the analysis of the simulated data. The model proposed in this study offers higher reliability to the breeder's decision-making by avoiding hybrids of average performance be selected as superior hybrids.

GCA and SCA effects not different from zero obtained through analysis of real data can be occurred due to the no orthogonality of these effects. The model used in this work only contemplates the additive and dominance effects. Thus, there can have been a confounding of these effects with the epistatic ones. In this case, the additive effects of GCA are mistaken with additive x additive epistatic effects. The same thing occurs for dominance deviation related to SCA effects. Therefore, the decomposition of genetic effects in GCA and SCA can be a little informative (Melo et al., 2014).

This confounding is quite evident in results obtained by Melo et al. (2014). 62 models in relation to description of genetic effects were evaluated. Among them, 31 models were without information about relationship, and 31 recovered this piece of information. In general, the epistatic models were more efficient than those additive and dominant, evaluated together or separately. The results suggest that a hard confounding of additive, dominant and epistatic effects occurred in complex pedigrees. It can occur for instance, in cases of unbalance, such as in the case of real data analysis realized in this study.

Recent studies show the difficulty with predicting the behavior of environments in different years in tropical regions (de Figueiredo et al., 2015; Bernardo Júnior et al., 2018). The same local can provide completely different conditions from one year to another. Even the hybrids with the best performance are influenced by this environmental oscillation and can present different results in the same local in different years. The estimates obtained for an only year can represent one of the many possible situations due to environmental variations that cause fluctuations in genotypic and environmental scores, and conceal the standard graphically obtained (Yang et al., 2009). Likewise, the breeder should not only realize crossings between lines and select the hybrids with the highest effects of GCA and SCA but also take into consideration the stability of these effects in order to have greater reliability of hybrid performance to a specific environment. This is clear when analyzing biplots obtained in this

study. The ellipses demonstrate great uncertainty around estimates of effects for GCA x environment and SCA x environment interactions, which can lead to a change of interaction standard in biplot and completely different inferences from those obtained from visual evaluation. For the analysis of real data, these results are more evident when comparing with biplots obtained through REML methodology. The point estimates do not indicate the uncertainty existent around genotypic and environmental scores, which can lead to completely mistaken inferences. In addition, the ellipses show how the unbalance affects the uncertainty of the predictions, indicated by the GEI effects not significantly different from zero.

The ellipses built into the biplot provide information about which environments provide the same information and which ones can be discarded (Bernardo Júnior et al., 2018), constituting an interesting tool to complement the results obtained by prediction models. The use of biplots is an excellent alternative for assessing the interaction in studies of combining ability, and the use of credible ellipses permit to realize more precise and less subjective inferences (Yang et al., 2009; Crossa et al., 2011; de Oliveira et al., 2015; Bernardo Júnior et al., 2018). This allows a better evaluation about stability of GCA and SCA effects. In the simulation realized in this work, almost all environments can provide the same information due to overlaps of ellipses in biplot. The E1 and E4 environments presented low contribution for GCA x environment interaction due to their ellipses included the origin of the biplot. The environments E2, E3 and E5 did not encompassed zero value in at least one principal component, which difficult infer about their contribution to GEI. Similarly, for SCA x environment interaction, all environments did not present significant contribution for interaction because of ellipses contain origin [0,0]. The same inferences can be done for the analysis of real data. For GCA x environment interaction, all environments covered the zero in the two first principal components. For SCA x environment interaction, the environments E2, E4 and E10 can provide the same information because all of them presented ellipses that encompass the origin, and can be considered as low contribution for the interaction. Moreover, most part of the environments present overlaps of the ellipses. On the other hand, the environments E6, E7 and E12 showed ellipses completely inside a unique quadrant. This is the ideal scenario for realizing the recommendation of genotypes because there is no chance of occurring change in the interaction standard of these environments.

The advantages of realizing analyses under Bayesian approach are widely discussed by Viele and Srinivasan (2000), Crossa et al. (2011), de Oliveira et al. (2015) and Bernardo Júnior et al. (2018). Among these advantages, it is possible to cite the ease to analyze sets of unbalanced data and deal with heterogeneity of variances; produce estimates of the main effects

and GEI with shrinkage effect for uninformative priors; possibility of including information a priori to the analysis, such as historical of experiments and information about relationship; it does not demand determination of which effects are fixed and randomized, and provides a natural way to obtain credible regions from genotypic and environmental scores in the biplot. Although this study has not simulated all situations described, all advantages are applied for the model proposed. New studies can be carried out in order to evaluate the efficiency of the model under these circumstances.

The proposed method in this study shows to be efficient for prediction of GCA and SCA in multi-environmental trials, and represents promising tool for breeder's decision-making to select parents for synthesis of high stability hybrids for combining ability effects. The proposed method in this study shows to be efficient for prediction of GCA and SCA. Through the analysis of simulated data set, it is possible to evaluate the efficiency of the model to predict the effects of GCA and SCA through direct comparison between the true value of parameters and predictions obtained. Besides, the results obtained by REML methodology proved the efficiency of Bayesian method proposed and the applicability in a set of real data. Nevertheless, new studies are still necessary to evaluate the behavior of the model in scenarios of inclusion of epistasis effects.

ACKNOWLEDGEMENT

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

REFERENCES

- Abera, W., Hussein, S., Derera, J., Worku, M., & Laing, M. (2016). Heterosis and combining ability of elite maize inbred lines under northern corn leaf blight disease prone environments of the mid-altitude tropics. *Euphytica*, 208(2), 391-400.
- Acosta-Pech, R., Crossa, J., de los Campos, G., Teyssèdre, S., Claustres, B., Pérez-Elizalde, S., Pérez-Rodríguez, P. (2017). Genomic models with genotype x environment interaction for predicting hybrid performance: an application in maize hybrids. *Theoretical and Applied Genetics*, 130(7), 1431-1440.
- Adebayo, M. A., Menkir, A., Blay, E., Gracen, V., & Danquah, E. (2017). Combining ability and heterosis of elite drought-tolerant maize inbred lines evaluated in diverse environments of lowland tropics. *Euphytica*, 213(2), 43.

- Amegbor, I. K., Badu-Apraku, B., & Annor, B. (2017). Combining ability and heterotic patterns of extra-early maturing white maize inbreds with genes from *Zea diploperennis* under multiple environments. *Euphytica*, 213(1), 24.
- Balestre, M., Machado, J. C., Lima, J. L., Souza, J. C., & Nóbrega Filho, L. (2008). Genetic distance estimates among single cross hybrids and correlation with specific combining ability and yield in corn double cross hybrids. *Genetics and Molecular Research*, 7(1), 65-73.
- Beck, D. L., Vasal, S. K., & Crossa, J. (1990). Heterosis and combining ability of CIMMYT's tropical early and intermediate maturity maize (*Zea mays* L.) germplasm. *Maydica*, 35(3), 279-285.
- Bernardo Júnior, L. A. Y., da Silva, C. P., de Oliveira, L. A., Nuvunga, J. J., Pires, L. P. M., Von Pinho, R. G., & Balestre, M. (2018). AMMI bayesian models to study stability and adaptability in maize. *Agronomy Journal*, 110(5), 1765-1776.
- Bradu, D., & Gabriel, K. R. (1978). The biplot as a diagnostic tool for models of two-way tables. *Technometrics*, 20(1), 47-68.
- Burgueño, J., de los Campos, G., Weigel, K., & Crossa, J. (2012). Genomic prediction of breeding values when modeling genotype \times environment interaction using pedigree and dense molecular markers. *Crop Science*, 52(2), 707-719.
- Camargos, R. B., Von Pinho, R. G., Balestre, M., Ferreira, R. A. D. C., Dias, K. O. G., Pires, L. P. M., & Souza, V. F. (2017). Efficiency of selection per se and in single-cross hybrids for disease resistance in maize. *Genetics and Molecular Research*, 16(3).
- Chen, M. H., & Shao, Q. M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, 8(1), 69-92.
- Crossa, J., Perez-Elizalde, S., Jarquin, D., Cotes, J. M., Viele, K., Liu, G., & Cornelius, P. L. (2011). Bayesian estimation of the additive main effects and multiplicative interaction model. *Crop Science*, 51(4), 1458-1469.
- Cruz, C. D., Regazzi, A. J., & Carneiro, P. C. S. (2012). *Modelos biométricos aplicados ao melhoramento genético*. Viçosa: Ed. UFV.
- da Silva, C. P., de Oliveira, L. A., Nuvunga, J. J., Pamplona, A. K. A., & Balestre, M. (2015). A Bayesian shrinkage approach for AMMI models. *PloS one*, 10(7), 1-27.
- de Figueiredo, A. G., Von Pinho, R. G., Silva, H. D., & Balestre, M. (2015). Application of mixed models for evaluating stability and adaptability of maize using unbalanced data. *Euphytica*, 202(3), 393-409.
- de Oliveira, L. A., da Silva, C. P., Nuvunga, J. J., da Silva, A. Q., & Balestre, M. (2015). Credible intervals for scores in the AMMI with random effects for genotype. *Crop Science*, 55(2), 465-476.
- Derera, J., Tongoona, P., Vivek, B. S., & Laing, M. D. (2008). Gene action controlling grain yield and secondary traits in southern African maize hybrids under drought and non-drought environments. *Euphytica*, 162(3), 411-422.
- Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58(3), 453-467.

- Giraud, H., Bauland, C., Falque, M., Madur, D., Combes, V., Jamin, P., ... & Moreau, L. (2017). Reciprocal genetics: Identifying QTL for general and specific combining abilities in hybrids between multiparental populations from two maize (*Zea mays* L.) heterotic groups. *Genetics*, 207(3), 1167-1180.
- Griffing, B. (1956). A generalised treatment of the use of diallel crosses in quantitative inheritance. *Heredity*, 10(1), 31.
- Guedes, F. L. (2012). *Desempenho de híbridos de milho a partir de progênies contrastantes em relação ao uso de nitrogênio* (Doctoral thesis). Retrieved from Federal University of Lavras Digital Repository <http://repositorio.ufla.br/jspui/handle/1/374>
- Heidelberger, P., & Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, 31(6), 1109-1144.
- Heslot, N., Akdemir, D., Sorrells, M. E., & Jannink, J. L. (2014). Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theoretical and applied genetics*, 127(2), 463-480.
- Holland, J. B., Nyquist, W. E., & Cervantes-Martínez, C. T. (2003). Estimating and interpreting heritability for plant breeding: an update. *Plant breeding reviews*, 22.
- Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., ... & de los Campos, G. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and applied genetics*, 127(3), 595-607.
- Jumbo, M. B., & Carena, M. J. (2008). Combining ability, maternal, and reciprocal effects of elite early-maturing maize population hybrids. *Euphytica*, 162(3), 325-333.
- Kadam, D. C., Potts, S. M., Bohn, M. O., Lipka, A. E., & Lorenz, A. J. (2016). Genomic prediction of single crosses in the early stages of a maize hybrid breeding pipeline. *G3: Genes, Genomes, Genetics*, 6(11), 3443-3453.
- Kulka, V. P., da Silva, T. A., Contreras-Soto, R. I., Maldonado, C., Mora, F., & Scapim, C. A. (2018). Diallel analysis and genetic differentiation of tropical and temperate maize inbred lines. *Crop Breeding and Applied Biotechnology*, 18(1), 31-38.
- Lopez-Cruz, M., Crossa, J., Bonnett, D., Dreisigacker, S., Poland, J., Jannink, J. L., ... & de los Campos, G. (2015). Increased prediction accuracy in wheat breeding trials using a marker \times environment interaction genomic selection model. *G3: Genes, Genomes, Genetics*, 5(4), 569-582.
- Massman, J. M., Gordillo, A., Lorenzana, R. E., & Bernardo, R. (2013). Genomewide predictions from maize single-cross data. *Theoretical and applied genetics*, 126(1), 13-22.
- Matzinger, D. F., Sprague, G. F., & Cockerham, C. C. (1959). Diallel crosses of maize in experiments repeated over locations and years. *Agronomy Journal*, 51(3), 346-350.
- Melo, W. M. C., Balestre, M., Von Pinho, R. G., & Bueno Filho, J. S. S. (2014). Genetic control of the performance of maize hybrids using complex pedigrees and microsatellite markers. *Euphytica*, 195(3), 331-344.
- Nass, L. L., Lima, M., Vencovsky, R., & Gallo, P. B. (2000). Combining ability of maize inbred lines evaluated in three environments in Brazil. *Scientia Agricola*, 57(1), 129-134.

- Pires, L. P. M. (2017). *Predição genômica do desempenho de híbridos de milho considerando a interação genótipos por ambientes* (Doctoral thesis). Retrieved from Federal University of Lavras Digital Repository <http://repositorio.ufla.br/jspui/handle/1/29463>
- R Development Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2018.
- Raftery, A. E., Lewis, S. M (1992). Comment: One long run with diagnostics: Implementation strategies for Markov Chain Monte Carlo. *Statistical Science*, 7:493–497.
- Rojas, B. A.; Sprague, G.F. (1952). A comparison of variance components in corn yield trials: III. General and specific combining ability and their interaction with locations and years. *Agronomy Journal*, 44(9), 462.
- SAS Institute Inc. (2002). *SAS/STAT® software: version 9*.
- Schrag, T. A., Möhring, J., Maurer, H. P., Dhillon, B. S., Melchinger, A. E., Piepho, H. P., ... & Frisch, M. (2009). Molecular marker-based prediction of hybrid performance in maize using unbalanced data from multiple experiments with factorial crosses. *Theoretical and applied genetics*, 118(4), 741-751.
- Schmidt, P., Hartung, J., Rath, J., & Piepho, H. P. (2019). Estimating Broad-Sense Heritability with Unbalanced Data from Agricultural Cultivar Trials. *Crop Science*, 59(2), 525-536.
- Smith, B. J. (2007). boa: an R package for MCMC output convergence assessment and posterior inference. *Journal of Statistical Software*, 21(11), 1-37.
- 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(1-2), 201-219.
- Yang, R. C., Crossa, J., Cornelius, P. L., & Burgueño, J. (2009). Biplot analysis of genotype × environment interaction: Proceed with caution. *Crop Science*, 49(5), 1564-1576.
- Zobel, R. W., Wright, M. J., & Gauch, H. G. (1988). Statistical analysis of a yield trial. *Agronomy journal*, 80(3), 388-393.

SUPPLEMENTAL MATERIAL

Table S1. Mean effects, standard deviation and credible intervals for 7 lines in 5 environments.

Line	Simulated Effect	Mean Effect	σ	Lower Limit	Upper Limit
P1	2.9890	4.8569	2.3421	0.1093	9.4519
P4	0.3814	0.8335	2.1590	-3.5554	4.9900
P2	1.0281	0.6149	2.1777	-3.9039	4.8578
P3	0.3973	-0.1070	2.1738	-4.7285	3.8003
P5	0.2280	-0.4654	2.1830	-5.0506	3.5353
P6	-0.5290	-0.6116	2.1917	-5.0101	3.4474
P7	-3.2706	-5.1938	2.3548	-10.1465	-0.7640

Table S2. Mean effects, standard deviation and credible intervals for 21 simulated hybrids in complete diallel with 7 parents in 5 environments.

Hybrid	Simulated Effect	Mean Effect	σ	Lower Limit	Upper Limit
1x3	5.1155	4.0183	1.9781	0.4705	8.3301
1x2	5.8844	3.7941	2.0287	0.0895	8.3486
6x7	0.0254	2.6135	1.9071	-1.6704	6.1037
1x5	4.0510	1.9095	1.8872	-1.6094	6.0327
1x4	4.4895	1.2467	1.9582	-2.2293	5.7132
2x4	0.7016	1.1509	1.7015	-2.1099	4.4633
3x6	-1.7255	0.9966	1.6881	-2.3218	4.4396
5x7	-1.3314	0.7561	1.9486	-3.5960	4.2689
2x6	-0.0144	0.5003	1.6682	-2.7159	3.8013
3x5	0.3889	0.4410	1.7244	-2.9756	3.8775
4x5	-0.2488	0.0721	1.6915	-3.2258	3.4551
3x4	-0.1719	-0.0049	1.7180	-3.3445	3.3947
4x6	0.7706	-0.1332	1.6942	-3.4749	3.3080
2x7	-2.9836	-0.2485	1.8853	-4.4473	3.1260
4x7	-2.7821	-1.3376	1.8540	-5.3957	2.0251
2x5	-3.0445	-1.5345	1.7341	-4.9793	1.8878
5x6	-2.7349	-1.7913	1.7084	-5.2305	1.5333
3x7	-4.1325	-2.0526	1.9636	-6.5109	1.4120
1x6	-1.5781	-2.6414	1.8641	-6.1819	1.3468
2x3	-4.0087	-3.0982	1.7622	-6.5141	0.3961
1x7	-4.7071	-4.0941	1.7355	-7.4779	-0.6528

Table S3. Simulated and predicted values of singular vectors from genotypic effects of GCA x environment interaction for the first principal component.

Line	Simulated Value	Corrected Mean ¹	Median ²	σ	Lower Limit	Upper Limit
1	-0.2688	-0.609	-0.5933	0.2282	-0.7059	0.2417
2	-0.3139	-0.0902	-0.0839	0.1993	-0.6416	0.4503
3	-0.5109	-0.4157	-0.3978	0.1763	-0.5438	0.1901
4	0.5553	0.5555	0.5423	0.2239	-0.3154	0.6809
5	0.4863	0.3283	0.3189	0.1827	-0.3217	0.4962
6	-0.0791	0.0652	0.0603	0.1410	-0.2443	0.3407
7	0.1311	0.1658	0.1556	0.1436	-0.2416	0.3791

¹ Correlation between real and predicted values: 0.8891² Correlation between real and predicted values: 0.8875

Table S4. Simulated and predicted values of singular vectors from genotypic effects of GCA x environment interaction for the second principal component.

Line	Simulated Value	Corrected Mean ¹	Median ²	σ	Lower Limit	Upper Limit
1	-0.1795	-0.2114	-0.1847	0.2096	-0.5167	0.4040
2	0.7475	0.7475	0.6826	0.2822	-0.2966	0.8505
3	-0.2939	-0.1425	-0.1250	0.2137	-0.4784	0.4314
4	-0.2762	-0.3568	-0.3345	0.2325	-0.5994	0.4143
5	0.4057	0.3941	0.3545	0.2391	-0.4124	0.6293
6	-0.2150	-0.1961	-0.1732	0.2204	-0.5374	0.4032
7	-0.1887	-0.2348	-0.2091	0.2149	-0.5479	0.4024

¹ Correlation between real and predicted values: 0.9835² Correlation between real and predicted values: 0.9820

Table S5. Simulated and predicted values of singular vectors from environmental effects of GCA x environment interaction for the first principal component.

Line	Simulated Value	Corrected Mean ¹	Median ²	σ	Lower Limit	Upper Limit
1	-0.2311	-0.0280	-0.1620	0.1954	-0.5531	0.4865
2	-0.5044	-0.6227	-0.5103	0.2102	-0.6314	0.2785
3	-0.3227	-0.1234	-0.3482	0.2126	-0.7652	0.3810
4	0.4103	0.0019	0.3639	0.1976	-0.3031	0.5211
5	0.6478	0.7722	0.6486	0.1432	0.1220	0.7463

¹ Correlation between real and predicted values: 0.8614² Correlation between real and predicted values: 0.9963

Table S6. Simulated and predicted values of singular vectors from environmental effects of GCA x environment interaction for the second principal component.

Line	Simulated Value	Corrected Mean ¹	Median ²	σ	Lower Limit	Upper Limit
1	-0.5303	-0.6793	-0.5359	0.2902	-0.7845	0.5088
2	-0.2298	-0.2000	-0.2662	0.2312	-0.6560	0.4027
3	0.7912	0.5719	0.7332	0.1903	0.0877	0.8368
4	-0.1560	0.4029	0.0729	0.2511	-0.6247	0.4998
5	0.1248	-0.0955	-0.0088	0.2171	-0.5835	0.4672

¹ Correlation between real and predicted values: 0.7839² Correlation between real and predicted values: 0.9619

Table S7. Simulated and predicted values of singular vectors from genotypic effects of SCA x environment interaction for the first principal component.

Line	Simulated Value	Corrected Mean ¹	Median ²	σ	Lower Limit	Upper Limit
1	-0.2530	0.1685	0.1394	0.2391	-0.3408	0.4958
2	-0.3307	-0.0764	-0.0255	0.2023	-0.4135	0.3193
3	-0.2267	0.1725	0.0959	0.1686	-0.2671	0.4052
4	-0.4132	0.0458	0.0370	0.1940	-0.3190	0.4011
5	-0.3071	-0.4069	-0.2536	0.1810	-0.5088	0.2136
6	-0.1880	-0.1677	-0.1108	0.1800	-0.4337	0.2998
7	0.2800	-0.2608	-0.1528	0.1733	-0.4239	0.2457
8	0.2995	0.0831	0.0506	0.1693	-0.2994	0.3318
9	0.1657	0.3138	0.1829	0.1692	-0.2211	0.4622
10	0.2451	-0.2415	-0.1308	0.1731	-0.4536	0.2575
11	0.2461	-0.0395	-0.0171	0.1783	-0.3785	0.3361
12	0.2714	-0.0625	-0.0441	0.1884	-0.4036	0.3313
13	0.0929	-0.2106	-0.1150	0.1586	-0.3994	0.2366
14	0.2113	0.2926	0.1420	0.1886	-0.2434	0.5131
15	-0.1619	-0.2163	-0.1241	0.2014	-0.4407	0.3314
16	0.0065	0.2898	0.1573	0.1892	-0.2521	0.4659
17	0.0642	-0.1746	-0.0894	0.1819	-0.4395	0.2594
18	0.0271	0.1707	0.1081	0.1679	-0.2634	0.3777
19	0.0059	0.1043	0.0623	0.1600	-0.2869	0.3334
20	-0.0198	-0.1647	-0.1228	0.2133	-0.4619	0.3030
21	-0.0151	0.3804	0.2158	0.1795	-0.2180	0.5122

¹ Correlation between real and predicted values: 0.0698

² Correlation between real and predicted values: 0.045

Table S8. Simulated and predicted values of singular vectors from genotypic effects of SCA x environment interaction for the second principal component.

Line	Simulated Value	Corrected Mean ¹	Median ²	σ	Lower Limit	Upper Limit
1	0.2197	-0.3902	-0.1577	0.2195	-0.5239	0.3041
2	0.2646	0.0359	0.0120	0.1908	-0.3495	0.3990
3	0.2076	-0.1035	-0.0474	0.1923	-0.4005	0.3491
4	-0.0017	-0.3221	-0.1321	0.2068	-0.4863	0.3333
5	-0.5487	0.0345	0.0215	0.2170	-0.3859	0.4326
6	-0.3516	0.2971	0.1276	0.2115	-0.3365	0.4820
7	0.0851	-0.1461	-0.0507	0.1982	-0.4235	0.3508
8	-0.0044	0.0993	0.0315	0.1884	-0.3458	0.4040
9	0.3995	0.0405	0.0146	0.2206	-0.4530	0.4186
10	-0.2303	0.0772	0.0400	0.2028	-0.3623	0.4222
11	-0.0493	0.2463	0.1014	0.2032	-0.3277	0.4596
12	-0.0518	0.3498	0.1406	0.1981	-0.3054	0.4714
13	-0.1554	-0.1092	-0.0308	0.1923	-0.4057	0.3580
14	-0.1442	0.3020	0.1063	0.2277	-0.3713	0.5128
15	-0.0354	-0.4096	-0.1531	0.2225	-0.5272	0.3021
16	0.1015	0.0436	0.0077	0.1861	-0.3623	0.3729
17	-0.1556	-0.2091	-0.0744	0.2175	-0.4744	0.3585
18	0.0428	-0.0926	-0.0415	0.1866	-0.3911	0.3315
19	0.1056	0.0721	0.0223	0.1868	-0.3473	0.3816
20	-0.0063	0.2837	0.1153	0.2004	-0.3123	0.4644
21	0.3085	-0.0999	-0.0494	0.2037	-0.4340	0.3567

¹Correlation between real and predicted values: -0.2694²Correlation between real and predicted values: -0.3189

Table S9. Simulated and predicted values of singular vectors from environmental effects of SCA x environment interaction for the first principal component.

Line	Simulated Value	Corrected Mean ¹	Median ²	σ	Lower Limit	Upper Limit
1	-0.3292	-0.0639	-0.0566	0.2752	-0.6120	0.5121
2	-0.2809	-0.5267	-0.4470	0.2706	-0.7374	0.3878
3	-0.4681	0.0279	0.0572	0.2987	-0.6007	0.5606
4	0.6175	-0.2475	-0.2916	0.4214	-0.8188	0.6604
5	0.4607	0.8102	0.6599	0.2099	0.0637	0.8598

¹Correlation between real and predicted values: 0.3764²Correlation between real and predicted values: 0.2834

Table S10. Simulated and predicted values of singular vectors from environmental effects of SCA x environment interaction for the second principal component.

Line	Simulated Value	Corrected Mean ¹	Median ²	σ	Lower Limit	Upper Limit
1	0.2012	-0.3398	-0.2613	0.3766	-0.7956	0.6447
2	-0.8344	-0.2251	-0.1489	0.3671	-0.7620	0.5775
3	0.5051	-0.3620	-0.2740	0.3838	-0.8182	0.6697
4	0.0588	0.8331	0.5422	0.2428	0.0369	0.8526
5	0.0694	0.0938	0.0601	0.3945	-0.7530	0.7128

¹Correlation between real and predicted values: -0.0079²Correlation between real and predicted values: -0.0451

Table S11. Mean effects, standard deviation and credible interval for 67 lines in 14 environments obtained from Bayesian model.

Line	Mean Effect	σ	Lower Limit	Upper Limit
P1	0.3061	0.2095	-0.0416	0.6871
P22	0.2433	0.2584	-0.1452	0.7857
P50	0.2240	0.1939	-0.1027	0.5950
P12	0.1972	0.2349	-0.1989	0.6770
P7	0.1962	0.2343	-0.2172	0.6547
P11	0.1917	0.2352	-0.1906	0.6916
P42	0.1820	0.2268	-0.1824	0.6717
P55	0.1787	0.2275	-0.1888	0.6718
P43	0.1633	0.2230	-0.2110	0.6448
P62	0.1552	0.2145	-0.2272	0.6040
P20	0.1536	0.2146	-0.2110	0.6169
P38	0.1480	0.2123	-0.2037	0.6038
P32	0.1377	0.2072	-0.2427	0.5754
P46	0.1368	0.1671	-0.1395	0.4943
P47	0.1350	0.2117	-0.2401	0.5914
P67	0.1140	0.2108	-0.2539	0.5785
P2	0.1094	0.1559	-0.1381	0.4585
P54	0.1081	0.1986	-0.2544	0.5325
P33	0.1032	0.2014	-0.2593	0.5431
P37	0.0976	0.1620	-0.1876	0.4435
P19	0.0945	0.1991	-0.2660	0.5252
P66	0.0718	0.2204	-0.3322	0.5606
P39	0.0609	0.2174	-0.3582	0.5296
P25	0.0572	0.1870	-0.3249	0.4405
P21	0.0542	0.1879	-0.3196	0.4534
P3	0.0474	0.1845	-0.3214	0.4323
P52	0.0471	0.1629	-0.2472	0.4176
P14	0.0321	0.1677	-0.3054	0.3769
P28	0.0231	0.1835	-0.3396	0.4115
P24	0.0213	0.1506	-0.2888	0.3277
P41	0.0186	0.1908	-0.3750	0.4263
P49	0.0156	0.1439	-0.2691	0.3124
P40	0.0072	0.1922	-0.3864	0.4169
P18	0.0010	0.1854	-0.3671	0.4045
P4	-0.0029	0.1855	-0.3768	0.3929
P44	-0.0050	0.1262	-0.2640	0.2528
P34	-0.0083	0.1954	-0.4135	0.3941
P16	-0.0125	0.1856	-0.4046	0.3650
P63	-0.0272	0.1809	-0.4134	0.3317
P57	-0.0344	0.1854	-0.4396	0.3299
P45	-0.0366	0.1856	-0.4297	0.3379
P56	-0.0528	0.1874	-0.4500	0.3167
P5	-0.0638	0.1886	-0.4870	0.2856
P53	-0.0752	0.1944	-0.4960	0.3004
P60	-0.0829	0.1952	-0.5030	0.2841

P51	-0.0882	0.1518	-0.4047	0.2017
P15	-0.0924	0.1952	-0.5322	0.2555
P35	-0.0938	0.1983	-0.5122	0.2895
P17	-0.0956	0.1968	-0.5272	0.2727
P59	-0.0979	0.1762	-0.4846	0.2364
P27	-0.1016	0.1998	-0.5346	0.2634
P48	-0.1068	0.1580	-0.4251	0.1884
P30	-0.1069	0.2015	-0.5505	0.2495
P58	-0.1189	0.2047	-0.5780	0.2365
P31	-0.1214	0.2037	-0.5587	0.2460
P61	-0.1319	0.2098	-0.5802	0.2379
P29	-0.1447	0.2147	-0.6277	0.2177
P10	-0.1490	0.2164	-0.5889	0.2551
P26	-0.1535	0.2207	-0.6273	0.2162
P6	-0.1545	0.2211	-0.6337	0.2112
P13	-0.1701	0.2311	-0.6830	0.2051
P9	-0.1878	0.2358	-0.6901	0.1946
P36	-0.2307	0.2573	-0.7586	0.1863
P23	-0.2316	0.2596	-0.7775	0.1754
P64	-0.2358	0.2620	-0.8004	0.1606
P65	-0.3067	0.3032	-0.9011	0.1754
P8	-0.3646	0.3357	-1.0121	0.1449

Table S12. Mean effects, standard deviation and interval of credibility for 73 hybrids in 14 environments obtained from Bayesian model.

Hybrid	Mean Effect	σ	Lower Limit	Upper Limit
38x17	0.4496	0.2272	-0.0110	0.8596
36x17	0.4154	0.2774	-0.1087	0.9201
21x17	0.3435	0.2524	-0.1317	0.8100
25x17	0.3355	0.2527	-0.1710	0.7775
26x17	0.3349	0.2582	-0.1328	0.8290
63x51	0.3136	0.2461	-0.1606	0.7667
9x12	0.2927	0.2426	-0.1736	0.7389
9x6	0.2911	0.2277	-0.1154	0.7384
64x51	0.2877	0.2408	-0.1616	0.7447
1x3	0.2786	0.2571	-0.2003	0.7645
65x51	0.2772	0.1882	-0.0709	0.6512
28x17	0.2713	0.2136	-0.1317	0.6967
34x17	0.2593	0.2336	-0.1786	0.7131
51x54	0.2574	0.2338	-0.1769	0.7271
58x51	0.2573	0.2295	-0.1816	0.6975
1x7	0.2408	0.1997	-0.1342	0.6306
46x17	0.2362	0.2252	-0.1955	0.6686
1x2	0.1941	0.2372	-0.2295	0.6632
9x11	0.1834	0.2140	-0.2443	0.5863
47x17	0.1748	0.2248	-0.2778	0.6151
33x17	0.1598	0.2187	-0.2633	0.5804
65x66	0.1380	0.2090	-0.2689	0.5499
65x6	0.1289	0.1966	-0.2593	0.5131
1x6	0.1165	0.2555	-0.3322	0.6508
35x17	0.0985	0.2090	-0.2975	0.5257
39x17	0.0882	0.2089	-0.3221	0.4968
59x60	0.0862	0.2221	-0.3462	0.5378
16x17	0.0848	0.2056	-0.3429	0.4641
9x8	0.0782	0.1951	-0.3155	0.4583
1x8	0.0672	0.2170	-0.3464	0.4984
65x2	0.0568	0.1893	-0.3251	0.4314
65x5	0.0380	0.1916	-0.3399	0.4070
62x2	0.0283	0.2103	-0.4031	0.4354
1x5	0.0258	0.2031	-0.3745	0.4215
42x17	0.0224	0.2096	-0.3818	0.4450
61x2	0.0203	0.2093	-0.3773	0.4542
18x17	0.0053	0.2023	-0.3639	0.4413
32x17	-0.0005	0.2031	-0.3880	0.3987
9x7	-0.0137	0.1899	-0.3797	0.3706
30x17	-0.0170	0.2095	-0.4437	0.3816
9x5	-0.0173	0.1843	-0.3703	0.3486
48x28	-0.0319	0.2143	-0.4588	0.3796
51x55	-0.0386	0.1948	-0.4038	0.3641
1x4	-0.0420	0.1928	-0.4383	0.3198
65x4	-0.0444	0.2122	-0.4821	0.3455

67x51	-0.0511	0.2013	-0.4427	0.3501
9x14	-0.0644	0.2050	-0.4698	0.3373
9x13	-0.0898	0.2052	-0.5163	0.2895
19x17	-0.0988	0.2097	-0.5253	0.2964
9x10	-0.1137	0.2089	-0.5450	0.2846
51x52	-0.1378	0.2103	-0.5441	0.2859
31x17	-0.1420	0.2128	-0.5660	0.2611
29x17	-0.1567	0.2128	-0.5726	0.2706
41x17	-0.1631	0.2192	-0.5913	0.2616
9x4	-0.1641	0.1953	-0.5496	0.2124
49x38	-0.1732	0.2197	-0.6043	0.2465
44x17	-0.1860	0.2199	-0.6294	0.2267
9x15	-0.2012	0.2167	-0.6130	0.2219
45x17	-0.2032	0.2214	-0.6319	0.2278
51x53	-0.2087	0.2209	-0.6546	0.1968
43x17	-0.2384	0.2315	-0.6991	0.1924
24x17	-0.2414	0.2334	-0.7103	0.1861
20x17	-0.2573	0.2339	-0.6919	0.2040
40x17	-0.2612	0.2358	-0.6979	0.2068
27x28	-0.3067	0.2638	-0.8129	0.1769
23x17	-0.3077	0.2505	-0.7972	0.1556
65x38	-0.3079	0.2046	-0.7325	0.0602
37x17	-0.3838	0.2676	-0.8669	0.1150
50x17	-0.3879	0.2735	-0.8914	0.1276
51x56	-0.3885	0.2662	-0.8771	0.1004
65x7	-0.4370	0.2352	-0.8709	0.0156
51x57	-0.5120	0.3063	-1.0490	0.0569
22x17	-0.6032	0.3423	-1.1873	0.0356

Table S13. Mean effects of 67 evaluated lines in 14 environments obtained through REML model.

Line	Mean Effect	Line	Mean Effect	Line	Mean Effect
P1	0.1756	P59	0.0184	P49	-0.0352
P6	0.1147	P60	0.0184	P44	-0.0374
P36	0.0872	P16	0.0174	P15	-0.0411
P26	0.0706	P42	0.0064	P45	-0.0413
P21	0.0700	P62	0.0058	P53	-0.0414
P25	0.0696	P5	0.0056	P7	-0.0464
P63	0.0666	P61	0.0031	P43	-0.0496
P2	0.0641	P32	0.0010	P24	-0.0500
P12	0.0631	P18	0.0001	P20	-0.0537
P64	0.0608	P38	-0.0029	P40	-0.0541
P3	0.0556	P30	-0.0033	P4	-0.0558
P54	0.0554	P48	-0.0053	P27	-0.0634
P58	0.0551	P55	-0.0073	P23	-0.0654
P34	0.0536	P67	-0.0100	P37	-0.0801
P46	0.0519	P14	-0.0112	P50	-0.0810
P9	0.0407	P28	-0.0143	P56	-0.0835
P11	0.0379	P13	-0.0180	P17	-0.0844
P47	0.0367	P19	-0.0208	P57	-0.1039
P33	0.0334	P65	-0.0211	P22	-0.1277
P8	0.0290	P10	-0.0235		
P66	0.0289	P52	-0.0261		
P35	0.0234	P31	-0.0315		
P39	0.0198	P41	-0.0336		
P51	0.0187	P29	-0.0343		

Table S14. Mean effects of 73 evaluated hybrids in 14 environments obtained through REML model.

Hybrid	Mean Effect	Hybrid	Mean Effect	Hybrid	Mean Effect
36x17	0.2998	39x17	0.0679	51x52	-0.0897
38x17	0.2887	59x60	0.0631	9x4	-0.1073
26x17	0.2426	16x17	0.0597	31x17	-0.1083
21x17	0.2408	1x8	0.0506	41x17	-0.1155
25x17	0.2392	9x8	0.049	29x17	-0.1178
63x51	0.2291	65x2	0.0472	49x38	-0.1209
9x12	0.2168	42x17	0.0221	44x17	-0.1287
64x51	0.2091	62x2	0.0198	9x15	-0.1414
9x6	0.1989	65x5	0.0143	45x17	-0.142
1x3	0.1912	1x5	0.0134	51x53	-0.1424
51x54	0.1906	61x2	0.0107	43x17	-0.1704
58x51	0.1895	32x17	0.0035	24x17	-0.1717
28x17	0.1868	18x17	0.0003	65x38	-0.1779
34x17	0.1844	9x5	-0.0085	20x17	-0.1847
65x51	0.1816	30x17	-0.0114	40x17	-0.186
46x17	0.1783	9x7	-0.017	27x28	-0.2179
1x2	0.1428	48x28	-0.0182	23x17	-0.225
1x7	0.1382	51x55	-0.025	37x17	-0.2753
9x11	0.1304	67x51	-0.0344	50x17	-0.2786
47x17	0.1263	1x4	-0.0357	65x7	-0.2807
33x17	0.1147	9x14	-0.0385	51x56	-0.287
1x6	0.103	65x4	-0.0487	51x57	-0.3571
65x66	0.0994	9x13	-0.0618	22x17	-0.4392
65x6	0.0922	19x17	-0.0715		
35x17	0.0805	9x10	-0.0808		

**CAPÍTULO 2 – USO DE VARIÁVEIS AMBIENTAIS NO ESTUDO DA INTERAÇÃO
GENÓTIPOS x AMBIENTES**

Versão preliminar de artigo redigido conforme norma do periódico científico Agronomy Journal. O conselho editorial do periódico poderá sugerir alterações para adequá-lo ao seu próprio estilo.

Ideias principais

USO DE VARIÁVEIS AMBIENTAIS NO ESTUDO DA INTERAÇÃO GENÓTIPOS x AMBIENTES

Luiz Antonio Yanes Bernardo Júnior, Carlos Pereira da Silva, Luciano Antonio de Oliveira, Vitor Passos da Silva Júnior, Marcio Balestre, Renzo Garcia Von Pinho

L.A.Y. Bernardo Júnior, V. P. da Silva Júnior, Universidade Federal de Lavras, Departamento de Biologia, Lavras, Minas Gerais, Brasil; C. P. da Silva, M. Balestre, Universidade Federal de Lavras, Departamento de Estatística, Lavras, Minas Gerais, Brasil; L.A. de Oliveira, Faculdade de Ciências Exatas e Tecnologias, Universidade Federal de Grande Dourados, Dourados, Mato Grosso do Sul, Brasil; R.G. Von Pinho, Universidade Federal de Lavras, Departamento de Agricultura, Lavras, Minas Gerais, Brasil.

Abreviações: AMMI, Additive Main Effects and Multiplicative Interaction; AOV, análise de variância; FDA, análise de dados funcionais; FPCA, análise de componentes principais funcional; GEI, interação genótipos x ambientes; PCA, análise de componentes principais; SVD, decomposição por valores singulares.

RESUMO

A análise de dados funcionais (FDA) é capaz de lidar com a análise e teoria de dados que se apresentam na forma de funções. Dados funcionais são intrinsecamente infinitos, e a alta ou infinita dimensionalidade dos dados constitui uma fonte rica de informações. O objetivo com este trabalho foi estudar a interação genótipos x ambientes (GEI) por meio do uso de variáveis ambientais sob abordagem funcional com diferentes volumes de dados e níveis de desbalanceamento. Foram simulados 20 conjuntos de dados para testar o comportamento do método em diferentes cenários. Para cada cenário, foram obtidas as correlações entre os valores preditos pelo modelo e os valores simulados em situações de dados completos e de desbalanceamento de dados. O método EM-AMMI (Expectation-Maximization Additive Main Effects and Multiplicative Interaction) sob o enfoque funcional foi aplicado através da integração do método smooth SVD (Singular Value Decomposition) ou SSVD ao método EM-AMMI. O seu comportamento foi testado via desbalanceamento por meio da validação cruzada. A análise de dados funcionais foi eficiente quando testada sob situações de aumento de genótipos, ambientes e número de componentes principais ou ao lidar com altos níveis de

desbalanceamento. A integração do método SSVD ao método EM-AMMI pode alterar a perspectiva de estudo da interação genótipos por ambientes. A identificação de padrões de efeitos da interação por meio do método EM-AMMI Funcional representa uma nova forma de entender a interação, e abre diversas possibilidades para aplicação no melhoramento de plantas.

INTRODUÇÃO

A interação genótipos por ambientes (GEI) é alvo de estudos de pesquisadores há décadas. Em países de clima tropical a GEI assume maior importância. No caso do Brasil, é possível realizar o plantio de duas safras em grande parte do país. Essas safras apresentam grande variação ambiental, o que exige a adoção de híbridos adaptados tanto para a região quanto para a época de plantio. Para isso, os ensaios multiambientais são essenciais para que a recomendação do híbrido seja realizada de maneira precisa.

A interação pode ocorrer por diversas razões, e entender as causas relacionadas a ela pode auxiliar na escolha do melhor modelo estatístico para o seu estudo. Os fatores ambientais que podem influenciar o desempenho genotípico já foram classificados por Allard e Bradshaw (1964). Fatores permanentes como clima, tipo de solo e comprimento do dia são classificados como previsíveis. Já fatores como quantidade e distribuição de chuvas e temperatura são não previsíveis. Essas duas categorias apresentam impactos bem distintos. O correto entendimento desses fatores pode permitir propor modelos capazes de estimar e isolar esses efeitos e melhorar a acurácia preditiva da análise.

Em alguns trabalhos clássicos sobre metodologias de estudo da GEI foram propostos modelos que levavam em conta um índice ou efeito ambiental. Eberhart e Russel (1966) propuseram o uso de regressão linear não-segmentada. O modelo considera um índice ambiental, o qual utiliza a média dos ambientes como uma medida de flutuação ambiental. Porém, não há independência da variável utilizada para estimar o índice ambiental e os dados das cultivares (Cossa, 1990; Ramalho et al., 2012).

Toler e Burrows (1998) propuseram um modelo de regressão não linear. A principal diferença para o método de Eberhart e Russel (1966) é que o modelo considera um efeito ambiental, o qual é um parâmetro do modelo obtido por meio de um método iterativo (Ramalho et al., 2012). Porém esse método se torna limitado, quando se deseja uma representação gráfica mais detalhada de estabilidade e adaptabilidade dos genótipos e ambientes em estudo.

Outros modelos que podem ser citados são os modelos de norma de reação, os quais permitem descrever a mudança na expressão fenotípica de um determinado genótipo ao longo

dos ambientes (Gravilets e Scheiner, 1993). É necessário ajustar modelos estatísticos que representem a norma de reação de genótipos para prever a performance fenotípica. Ou seja, os principais fatores ambientais responsáveis pelas diferenças fenotípicas precisam ser identificados em conjunto com formas funcionais adequadas para normas de reação (van Eeuwijk, Bustos-Korts, & Malosetti, 2016).

O método AMMI (Additive Main Effects and Multiplicative Interaction) é um dos métodos mais empregados nos últimos anos para estudo da GEI (Gauch, 2006; Gauch, Piepho, & Annicchiarico, 2008; van Eeuwijk, Bustos-Korts, & Malosetti, 2016; Bernardo Júnior et al., 2018). Esse método primeiramente utiliza análise de variância (AOV) para particionar a variação em efeitos de genótipos, ambientes e GEI. Em seguida, realiza a análise de componentes principais (PCA) para avaliar GEI (Zobel, Wright, & Gauch 1988; Gauch, 2013). O método AMMI também permite realizar a análise gráfica por meio de biplots. O biplot (Gabriel, 1971) é um gráfico de dispersão o qual fornece a visualização gráfica de uma tabela de dupla entrada. Esse tipo de gráfico permite visualizar simultaneamente as relações entre seus fatores de linha, entre seus fatores de coluna e as interações entre esses dois fatores (Yan e Tinker, 2006). O método AMMI, através do biplot, permite visualizar os padrões e interações específicas existentes nos dados (Zobel, Wright, & Gauch, 1988).

Um programa de melhoramento é normalmente desbalanceado (de Figueiredo et al., 2015). Vários genótipos são descartados e outros vários são adicionados ao programa de melhoramento ao longo dos ciclos de seleção. Dessa forma, o melhorista deve dispor de métodos que sejam capazes de lidar com esse dinamismo de um programa de melhoramento. Entre esses métodos, pode-se citar o EM-AMMI, proposto por Gauch e Zobel (1990). Esse método implementa o algoritmo EM (Expectation-Maximization) para realizar a imputação de dados e permitir o ajustamento do modelo AMMI em cenários de desbalanceamento de dados. Esse método permite uma exata e completa decomposição de toda informação retirada de um ensaio em informações diretas e informações indiretas. Para um determinado tratamento, as informações diretas representam suas repetições, enquanto as informações indiretas representam todos os outros dados do experimento.

O aumento do volume de informações fenotípicas, genotípicas e ambientais disponíveis estimulam o desenvolvimento de novas técnicas estatísticas que possibilitem uma descrição e predição mais precisa de GEI e uma maior eficiência nos programas de melhoramento (van Eeuwijk, Bustos-Korts, & Malosetti, 2016). Nesse contexto, uma alternativa ao método AMMI poderia ser aplicar a PCA sob uma perspectiva de análise de dados funcionais (FDA).

A FDA é capaz de lidar com a análise e teoria de dados que se apresentam na forma de funções (Wang, Chiou, & Müller, 2016). A FDA é uma metodologia a qual considera as funções dos dados observados como entidades únicas, e não como uma simples sequência de observações. O termo funcional refere-se à estrutura intrínseca dos dados. Na prática, dados funcionais são observados discretamente como n pares (t_j, y_j) , em que y_j representa um instante da função no tempo t_j . Essa função t_j comumente representa tempo, mas outras medidas contínuas podem ser consideradas, como posição espacial, frequência, peso, entre outras (Ramsay e Silverman, 2005, p. 38). Dados funcionais são intrinsecamente infinitos, e trazem desafios tanto para a teoria quanto para a computação, os quais variam de acordo com a sua amostragem. Por outro lado, a alta ou infinita dimensionalidade dos dados constitui uma fonte rica de informações, e pode criar inúmeras oportunidades para pesquisa e análise de dados (Wang, Chiou, & Müller, 2016).

Os objetivos da FDA são os mesmos existentes em qualquer outro ramo da estatística. Esse tipo de análise dá destaque a diversas características do conjunto de dados, permite estudar importantes fontes de variação e padrão dos dados e explica a variação existente em uma variável dependente por meio da informação de uma variável independente (Ramsay e Silverman, 2005, p. 9-10). Esse tipo de análise é extensivamente utilizado em diversas áreas da matemática, porém foi pouco explorado nos estudos da interação genótipos por ambientes. Krzyśko et al. (2013) demonstraram a aplicabilidade e utilidade da PCA de dados funcionais (FPCA) associada à análise de agrupamento para analisar a produtividade de cultivares de trigo de inverno. A metodologia mostrou-se uma ferramenta eficiente para realizar uma classificação confiável das cultivares avaliadas.

O tipo mais comum de dados adequados para a análise por FDA são séries temporais não estacionárias, cujas observações possivelmente são espaçadas de forma desigual, como médias de temperatura e precipitação (Ramsay e Dalzell, 1991). Esses tipos de dados são utilizados para compor o índice ambiental de métodos de regressão como o de Eberhart e Russel (1966), ou seja, tratam-se dos próprios dados analisados. Portanto, esses dados apresentam grande variação entre anos e entre safras e são não previsíveis. Essa variação desmonta todo o padrão de interação identificado em determinado ano ou safra, ou seja, esse padrão deixa de ter sentido. Isso se torna uma dificuldade no contexto da predição da estabilidade e adaptabilidade genotípica e ambiental. Portanto, o modelo FDA adequado deve ser aquele que possui parâmetros não variáveis e de fácil mensuração, como dados de latitude e longitude, os quais são constantes ao longo de anos e safras.

No contexto de GEI, o enfoque funcional permite ordenar os ambientes sequencialmente dentro do intervalo real T para se realizar a análise. Dessa forma, é possível propor uma relação funcional entre o comportamento genotípico e o efeito de GEI. Essa interação pode ser decomposta em interações com diferentes variáveis ambientais, como genótipos x latitude ou genótipos x longitude. Em um país de dimensões continentais como o Brasil, é esperado que essas interações tenham grande influência no desempenho genotípico. Isso permitiria explorar interações mais específicas entre genótipos e ambientes, as quais vão muito além de regiões delimitadas geograficamente. Essa estratégia seria interessante do ponto de vista da recomendação de cultivares por permitir uma estratificação ambiental mais precisa. Portanto, o enfoque difere bastante dos métodos de Eberhart e Russel (1966) e Toler e Burrows (1998).

O estudo de GEI sob uma abordagem funcional pode trazer uma nova forma de entender a interação e trazer novas perspectivas na determinação da estabilidade e adaptabilidade por meio de um novo enfoque. Com isso, o objetivo com este trabalho foi estudar GEI por meio do uso de variáveis ambientais no modelo AMMI sob abordagem funcional com diferentes volumes de dados e níveis de desbalanceamento. Para isso, visou-se avaliar o comportamento da FPCA em cenários de desbalanceamento de dados e integrá-la ao método EM-AMMI para realizar a predição genotípica sob uma perspectiva funcional.

MATERIAL E MÉTODOS

Nesta seção, serão descritos os métodos utilizados para o estudo da interação genótipos por ambientes sob uma perspectiva funcional. Primeiramente, será demonstrado como foi feita a simulação dos dados dos diferentes cenários para validação do método. Então, serão demonstrados os métodos de avaliação do comportamento da FPCA sob condições de dados completos e dados desbalanceados. Posteriormente, será demonstrado como foi realizado o método EM-AMMI associado à FPCA.

Simulação dos dados

Os dados simulados foram gerados com estrutura autorregressiva de primeira ordem. Os ambientes possuem distribuição normal multivariada com a estrutura de covariância entre ambientes representada por

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho^{J-1} \\ \rho & 1 & \dots & \rho^{J-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{J-1} & \rho^{J-2} & \dots & 1 \end{bmatrix}$$

e com média igual à um vetor de zeros $\mathbf{0}$, em que J é o número de ambientes. Os J ambientes receberam as coordenadas especificadas por

$$\mathbf{t} = \begin{bmatrix} 1/J \\ 2/J \\ \vdots \\ 1 \end{bmatrix}$$

de forma a simular um intervalo contendo coordenadas geográficas de latitude.

Análise de dados funcionais (FDA)

Para verificar a aplicabilidade do modelo proposto, foram simulados 20 conjuntos de dados para testar o comportamento do método em diferentes cenários (Tabela 1). Esses cenários foram simulados com diferentes números de genótipos e de ambientes, além de diferentes valores de correlação entre os ambientes simulados. Também foram testadas análises com 10 e 30 componentes principais retidos pelo modelo. Em alguns cenários, ocorrem maior número de genótipos do que de ambientes, enquanto em outros ocorre o oposto, em diferentes proporções.

Tabela 1. Cenários simulados com diferentes números de genótipos (NG) e de ambientes (NA), com correlações entre os ambientes (r) de 0.2 e 0.8 e número de componentes principais (NPC) de 10 e 30.

Cenário	NG	NA	r	NPC
1	50	100	0.8	10
2	50	100	0.2	10
3	1000	100	0.8	10
4	1000	100	0.2	10
5	2000	100	0.8	10
6	2000	100	0.2	10
7	3000	100	0.8	10
8	3000	100	0.2	10
9	50	200	0.8	10
10	50	200	0.2	10
11	50	300	0.8	10
12	50	300	0.2	10
13	300	500	0.8	10
14	300	500	0.2	10
15	100	100	0.8	10
16	100	100	0.2	10
17	300	300	0.8	10
18	300	300	0.2	10
19	2000	100	0.8	30
20	2000	100	0.2	30

Seguindo as sugestões de Zobel, Wright, & Gauch (1988), o ajuste por quadrados mínimos na análise AMMI para dados balanceados foi obtido em dois passos. Primeiro, os efeitos principais da parte aditiva do modelo, ou seja, a média geral e os efeitos genotípicos e ambientais, são obtidos por AOV comum. Isso gera um resíduo não aditivo, que constitui a interação. O segundo passo é a análise do componente multiplicativo do modelo, representado pela interação, por meio de PCA. Se todos os componentes principais forem retidos pelo modelo, o modelo completo irá possuir o mesmo número de graus de liberdade que os dados originais possuem. Dessa forma, o modelo se ajustará aos dados perfeitamente. Porém, o objetivo é obter a maior parte da interação em alguns poucos componentes principais para se ter um modelo AMMI reduzido que gerará um resíduo, sendo que as estimativas dos parâmetros do modelo serão obtidas pelo método dos mínimos quadrados.

O modelo AMMI pode ser descrito da seguinte forma (Gauch, Piepho, & Annicchiarico, 2008):

$$Y_{ge} = \mu + \alpha_g + \beta_e + \sum_n \lambda_n \gamma_{gn} \delta_{en} + \rho_{ge}$$

em que

$$\alpha_g = \mu_g - \mu$$

$$\beta_e = \mu_e - \mu$$

e

$$\hat{\mu} = \frac{\sum_{ge} Y_{ge}}{rc}; \hat{\mu}_g = \frac{\sum_e Y_{ge}}{c}; \mu_e = \frac{\sum_g Y_{ge}}{r}$$

em que Y_{ge} é a produtividade do genótipo g no ambiente e ; μ , μ_g e μ_e são a média geral, a média do genótipo g sobre ambientes e a média do ambiente e sobre genótipos, respectivamente; α_g é o desvio genotípico; β_e é o desvio ambiental; λ_n é o valor singular do componente principal n ; γ_{gn} é o vetor singular do genótipo g ; δ_{en} é o vetor singular do ambiente e ; ρ_{ge} é o resíduo; r é o total de genótipos; e c é o total de ambientes.

Gauch, Piepho, e Annicchiarico (2008) demonstram que esta equação pode ser escrita separando-se as partes aditiva e multiplicativa. Isto é feito trocando os desvios por médias na parte aditiva do modelo para então isolá-los da parte multiplicativa:

$$Y_{ge} - \mu_g - \mu_e + \mu = \theta_{ge} = \sum_n \lambda_n \gamma_{gn} \delta_{en} + \rho_{ge}$$

em que

$$\hat{Y}_{ge} = \mu_g + \mu_e - \mu$$

$$\theta_{ge} = Y_{ge} - \hat{Y}_{ge}$$

em que θ_{ge} representa a interação.

Assim, a matriz de interação é dada por:

$$\begin{bmatrix} \theta_{11} & \theta_{12} & \dots & \theta_{1c} \\ \theta_{21} & \theta_{22} & \dots & \theta_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{r1} & \theta_{r2} & \dots & \theta_{rc} \end{bmatrix}$$

Através da decomposição por valores singulares (SVD), a matriz de interação, de dimensão $r \times c$, pode ser decomposta da seguinte forma

$$\theta_{ge} = USV' = \sum_n \lambda_n \gamma_{gn} \delta_{en}$$

em que S é uma matriz diagonal contendo os valores singulares λ_n ; e U e V são matrizes que contém em suas colunas os vetores singulares γ_{gn} e δ_{en} , respectivamente, com $p = \min(r, c)$ e $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_p$.

O modelo funcional proposto no presente estudo parte do pressuposto de que o efeito da interação foi descrito por uma função espacial dos ambientes. Dessa forma, foi possível sequenciar os ambientes ao simular coordenadas de latitude e analisar informações que vão além de pontos que representam observações discretas. Esse sequenciamento permite avaliar o efeito da variação da latitude no comportamento genotípico. Quanto mais próximos os locais, ou seja, quanto mais próximas forem as coordenadas de latitude entre dois ambientes, mais correlacionados eles serão e, conseqüentemente, maior a probabilidade de apresentarem o mesmo padrão de interação.

Como os dados não são coletados na forma de curvas, o processo de suavização exigiu que se realizasse a passagem dos dados da forma discreta para a forma de funções para que fosse possível obter valores das unidades amostrais para quaisquer valores dentro de um intervalo definido. Essa etapa foi parte do processo de modelagem e análise, pois no momento em que as funções são definidas em cima de um sistema de bases, as estimativas também ficaram assim definidas (da Silveira Neto, 2013).

Foi utilizada a proposta de Ramsay e Silverman (2005, p. 149-151), em que a PCA é definida dentro de um contexto funcional. No caso do modelo AMMI, utilizado no presente estudo, o valor singular λ_n e os vetores singulares γ_{gn} e δ_{en} são valores funcionais, em que o índice discreto n é substituído por um índice contínuo s . Portanto, os somatórios sobre n foram substituídos por integrais em s para definir o produto interno

$$\theta_{ge} = \int \lambda(s)\gamma_g(s)\delta_e(s)ds + \rho_{ge}$$

Portanto, o modelo AMMI Funcional pode ser descrito como

$$Y_{ge} = \mu + \alpha_g + \beta_e + \int \lambda(s)\gamma_g(s)\delta_e(s)ds + \rho_{ge}$$

Dessa forma, os ambientes são organizados com base no componente funcional para realização da PCA. Os pontos da curva, que podem representar diferentes variáveis ambientais, como dados de latitude ou longitude, são colocados em ordem para gerar uma autocorrelação entre as curvas e possibilitar a realização da PCA. No presente estudo, foram simuladas coordenadas de latitude na forma decimal, plotadas com duas casas decimais em ordem crescente no eixo x dos gráficos construídos. Para isso, foram estipuladas coordenadas em um intervalo com mínimo de -32.01 e máximo de 3.18, de forma a simular ambientes localizados entre os estados do Rio Grande do Sul e do Pará, respectivamente. Dessa forma, é possível representar as mais variadas condições ambientais presentes no Brasil.

O método funcional utilizado neste trabalho é o smooth SVD ou SSVD proposto por Xiao et al. (2014). Esse método utiliza splines suavizados dos autovetores obtidos de uma SVD da matriz de dados. Considerou-se que cada observação X_i é uma observação de um processo funcional X com função de covariância $K(s, t)$, $s, t \in [0,1]$, em que $i = 1, \dots, I$, e I é o número de genótipos. Os dados observados $Y_{ij} = X_i(t_j) + \varepsilon_{ij}$ são representantes da observação X_i nos pontos amostrados $\{t_1, \dots, t_j\}$. Cada ε_{ij} são erros independentes e identicamente distribuídos com média zero e variância 12.25, sem perda de generalidade, os quais são mutuamente independentes das observações X_i .

O método SSVD (Xiao et al., 2014) considera que a função de covariância amostral é computada para cada par de pontos amostrais (t_j, t_ℓ) por $\hat{K}(t_j, t_\ell) = I^{-1} \sum_i Y_{ij} Y_{i\ell}$. Cada observação Y_{ij} foi centralizada ao longo dos genótipos. A matriz $\hat{\mathbf{K}}$ é uma matriz $J \times J$ com a posição (j, ℓ) igual a $\hat{\mathbf{K}}(t_j, t_\ell)$, em que J é o número de ambientes. A suavização da covariância refere-se a suavizações bivariadas ao estimador $\hat{\mathbf{K}}$. Considere que $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ij})^T$, $i = 1, \dots, I$, então $\hat{\mathbf{K}} = I^{-1} \sum_{i=1} \mathbf{Y}_i \mathbf{Y}_i^T = I^{-1} \mathbf{Y} \mathbf{Y}^T$, em que \mathbf{Y} é uma matriz $I \times J$ com a i -ésima coluna igual a \mathbf{Y}_i .

A abordagem de Xiao et al. (2014) utilizada aqui para estimar as autofunções e os autovalores é a decomposição da matriz $\hat{\mathbf{K}}$ e a suavização dos autovetores. Inicialmente,

considere que $\mathbf{U}_y \mathbf{D}_y \mathbf{V}_y^T$ é a SVD da matriz \mathbf{Y} . \mathbf{U}_y é uma matriz $I \times J$ ortonormal por coluna. \mathbf{V}_y é uma matriz $J \times J$ ortogonal. \mathbf{D}_y é uma matriz $J \times J$ diagonal. \mathbf{U}_y possui em suas colunas todos os autovetores de $\hat{\mathbf{K}}$ que são associados com autovetores não nulos. Os elementos da diagonal de $I^{-1} \mathbf{D}_y^2$ contém todos os autovalores não nulos de $\hat{\mathbf{K}}$. Dessa forma, a obtenção das matrizes \mathbf{U}_y e $I^{-1} \mathbf{D}_y^2$ é equivalente à decomposição espectral da matriz de covariância amostral $\hat{\mathbf{K}}$. Então, a suavização dos autovetores retidos foi realizada através de splines suavizados por meio da função “smooth.spline” do software R, versão 2.15.2 (R Development Core Team, 2018). Nesse caso, todos os pontos distintos representantes dos ambientes são utilizados como nós do spline.

Predição dos dados faltantes

Para obtenção das predições, primeiramente foi realizada a SVD. Depois, com base na SVD, é utilizada a decomposição realizada por análise de covariância da predição do vetor do genótipo desbalanceado. As duas decomposições são combinadas para se realizar a predição. Nesse caso, são utilizados os ajustes dos splines para se realizar uma regressão sobre os efeitos ambientais e obter os escores dos genótipos. Os escores genotípicos são utilizados para se realizar uma combinação linear com os escores ambientais para se obter as predições.

A capacidade preditiva do método SSVD no contexto de GEI foi testada utilizando dados completos e dados desbalanceados.

Capacidade preditiva de ajuste de modelo

Análise com dados completos

Inicialmente, foi avaliado o comportamento do método SSVD em uma análise de dados balanceados, sem perdas de genótipos ou de ambientes. Considere uma matriz de interações, na qual os genótipos estão dispostos nas linhas e os ambientes nas colunas. Após o ajuste do modelo, foi obtida a correlação entre valores simulados e preditos das colunas, ou seja, dos valores simulados dos genótipos em cada ambiente com os valores preditos pelo modelo. Dessa forma, é possível verificar como o modelo se comporta na predição dos genótipos em cada ambiente. Também foi obtida a correlação entre os dados de interação e o ajuste das curvas dos genótipos, com o objetivo de avaliar o comportamento do método ao predizer o padrão de interação.

Nesse contexto, é oportuno demonstrar como ocorre a estrutura de interação de acordo com a correlação existente entre ambientes. Dessa forma, é possível evidenciar a relação funcional existente entre os ambientes ao ordená-los sequencialmente por meio de suas coordenadas de latitude. Para isso, foi realizada uma análise adicional em que foram obtidos os gráficos das curvas para os quatro primeiros componentes principais para os cenários 7 e 8. Esses cenários foram selecionados por simularem uma situação com grande número de genótipos, de forma a representar uma situação real de programas de melhoramento. Porém, é comum melhoristas optarem por biplots construídos com dois componentes principais em análises de ensaios multiambientais. Para que essa situação fosse comparada ao método proposto no presente estudo, foram adicionados apenas a essa análise adicional os cenários 21, 22, 23 e 24 (Tabela 2), os quais simulam as mesmas condições dos cenários 7 e 8, porém com 2 e 6 componentes principais. Assim, é possível demonstrar as consequências de uma visão tradicional ao se utilizar a FPCA em ensaios multiambientais.

Tabela 2. Cenários selecionados para representar a estrutura de interação existente ao se variar a correlação entre ambientes (r) e número de componentes principais (NPC) na análise de dados completos, sem desbalanceamento de dados.

Cenário	Número de Genótipos	Número de Ambientes	r	NPC
7	3000	100	0.8	10
8	3000	100	0.2	10
21	3000	100	0.8	2
22	3000	100	0.2	2
23	3000	100	0.8	6
24	3000	100	0.2	6

Capacidade preditiva com dados faltantes

Desbalanceamento de 1 ambiente

A primeira situação de desbalanceamento testada foi a de retirada de um ambiente inteiro, ou seja, foi feita a retirada de uma coluna da matriz de interações. O objetivo foi simular a perda de um ambiente completo e prever o comportamento de todos os genótipos em um ambiente não avaliado, mas com latitude conhecida. Para cada cenário (Tabela 1), foi realizado o desbalanceamento de 1 ambiente e obtida a correlação entre os valores simulados e preditos dos genótipos do ambiente desbalanceado. Isso foi repetido realizando-se o desbalanceamento de cada um dos ambientes para obter sua predição e correlação entre valores simulados e

preditos. Então, foi obtida a média de todas as correlações obtidas para cada ambiente desbalanceado para assim ter uma correlação média para cada um dos 20 cenários simulados.

Desbalanceamento Sistemático

O objetivo foi realizar o desbalanceamento de mais de um ambiente com coordenadas de latitude conhecidas, sob perdas sistemáticas, e verificar a capacidade do modelo em prever todos os genótipos nesses ambientes. Ou seja, foram retiradas colunas inteiras da matriz de interações de maneira sistemática. Como comentado anteriormente, quanto mais próximos os locais, ou seja, quanto mais próximas forem as coordenadas de latitude entre dois ambientes, mais correlacionados estes ambientes serão e, conseqüentemente, maior a probabilidade de apresentarem o mesmo padrão de interação. Ao se realizar o desbalanceamento de maneira sistemática o padrão de interação não é quebrado. Dessa forma, é possível avaliar o padrão dos dados ambientais na curva suavizada mesmo sob situação de desbalanceamento experimental. O desbalanceamento realizado de forma aleatória pode retirar genótipos de um único subconjunto de dados e quebrar o padrão de interação existente, dificultando as predições. Portanto, foi feito o desbalanceamento de 10% dos ambientes ($k = 10$) através de uma amostragem realizada de maneira sistemática para cada cenário (Tabela 1). Os dados foram divididos em k subconjuntos de dados de tamanho r . Entre os r primeiros foi sorteado 1 ambiente, digamos X_j , em que $1 \leq j \leq r$. Os demais valores $X_{i's}$ são dados por índices $i = j + r, j + 2r, \dots, j + kr$, com $i \leq k$, sendo $r \approx J/k$.

Para cada amostragem, foi obtida a correlação média entre valores simulados e preditos dos genótipos em cada ambiente, ou seja, entre valores simulados e preditos das colunas da matriz de interação. Então, foram obtidas as médias e variâncias das correlações entre as amostragens, ou seja, a média e a variância da correlação média.

Desbalanceamento Aleatório

Assim como na realização do desbalanceamento sistemático, o objetivo foi realizar o desbalanceamento de mais de um ambiente com coordenadas de latitude conhecidas e verificar a capacidade do modelo em prever todos os genótipos nesses ambientes. Porém, a retirada de ambientes foi realizada totalmente ao acaso. O desbalanceamento com amostragem aleatória de ambientes permite avaliar o comportamento do modelo ao quebrar o padrão de interação da variável ambiental e, conseqüentemente, dificultar a realização das predições, conforme

explicado no tópico anterior. Dessa forma, foi possível verificar a robustez do método sob uma possível quebra de determinado padrão de latitude. Para cada cenário (Tabela 1), foi feito o desbalanceamento de 5, 10, 20 e 30% dos ambientes ($k = 5$, $k = 10$, $k = 20$ e $k = 30$, respectivamente), totalmente ao acaso. Esse desbalanceamento aleatório foi repetido 1000 vezes para cada cenário. Para cada uma das 1000 vezes que o processo foi repetido, foi obtida a correlação média entre valores simulados e preditos dos genótipos em cada ambiente, ou entre colunas da matriz de interações, tanto para os ambientes eliminados quanto para os ambientes não eliminados. Então, foram obtidas as médias e as variâncias das correlações entre as amostragens, ou seja, a média e a variância da correlação média entre valores simulados e preditos dos genótipos em cada ambiente.

Capacidade preditiva de GEI

EM-AMMI Funcional

A proposta do presente estudo é realizar a análise EM-AMMI (Gauch e Zobel, 1990) sob um enfoque funcional. Foi aplicado o modelo funcional sob cenário de perdas ao acaso de alguns genótipos em alguns ambientes. A predição do comportamento genotípico em ensaios multiambientais, em que alguns genótipos são avaliados em alguns ambientes, mas não em outros, é uma situação a qual o melhorista comumente deve lidar (Burgeño et al., 2012). No presente estudo, para os genótipos avaliados em todos os ambientes, já existe um padrão de interação nas curvas suavizadas, conforme mencionado anteriormente. Porém, para os genótipos desbalanceados ocorre a quebra deste padrão de interação. Portanto, o objetivo foi verificar o quanto o EM-AMMI Funcional é capaz de prever o genótipo em determinado ambiente o qual não foi avaliado.

O método EM-AMMI busca substituir a informação perdida pela esperança dada pela média marginal. O algoritmo do método sob o enfoque funcional pode ser descrito da seguinte forma:

(i) os ambientes onde houve perdas de genótipos tem suas caselas preenchidas pela média marginal do genótipo como um valor inicial. As caselas vazias também podem ser preenchidas por zero como um valor inicial, opção escolhida no presente estudo. Dessa forma, a matriz de interação não apresenta caselas vazias e permite o cálculo dos parâmetros multiplicativos do modelo normalmente por meio de FPCA;

(ii) através do método EM-AMMI Funcional, a predição é feita por meio do ajuste da curva do genótipo desbalanceado com base na posição da variável ambiental, no caso latitude. Os parâmetros aditivos são calculados através das médias não ponderadas de genótipos, ambientes e média geral. Depois, calcula-se os efeitos da interação da forma usual, ou seja, $\theta_{ge} = Y_{ge} - \mu_g - \mu_e + \mu$;

(iii) os valores são então reestimados, e as estimativas referentes aos dados perdidos são imputadas nas caselas vazias da matriz de interação inicial;

(iv) o processo se inicia novamente e continua até haver convergência, ou seja, até que os valores imputados apresentem pequena diferença de uma imputação para outra.

Para se realizar a análise EM-AMMI sob uma perspectiva funcional, o método de Xiao et al. (2014) foi incorporado ao modelo. As predições foram realizadas conforme descrito anteriormente. Para avaliar o comportamento do modelo, foram realizadas duas validações cruzadas. A primeira foi feita através do desbalanceamento de 5, 10, 20 e 30% dos genótipos, por meio da retirada aleatória de caselas da matriz original de observações. A segunda validação cruzada foi realizada por meio do desbalanceamento de 10% dos ambientes inteiros, ou seja, através da retirada de 10% das colunas da matriz de interações. Um k -fold consiste na partição aleatória dos dados em k subconjuntos não sobrepostos. A validação cruzada é então aplicada em $k-1$ grupos denominados conjuntos de treinamento, enquanto o grupo restante é o conjunto de validação (Burgeño et al., 2012). Foram obtidas as médias das correlações entre valores simulados e preditos dentro de cada subconjunto de dados. Destas médias de cada subconjunto, foi estimada a média dentro de cada cenário de desbalanceamento.

Todas as análises foram realizadas no software R, versão 2.15.2 (R Development Core Team, 2018).

RESULTADOS

Capacidade preditiva de ajuste de modelo

Análise com dados completos

A primeira situação em que os 20 cenários simulados foram testados foi com o modelo completo, sem desbalanceamento, por meio da correlação dos valores simulados dos genótipos em cada ambiente com os valores preditos pelo modelo, ou seja, por meio da correlação entre valores simulados e preditos nas colunas da matriz de interações (Tabela 3). Para todos os

cenários, é possível verificar que a maior correlação entre os ambientes leva à uma maior acurácia preditiva da análise. Porém, mesmo com uma baixa correlação entre ambientes, em geral a acurácia preditiva foi relativamente alta. Quanto aos ajustes (R^2), observa-se a mesma tendência das correlações, em que ocorre maiores valores para os cenários em que foi simulada uma correlação entre ambientes de 0.8. Portanto, a maior correlação entre ambientes permite uma maior previsibilidade de comportamento destes.

Tabela 3. Correlação média dentro de cada ambiente e ajuste da curva (R^2) entre valores simulados e preditos nas colunas da matriz de interações pelo modelo SSVD em condição de dados completos, ou seja, sem desbalanceamento de genótipos ou de ambientes.

Cenário	NG	NA	R	NPC	Correlação	R^2
1	50	100	0.8	10	0.8301	0.6891
2	50	100	0.2	10	0.5267	0.2774
3	1000	100	0.8	10	0.7739	0.5989
4	1000	100	0.2	10	0.4138	0.1712
5	2000	100	0.8	10	0.7771	0.6039
6	2000	100	0.2	10	0.3987	0.1590
7	3000	100	0.8	10	0.7783	0.6058
8	3000	100	0.2	10	0.3993	0.1594
9	50	200	0.8	10	0.7457	0.5561
10	50	200	0.2	10	0.3914	0.1532
11	50	300	0.8	10	0.6907	0.4771
12	50	300	0.2	10	0.3507	0.1230
13	300	500	0.8	10	0.4703	0.2212
14	300	500	0.2	10	0.2759	0.0761
15	100	100	0.8	10	0.8046	0.6474
16	100	100	0.2	10	0.4919	0.2420
17	300	300	0.8	10	0.5581	0.3115
18	300	300	0.2	10	0.3236	0.1047
19	2000	100	0.8	30	0.9292	0.8634
20	2000	100	0.2	30	0.6610	0.4369

NG: Número de Genótipos; NA: Número de Ambientes; r: Correlação entre Ambientes; NPC: Número de Componentes Principais; R^2 : ajuste do modelo.

Os resultados indicam que a capacidade preditiva do modelo pouco se altera com o aumento do número de genótipos, tanto em situações de correlação entre ambientes de 0.8 (cenários 1, 3, 5 e 7) quanto em cenários de correlação de 0.2 (cenários 2, 4, 6 e 8). O mesmo ocorre com o aumento do número de ambientes (cenários 1, 9, 11 e cenários 2, 10 e 12). Para os cenários com ambientes altamente correlacionados, mesmo ao se aumentar o número de genótipos para 3000 ou de ambientes para 300, a acurácia preditiva foi de alta magnitude. A previsibilidade de comportamento dos ambientes também foi pouco afetada, indicada pela baixa redução dos valores de R^2 . Esse resultado tem implicação direta na condução do programa de

melhoramento, que exige o uso de métodos que sejam capazes de lidar com um grande volume de dados. A escolha de ambientes correlacionados é determinante na qualidade das predições obtidas, o que reforça como a baixa correlação entre ambientes prejudica a previsibilidade de comportamento dos ambientes.

Os cenários 13 e 14 testam a situação de maior número de ambientes do que de genótipos, com valores altos para ambos. Nesses cenários, o aumento simultâneo de genótipos e ambientes levou a uma menor precisão nos resultados. Isso indica que, nesse tipo de situação, avaliar um grande volume de genótipos e ambientes simultaneamente poderá levar a predições pouco acuradas e, conseqüentemente, levar o melhorista a tomar decisões equivocadas.

Os cenários 15, 16, 17 e 18 testam uma situação em que ocorre o mesmo número de genótipos e de ambientes. O cenário 15 resultou em predições de alta magnitude. No cenário 16, onde reduziu-se a correlação entre ambientes para 0.2, a acurácia preditiva manteve-se relativamente alta, considerando a situação. Porém, a previsibilidade de comportamento foi muito reduzida. Vale ressaltar que o aumento do número de genótipos e ambientes também levou à uma piora da predição (Cenários 17 e 18).

Quanto ao número de componentes principais retidos pelo modelo, os resultados obtidos nos cenários 5, 6, 19 e 20 indicam uma maior precisão das predições com o aumento de 10 para 30 componentes principais. Esse aumento é mais evidente nos cenários 6 e 20, em que os ambientes são poucos correlacionados. Portanto, o uso de um maior número de componentes principais é uma possível alternativa para aumentar a precisão das predições quando os ambientes utilizados para condução dos ensaios são pouco correlacionados.

A Tabela 4 indica o comportamento do modelo SSVD ao predizer o desempenho genotípico por meio dos dados de interação e o ajuste das curvas dos genótipos. Os resultados indicam as mesmas tendências observadas na predição do comportamento ambiental em todos os cenários e comparações realizadas. Portanto, a previsibilidade do desempenho genotípico está intimamente ligada à estrutura de correlação entre ambientes. Adicionalmente, as estimativas de correlações e ajustes de cada genótipo apresentaram baixa variação (σ^2) em todos os cenários testados, o que mostra a qualidade das predições do método nas mais diversas situações possíveis.

Tabela 4. Médias das correlações entre genótipos e dos ajustes (R^2) entre valores simulados e preditos pelo modelo SSVD em condição de dados completos, ou seja, sem desbalanceamento de genótipos ou de ambientes.

Cenário	NG	NA	r	NPC	Correlação		R^2	
					Média	σ^2	Média	σ^2
1	50	100	0.8	10	0.8204	0.0051	0.6780	0.0123
2	50	100	0.2	10	0.5194	0.0066	0.2762	0.0071
3	1000	100	0.8	10	0.7304	0.0082	0.5417	0.0164
4	1000	100	0.2	10	0.4052	0.0066	0.1708	0.0045
5	2000	100	0.8	10	0.7304	0.0090	0.5425	0.0174
6	2000	100	0.2	10	0.3936	0.0066	0.1615	0.0042
7	3000	100	0.8	10	0.7316	0.0092	0.5444	0.0179
8	3000	100	0.2	10	0.3853	0.0068	0.1553	0.0042
9	50	200	0.8	10	0.7357	0.0066	0.5477	0.0139
10	50	200	0.2	10	0.4292	0.0053	0.1895	0.0042
11	50	300	0.8	10	0.6827	0.0088	0.4747	0.0151
12	50	300	0.2	10	0.3334	0.0054	0.1165	0.0028
13	300	500	0.8	10	0.4616	0.0080	0.2211	0.0063
14	300	500	0.2	10	0.2697	0.0034	0.0762	0.0011
15	100	100	0.8	10	0.7704	0.0071	0.6006	0.0159
16	100	100	0.2	10	0.4822	0.0072	0.2397	0.0071
17	300	300	0.8	10	0.5417	0.0092	0.3026	0.0101
18	300	300	0.2	10	0.3269	0.0046	0.1114	0.0020
19	2000	100	0.8	30	0.9150	0.0009	0.8382	0.0030
20	2000	100	0.2	30	0.6527	0.0034	0.4295	0.0057

NG: Número de Genótipos; NA: Número de Ambientes; r: Correlação entre Ambientes; NPC: Número de Componentes Principais; σ^2 : variância; R^2 : ajuste do modelo.

A estrutura de interação entre ambientes é determinante na qualidade das predições. Para todos os cenários, foram obtidas as curvas desta estrutura para melhor entendimento da influência da correlação entre ambientes na qualidade das predições. Para exemplificar como ocorre esta interação, foram selecionados os cenários indicados na Tabela 5 e Figuras 1 a 6. A Figura 1 indica um padrão de interação ao longo dos ambientes, indicado pela baixa variação das curvas. Fica claro que ambientes próximos apresentam efeitos da interação semelhantes. Por exemplo, entre as latitudes -32,01 e -24.19, considerando o primeiro componente principal, a maioria dos ambientes apresentam efeitos da interação que variam entre aproximadamente 0 e -1.0. Os gráficos da Figura 2 mostram comportamento oposto. Os três primeiros componentes principais apresentam grande variação de efeitos de interação, mesmo entre ambientes próximos, indicados pelos picos erráticos de efeitos da interação. Isso mostra uma total falta de correlação entre ambientes, não havendo ligação alguma entre um ambiente e outro. Tomando como exemplo novamente os ambientes situados no intervalo de latitude entre -32,01 e -24.19, para o primeiro componente principal, é possível verificar efeitos da interação que variam de menos de -1.0 a aproximadamente 2.0. Esse padrão tem impacto direto na previsibilidade do

comportamento ambiental e genotípico, e o método SSVD foi eficiente em captar essa estrutura de interação.

Tabela 5. Médias das correlações entre genótipos e dos ajustes (R^2) entre valores simulados e preditos pelo modelo SSVD em condição de dados completos, ou seja, sem desbalanceamento de genótipos ou de ambientes, para os cenários selecionados para representar a estrutura de interação existente ao se variar a correlação entre ambientes e número de componentes principais.

Cenário	NG	NA	r	NPC	Correlação		R^2	
					Média	σ^2	Média	σ^2
7	3000	100	0.8	10	0.7316	0.0092	0.5444	0.0179
8	3000	100	0.2	10	0.3853	0.0068	0.1553	0.0042
21	3000	100	0.8	2	0.2713	0.0395	0.1131	0.0121
22	3000	100	0.2	2	0.1618	0.0075	0.0337	0.0011
23	3000	100	0.8	6	0.5921	0.0181	0.3686	0.0231
24	3000	100	0.2	6	0.2943	0.0074	0.0940	0.0028

Ao se comparar os gráficos das Figuras 1 a 6, fica evidente a influência da escolha do número de componentes principais retidos pelo modelo na qualidade das predições. As Figuras 3 a 6 mostram que houve maior explicação da interação ao se utilizar 2 e 6 componentes principais em relação aos gráficos das Figuras 1 e 2, em que são retidos 10 componentes. Porém, a Tabela 5 indica baixa precisão na predição nos cenários 21, 22, 23 e 24. Portanto, ao se utilizar uma abordagem funcional, o uso de poucos componentes principais não é suficiente para a obtenção de boas predições. Isso se torna de grande relevância do contexto de estudo de GEI, em que comumente o melhorista opta por realizar inferências a partir dos dois primeiros componentes principais. Para obter maiores correlações e ajustes dentro de um contexto funcional e, conseqüentemente, melhorar a confiabilidade das predições, é preciso reter um maior número de componentes principais.

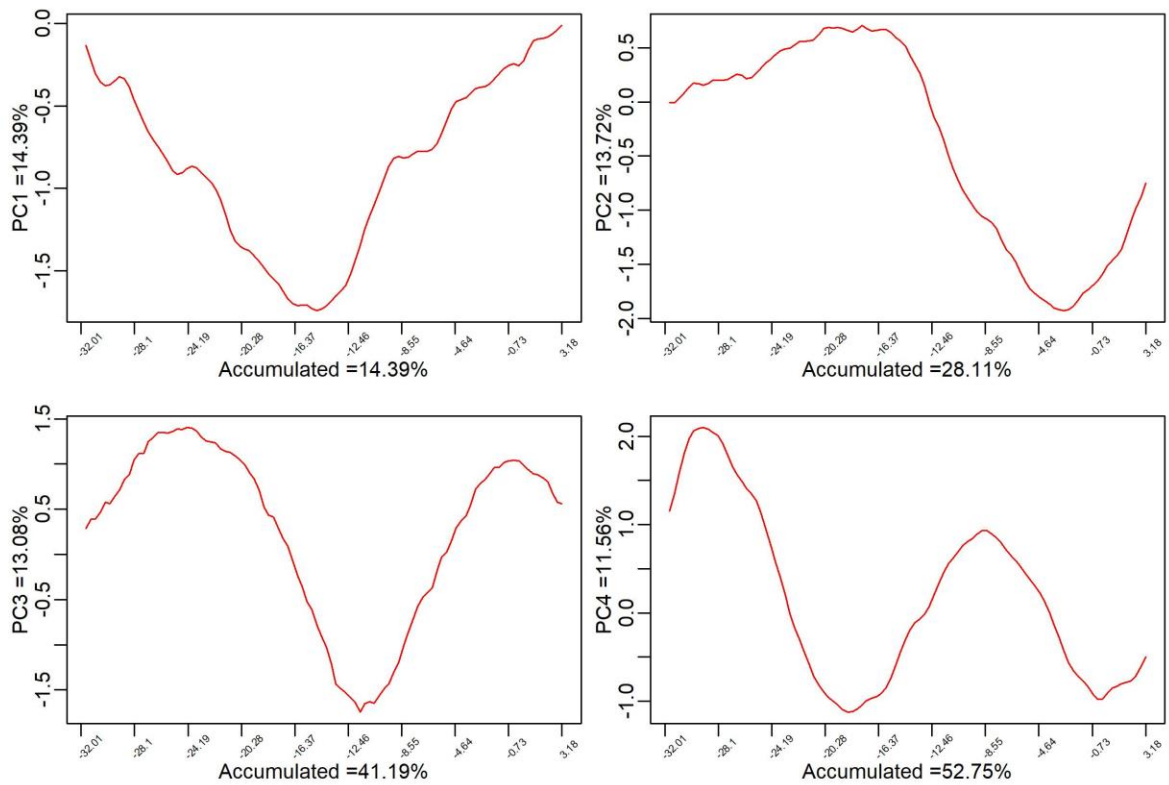


Figura 1. Estrutura de interação e porcentagem de explicação da interação para os quatro primeiros componentes principais relativos à análise realizada no cenário 7, com correlação entre ambientes de 0.8.

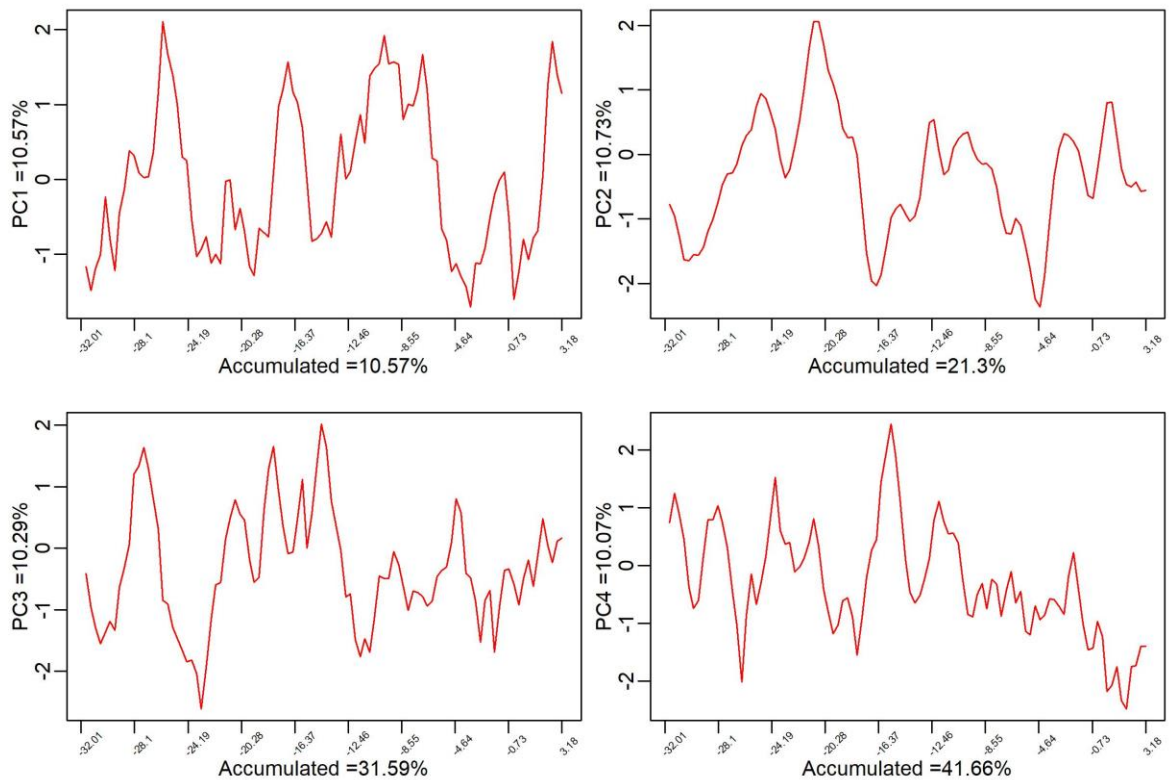


Figura 2. Estrutura de interação e porcentagem de explicação da interação para os quatro primeiros componentes principais relativos à análise realizada no cenário 8, com correlação entre ambientes de 0.2.

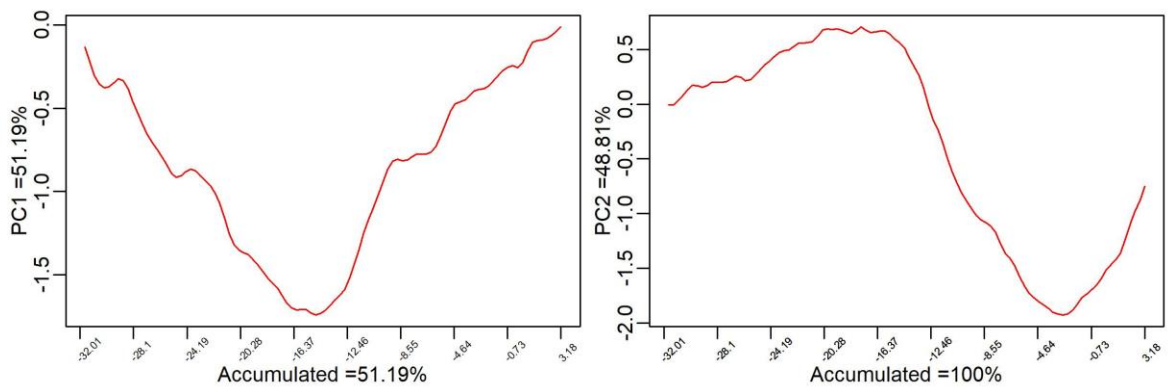


Figura 3. Estrutura de interação e porcentagem de explicação da interação para os dois primeiros componentes principais relativos à análise realizada no cenário 21, com correlação entre ambientes de 0.8.

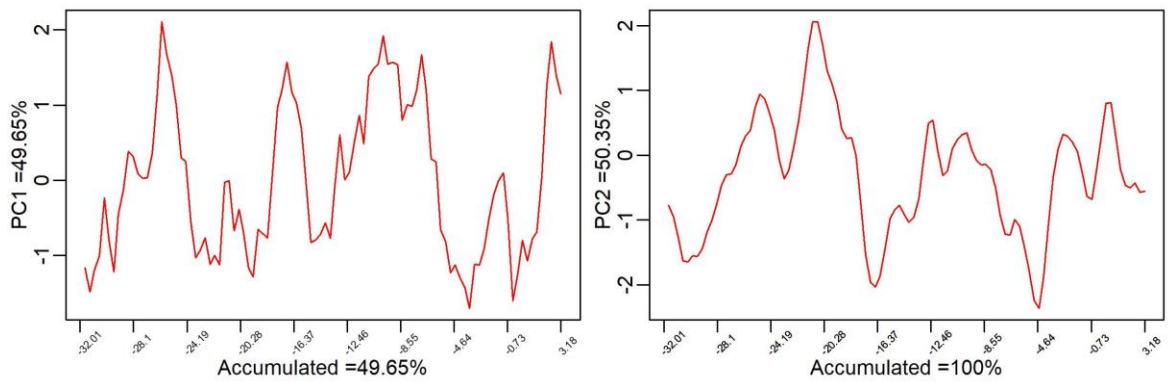


Figura 4. Estrutura de interação e porcentagem de explicação da interação para os dois primeiros componentes principais relativos à análise realizada no cenário 22, com correlação entre ambientes de 0.2.

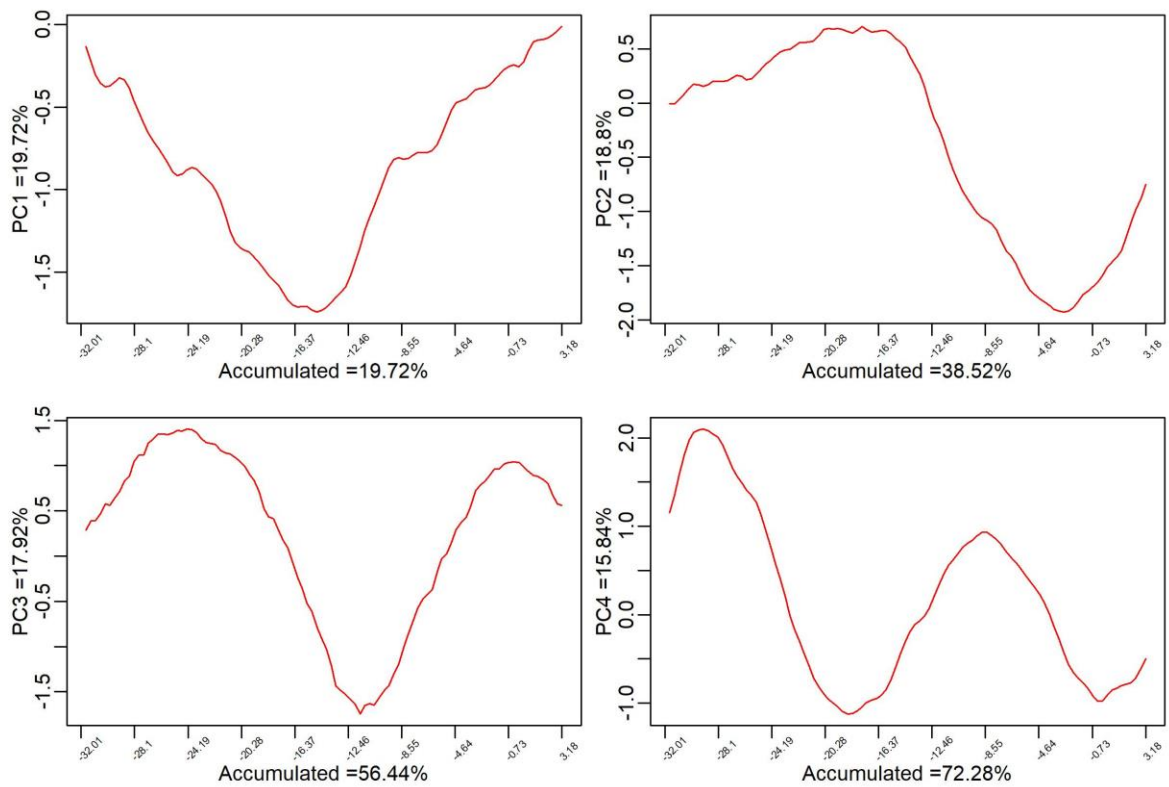


Figura 5. Estrutura de interação e porcentagem de explicação da interação para os quatro primeiros componentes principais relativos à análise realizada no cenário 23, com correlação entre ambientes de 0.8.

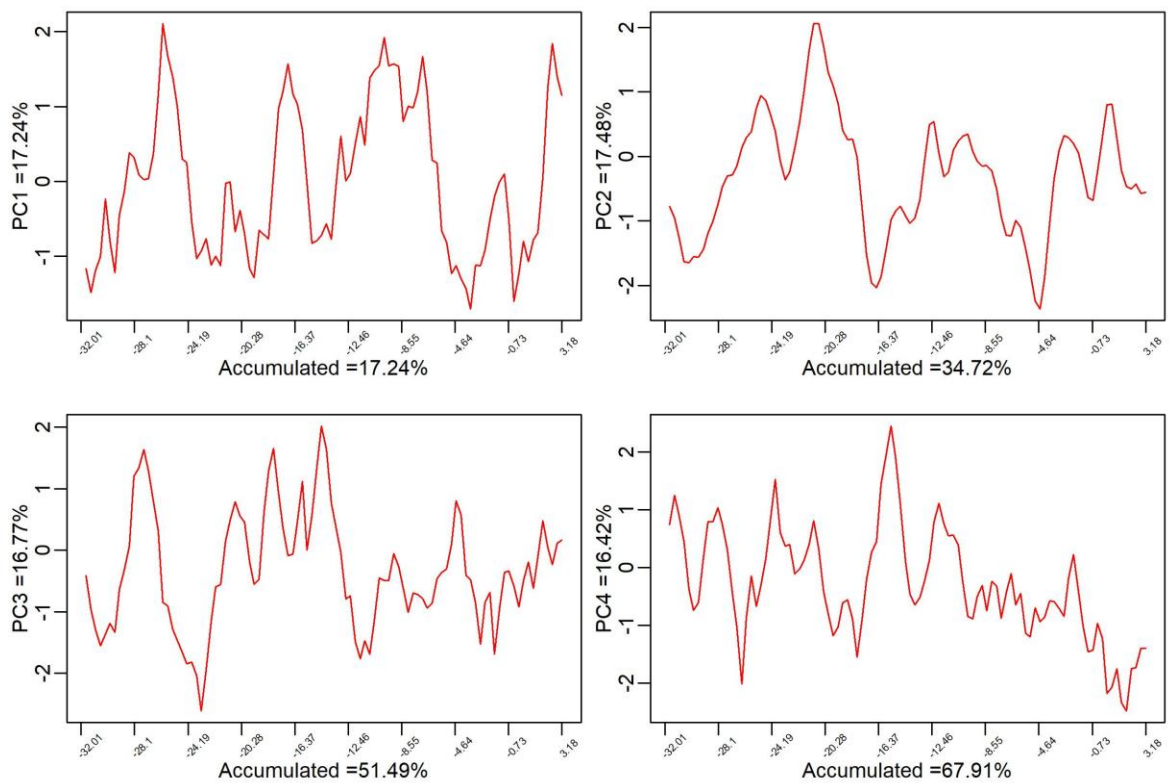


Figura 6. Estrutura de interação e porcentagem de explicação da interação para os quatro primeiros componentes principais relativos à análise realizada no cenário 24, com correlação entre ambientes de 0.2.

Capacidade preditiva com dados faltantes

Desbalanceamento de 1 ambiente

A Tabela 6 indica a acurácia preditiva de todos os genótipos em um ambiente nunca avaliado. Assim como para os resultados sem desbalanceamento, é possível verificar que a maior correlação entre os ambientes leva à uma maior acurácia preditiva da análise, independentemente do cenário.

Tabela 6. Médias de todas as correlações obtidas entre valores simulados e preditos dos genótipos de cada ambiente eliminado para todos os cenários.

Cenário	NG	NA	r	NPC	Média das Correlações
1	50	100	0.8	10	0.7495
2	50	100	0.2	10	0.1878
3	1000	100	0.8	10	0.6827
4	1000	100	0.2	10	0.0969
5	2000	100	0.8	10	0.6997
6	2000	100	0.2	10	0.0944
7	3000	100	0.8	10	0.7024
8	3000	100	0.2	10	0.0897
9	50	200	0.8	10	0.6715
10	50	200	0.2	10	0.1300
11	50	300	0.8	10	0.6229
12	50	300	0.2	10	0.1095
13	300	500	0.8	10	0.3929
14	300	500	0.2	10	0.0826
15	100	100	0.8	10	0.7154
16	100	100	0.2	10	0.1464
17	300	300	0.8	10	0.4646
18	300	300	0.2	10	0.0956
19	2000	100	0.8	30	0.8463
20	2000	100	0.2	30	0.1689

NG: Número de Genótipos; NA: Número de Ambientes; r: Correlação entre Ambientes; NPC: Número de Componentes Principais.

Os resultados apresentados na Tabela 6 revelam o impacto ocasionado pela perda de 1 ambiente inteiro na escolha do número de genótipos e ambientes a serem avaliados. A retirada de 1 ambiente completo não tem forte influência na capacidade preditiva do modelo ao se utilizar ambientes correlacionados, considerando desde 50 até 3000 genótipos (cenários 1, 3, 5 e 7). Porém, ao se conduzir ensaios em ambientes pouco correlacionados, a acurácia preditiva é reduzida drasticamente, e essa redução é maior ao se avaliar grande número de genótipos (cenários 2, 4, 6 e 8).

As mesmas tendências são observadas ao se variar o número de ambientes. A avaliação de 100, 200 e 300 ambientes forneceram acurácias preditivas semelhantes. Fica evidente que a correlação entre ambientes é o principal fator que influencia na qualidade das predições. Porém, para os cenários 13 e 14, assim como na análise de dados completos, a acurácia preditiva foi baixa.

Os cenários 15, 16, 17 e 18 representaram uma situação de mesmo número de genótipos e ambientes. O cenário 15 apresentou alta acurácia preditiva, porém houve uma queda brusca na qualidade das predições ao se avaliar com menor correlação entre ambientes (cenário 16). Independente da correlação existente entre os ambientes, o aumento do número de genótipos e ambientes de 100 para 300 levou à uma piora da predição.

Ao se aumentar o número de componentes principais, novamente houve um grande aumento na precisão (cenários 5, 6, 19 e 20). Assim como para o modelo em que não houve perda, esse aumento é mais evidente nos cenários em que a correlação entre ambientes é de 0.2, o que torna o uso de um maior número de componentes principais ainda mais atrativo.

Desbalanceamento Sistemático

As correlações obtidas entre valores simulados e preditos dos ambientes eliminados estão indicadas na Tabela 7. No Material Suplementar encontram-se as correlações obtidas pelo modelo para os ambientes não eliminados como indicativo do comportamento das predições pela quebra da estrutura pela retirada de 10% dos ambientes (Tabela S1).

A qualidade das predições sob retirada de vários ambientes inteiros de maneira sistemática também foi influenciada principalmente pela correlação entre ambientes. A qualidade das predições pouco se alterou com o aumento do número de genótipos ou de ambientes (cenários 1 a 12 das Tabelas 7 e S1 do Material Suplementar). Nos cenários com correlação entre ambientes de 0.2, a acurácia preditiva para os ambientes não eliminados (Tabela S1) foi muito superior a acurácia preditiva dos ambientes eliminados (Tabela 7). Nesse caso, deve-se levar em conta que, além dos ambientes terem sido retirados, trata-se de uma situação em que ocorrem ambientes pouco correlacionados. Portanto, havia dois fatores determinantes para a acurácia preditiva. Assim como nas análises anteriores, os cenários 13 e 14 mostram que o aumento do número de genótipos e de ambientes, com um maior número de ambientes, reduz a qualidade das predições.

Tabela 7. Médias e variâncias da correlação média obtidos entre valores simulados e preditos dos genótipos em cada ambiente em cada amostragem. Os resultados são referentes aos ambientes eliminados considerando uma retirada de 10 % dos ambientes realizada de maneira sistemática.

Cenário	NG	NA	r	NPC	Correlação Média	
					Média	σ^2
1	50	100	0.8	10	0.7571	0.00049
2	50	100	0.2	10	0.1871	0.00203
3	1000	100	0.8	10	0.6866	0.00020
4	1000	100	0.2	10	0.0991	0.00012
5	2000	100	0.8	10	0.6953	0.00024
6	2000	100	0.2	10	0.1007	0.00001
7	3000	100	0.8	10	0.6983	0.00018
8	3000	100	0.2	10	0.0969	0.00003
9	50	200	0.8	10	0.6869	0.00056
10	50	200	0.2	10	0.1346	0.00067
11	50	300	0.8	10	0.6395	0.00016
12	50	300	0.2	10	0.1064	0.00031
13	300	500	0.8	10	0.4371	0.00002
14	300	500	0.2	10	0.0870	0.00002
15	100	100	0.8	10	0.7196	0.00012
16	100	100	0.2	10	0.1508	0.00080
17	300	300	0.8	10	0.5175	0.00003
18	300	300	0.2	10	0.0976	0.00006
19	2000	100	0.8	30	0.8472	0.00003
20	2000	100	0.2	30	0.1804	0.00003

NG: Número de Genótipos; NA: Número de Ambientes; r: Correlação entre Ambientes; NPC: Número de Componentes Principais; σ^2 : variância.

Nos cenários em que foi testada uma situação de número igual de genótipos e ambientes (cenários 15, 16, 17 e 18), os resultados também indicam uma forte influência da correlação entre ambientes na precisão das predições. Quando a correlação foi de 0.2, ocorreu uma grande redução da acurácia preditiva, tanto para os ambientes eliminados quanto para os ambientes não eliminados. O mesmo ocorreu quando se passou de 100 para 300 genótipos e ambientes. Porém, a precisão obtida no cenário 15, com correlação entre ambientes de 0.8, foi de alta magnitude.

Por fim, ao se aumentar o número de componentes principais (5, 6, 19 e 20), a maior explicação da interação levou à melhores predições. As correlações obtidas com 30 componentes principais no cenário 19 foram de alta magnitude, em especial as obtidas para os ambientes não eliminados.

Desbalanceamento Aleatório

Os resultados obtidos ao se realizar a retirada de ambientes de maneira aleatória seguiram as mesmas tendências das análises anteriores, independentemente do nível de desbalanceamento avaliado (Tabela 8 e Tabela S2 do Material Suplementar). Em geral, a acurácia preditiva foi maior nos cenários em que a correlação entre ambientes foi de 0.8 e para as predições de ambientes não eliminados. Também houve pouca redução nos valores das correlações obtidas quando se testou o aumento do desbalanceamento simulado de 5 para 30% de ambientes eliminados.

Novamente, o aumento simultâneo do número de genótipos e de ambientes levou à uma redução da acurácia preditiva (cenários 13 e 14). Para os cenários em que foram testados números iguais de genótipos e de ambientes, novamente ficou claro a maior eficiência da predição do cenário 15 em todos os níveis de desbalanceamento testados. Ao se aumentar o número de genótipos e de ambientes, a precisão é reduzida (cenário 17). O mesmo ocorre quando se considera os ambientes com correlação de 0.2 ou os ambientes eliminados.

O aumento do número de componentes principais levou à uma grande melhora das predições (cenários 5, 6, 19 e 20). As correlações obtidas foram de alta magnitude até mesmo nos cenários de correlação entre ambientes de 0.2 para as predições de ambientes não eliminados. Para as predições dos ambientes eliminados, apesar da menor precisão das predições, os valores ainda foram altos para os ambientes com correlação de 0.8. Nas situações de ambientes pouco correlacionados, apesar de apresentarem baixa precisão, as correlações praticamente dobraram ao se aumentar o número de componentes principais de 10 para 30.

Tabela 8. Médias e variâncias da correlação média obtida entre valores simulados e preditos dos genótipos em cada ambiente em cada uma das 1000 amostragens realizadas. Os resultados são referentes aos ambientes eliminados considerando uma retirada de ambientes realizada de maneira aleatória repetida 1000 vezes, nos níveis de 5, 10, 20 e 30 % de desbalanceamento.

Cenário	NG	NA	r	NPC	5% de Perdas		10% de Perdas		20% de Perdas		30% de Perdas	
					Correlação Média		Correlação Média		Correlação Média		Correlação Média	
					Média	σ^2	Média	σ^2	Média	σ^2	Média	σ^2
1	50	100	0.8	10	0.7483	0.00210	0.7432	0.00126	0.7338	0.00073	0.7236	0.00055
2	50	100	0.2	10	0.1825	0.00295	0.1728	0.00158	0.1583	0.00087	0.1451	0.00058
3	1000	100	0.8	10	0.6847	0.00064	0.6833	0.00048	0.6789	0.00044	0.6730	0.00047
4	1000	100	0.2	10	0.0940	0.00029	0.0905	0.00017	0.0896	0.00011	0.0880	0.00009
5	2000	100	0.8	10	0.6981	0.00054	0.6947	0.00039	0.6887	0.00035	0.6811	0.00037
6	2000	100	0.2	10	0.0913	0.00020	0.0880	0.00012	0.0835	0.00009	0.0806	0.00007
7	3000	100	0.8	10	0.7006	0.00058	0.6970	0.00046	0.6912	0.00034	0.6846	0.00041
8	3000	100	0.2	10	0.0855	0.00014	0.0835	0.00009	0.0808	0.00008	0.0787	0.00007
9	50	200	0.8	10	0.6684	0.00084	0.6649	0.00046	0.6572	0.00033	0.6474	0.00028
10	50	200	0.2	10	0.1254	0.00173	0.1244	0.00085	0.1139	0.00045	0.1040	0.00029
11	50	300	0.8	10	0.6207	0.00066	0.6181	0.00033	0.6109	0.00021	0.6017	0.00016
12	50	300	0.2	10	0.1026	0.00117	0.0958	0.00050	0.0872	0.00028	0.0778	0.00019
13	300	500	0.8	10	0.3920	0.00034	0.3905	0.00019	0.3869	0.00012	0.3814	0.00011
14	300	500	0.2	10	0.0842	0.00014	0.0830	0.00007	0.0778	0.00004	0.0722	0.00002
15	100	100	0.8	10	0.7124	0.00086	0.7088	0.00058	0.7035	0.00045	0.6984	0.00043
16	100	100	0.2	10	0.1428	0.00178	0.1404	0.00086	0.1315	0.00046	0.1225	0.00036
17	300	300	0.8	10	0.4640	0.00063	0.4652	0.00034	0.4607	0.00026	0.4550	0.00023
18	300	300	0.2	10	0.0921	0.00025	0.0902	0.00012	0.0852	0.00006	0.0798	0.00004
19	2000	100	0.8	30	0.8432	0.00017	0.8396	0.00016	0.8324	0.00017	0.8213	0.00025
20	2000	100	0.2	30	0.1685	0.00030	0.1661	0.00022	0.1648	0.00018	0.1619	0.00016

NG: Número de Genótipos; NA: Número de Ambientes; r: Correlação entre Ambientes; NPC: Número de Componentes Principais; σ^2 : variância.

Capacidade preditiva de GEI

EM-AMMI Funcional

A acurácia preditiva obtida pelo modelo EM-AMMI Funcional ao se realizar o desbalanceamento ao acaso de genótipos está descrita na Tabela 9. Os gráficos que representam o padrão de interação e o comportamento genotípico ao longo dos ambientes estão representados nas Figuras 7 a 18. Devido ao grande número de gráficos gerados pela análise, foram selecionados para ilustrar o método spline os gráficos dos genótipos G1, G2, G3 e G4 do primeiro fold de cada porcentagem de desbalanceamento dos cenários 1, 5 e 19. Esses cenários foram selecionados por representarem, respectivamente, uma situação de maior número de ambientes do que de genótipos; maior número de genótipos do que de ambientes; e maior número de genótipos do que de ambientes, porém com 30 componentes principais retidos pelo modelo. Assim como em todas as análises realizadas, os cenários em que a correlação entre

ambientes foi de 0.2 resultou em menor qualidade das predições. Em geral, para todos os cenários, o aumento do nível de desbalanceamento pouco afetou a acurácia preditiva.

Os resultados novamente seguem a mesma tendência das análises anteriores. A qualidade das predições quando ocorre perda de alguns genótipos em alguns ambientes da rede de ensaios também é influenciada principalmente pela correlação entre ambientes. O aumento do número de genótipos ou de ambientes pouco altera a acurácia preditiva do modelo. Porém, assim como nos resultados das análises anteriores, o aumento do número de genótipos e ambientes para 300 e 500, respectivamente, levou à uma redução da precisão (cenários 13 e 14).

Os cenários em que foram testados os mesmos números de genótipos e de ambientes também apresentaram os mesmos comportamentos das análises anteriores. O aumento de 100 para 300 levou à uma redução da qualidade das predições. Porém, novamente, o cenário 15 apresentou alta acurácia preditiva.

O aumento do número de componentes principais mais uma vez se mostrou eficiente (cenários 5, 6, 19 e 20). A acurácia preditiva teve grande aumento ao se passar de 10 para 30 componentes principais retidos.

Tabela 9. Médias das correlações entre valores simulados e preditos dos genótipos ao se realizar a retirada ao acaso de alguns genótipos em alguns ambientes para 4 níveis de desbalanceamento de dados.

Cenário	NG	NA	r	NPC	Desbalanceamento			
					5%	10%	20%	30%
1	50	100	0.8	10	0.8036	0.8000	0.7906	0.7692
2	50	100	0.2	10	0.5150	0.5066	0.4849	0.4142
3	1000	100	0.8	10	0.7278	0.7257	0.7204	0.7101
4	1000	100	0.2	10	0.3969	0.3884	0.3690	0.3382
5	2000	100	0.8	10	0.7282	0.7262	0.7213	0.7119
6	2000	100	0.2	10	0.3854	0.3765	0.3585	0.3284
7	3000	100	0.8	10	0.7295	0.7276	0.7227	0.7132
8	3000	100	0.2	10	0.3776	0.3693	0.3521	0.3229
9	50	200	0.8	10	0.7166	0.7125	0.7037	0.6839
10	50	200	0.2	10	0.3942	0.3910	0.3731	0.3850
11	50	300	0.8	10	0.6636	0.6595	0.6498	0.6315
12	50	300	0.2	10	0.3490	0.3498	0.3523	0.3330
13	300	500	0.8	10	0.4564	0.4543	0.4495	0.4408
14	300	500	0.2	10	0.2688	0.2643	0.2569	0.2428
15	100	100	0.8	10	0.7600	0.7566	0.7497	0.7330
16	100	100	0.2	10	0.4692	0.4597	0.4432	0.3951
17	300	300	0.8	10	0.5355	0.5331	0.5278	0.5187
18	300	300	0.2	10	0.3144	0.3096	0.2956	0.2746
19	2000	100	0.8	30	0.9116	0.9078	0.8965	0.8596
20	2000	100	0.2	30	0.6356	0.6169	0.5725	0.4981

NG: Número de Genótipos; NA: Número de Ambientes; r: Correlação entre Ambientes; NPC: Número de Componentes Principais.

As Figuras 7 a 18 representam o comportamento dos genótipos avaliados em diferentes cenários e níveis de desbalanceamento. A linha preta representa os valores simulados dos efeitos da interação de cada genótipo ao longo dos ambientes. Essa linha permite avaliar como se comporta o padrão de interação ao se variar a latitude dos ambientes. A linha vermelha apresenta as mesmas interpretações, porém representa o comportamento das previsões dos efeitos da interação de cada genótipo ao longo dos ambientes. Os pontos azuis representam os pontos de desbalanceamento, ou seja, onde foi feita a previsão do efeito de GEI. Em todos os cenários avaliados, é possível observar que o efeito da interação segue um padrão de comportamento de acordo com a latitude do ambiente, o qual foi captado pelo método.

Para todos os cenários foi possível obter a curva suavizada do efeito de GEI. As correlações obtidas entre valores simulados e preditos (Tabela 9) mostram a alta capacidade preditiva do método em realizar as previsões em situações de desbalanceamento e a eficiência dos gráficos de curvas suavizadas em captar o padrão de comportamento e a influência da interação genótipos por ambientes.

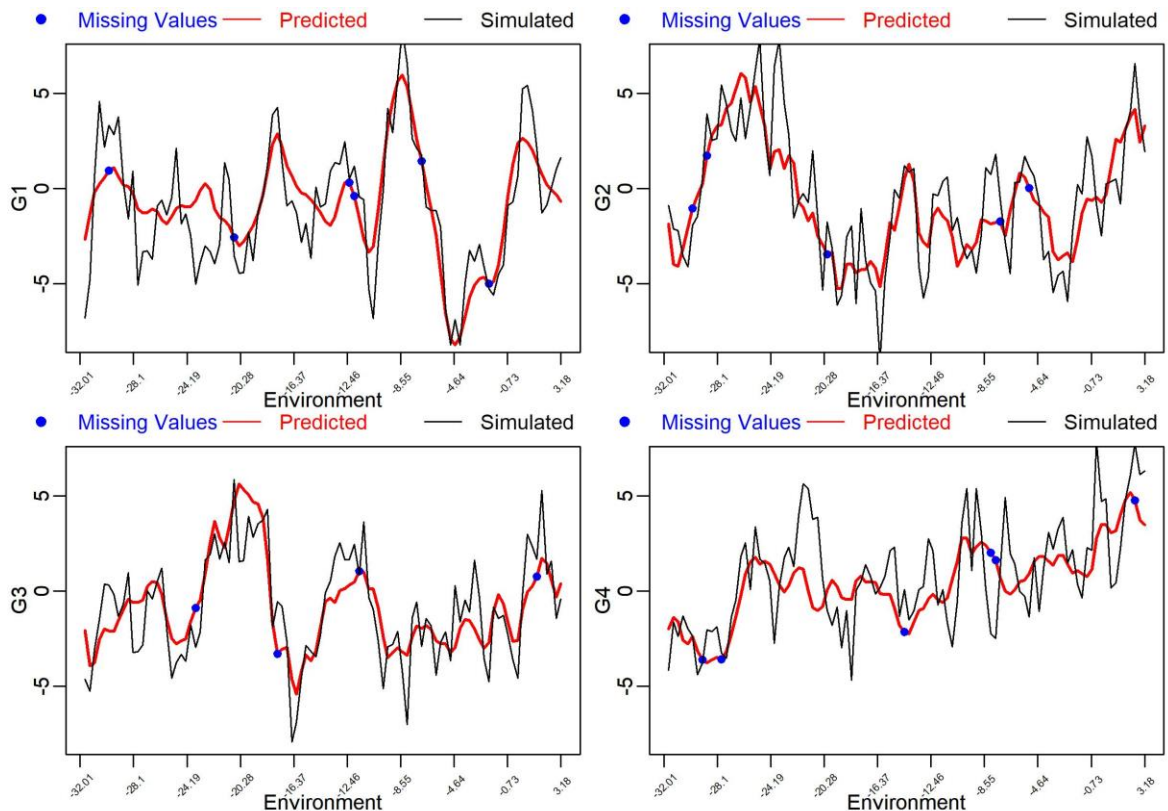


Figura 7. Padrão de interação dos genótipos G1, G2, G3 e G4 ao se variar a latitude dos ambientes, referente ao primeiro fold da análise considerando um desbalanceamento de 5% dos dados do cenário 1.

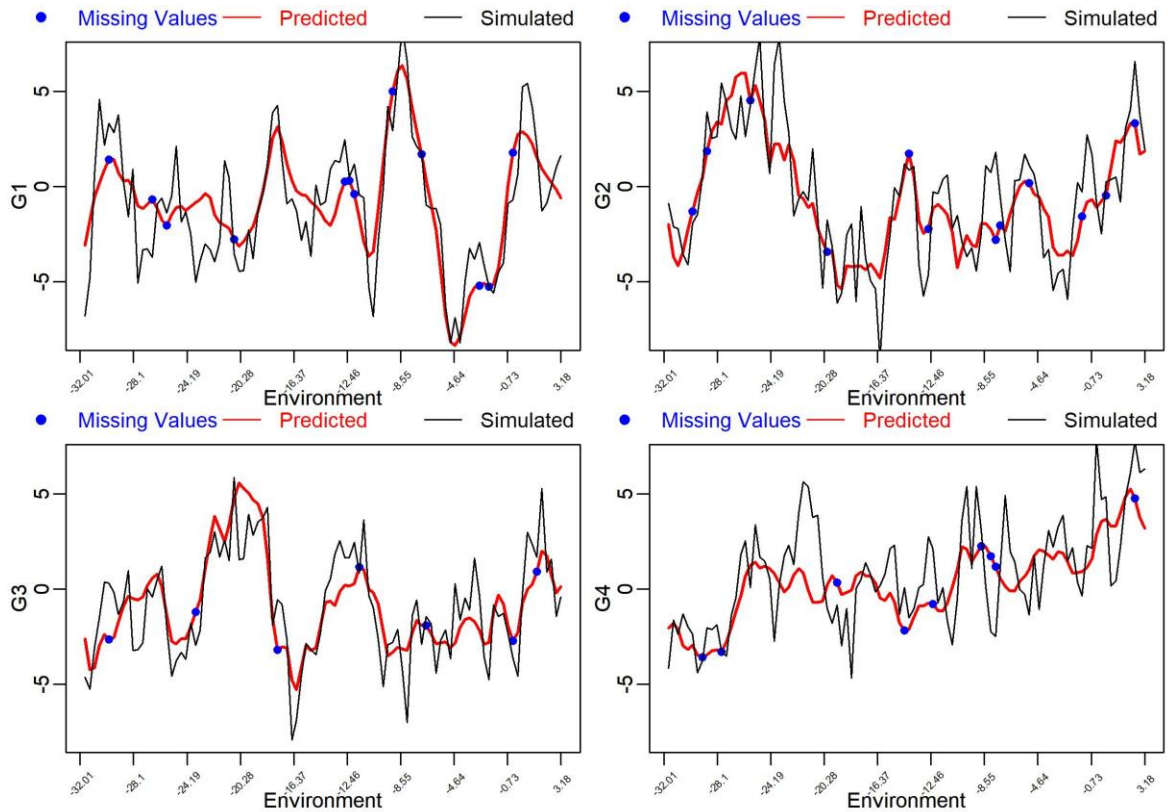


Figura 8. Padrão de interação dos genótipos G1, G2, G3 e G4 ao se variar a latitude dos ambientes, referente ao primeiro fold da análise considerando um desbalanceamento de 10% dos dados do cenário 1.

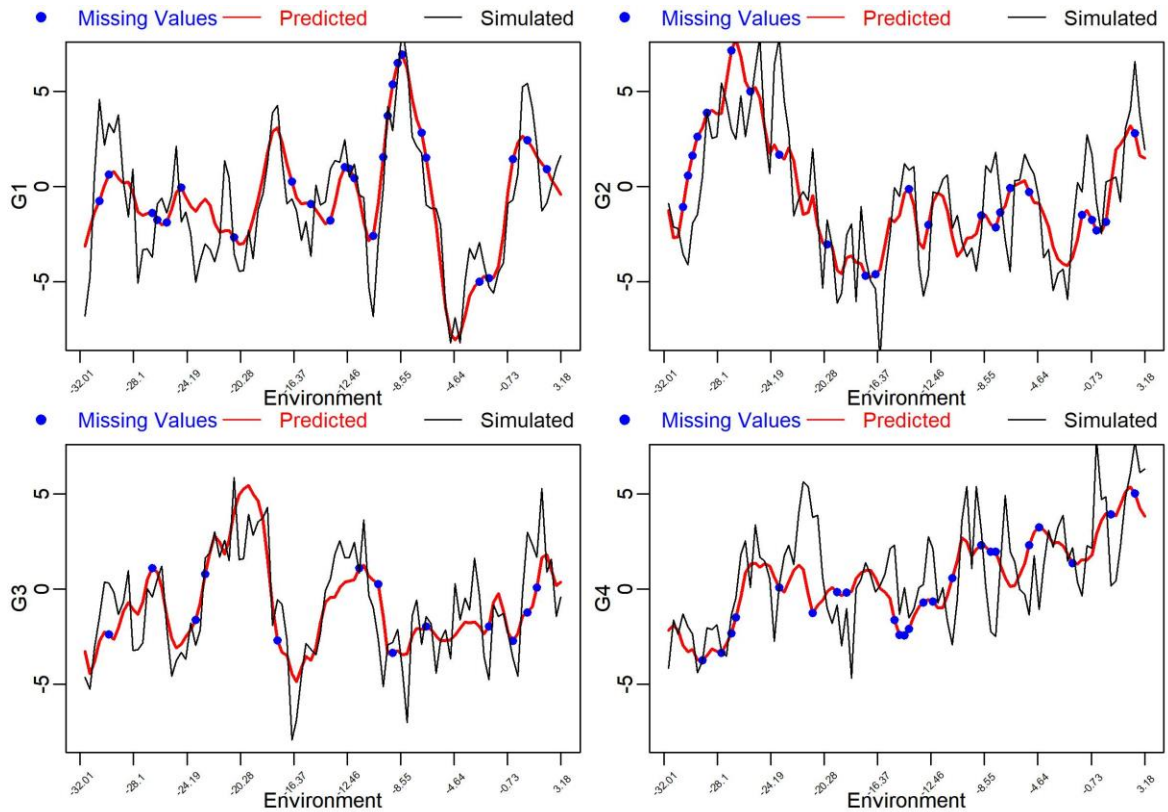


Figura 9. Padrão de interação dos genótipos G1, G2, G3 e G4 ao se variar a latitude dos ambientes, referente ao primeiro fold da análise considerando um desbalanceamento de 20% dos dados do cenário 1.

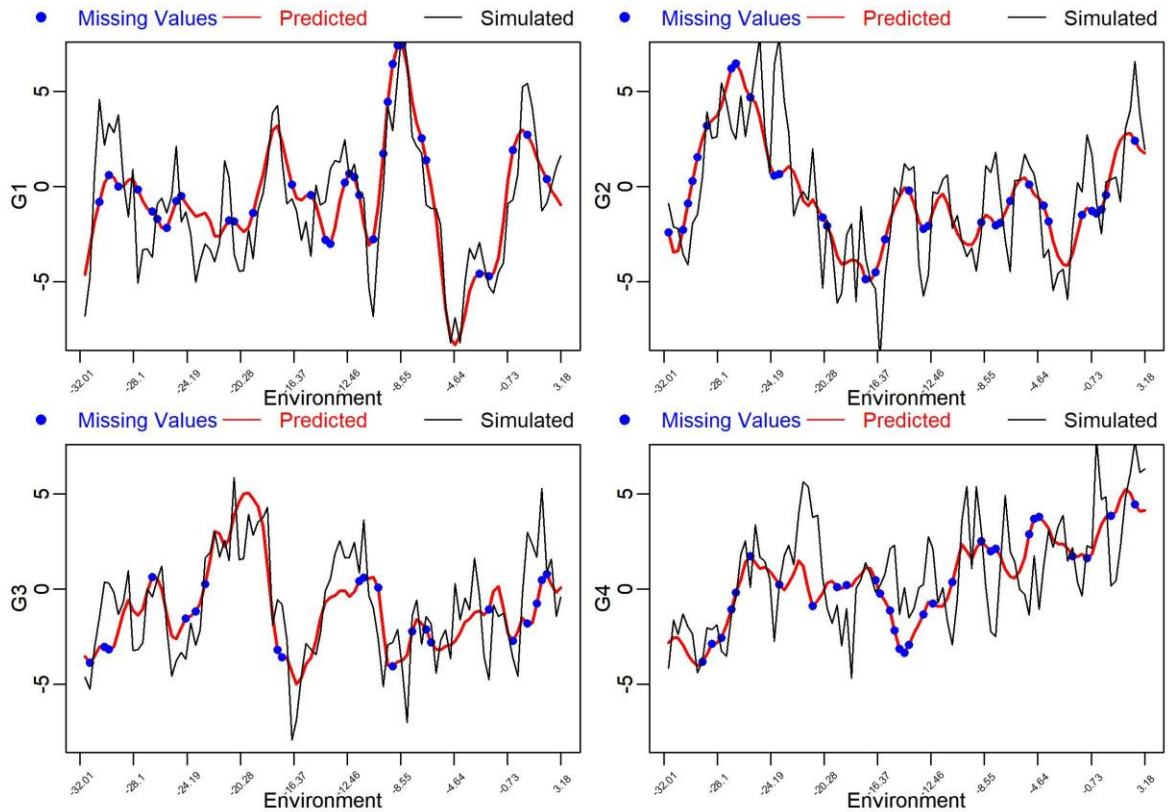


Figura 10. Padrão de interação dos genótipos G1, G2, G3 e G4 ao se variar a latitude dos ambientes, referente ao primeiro fold da análise considerando um desbalanceamento de 30% dos dados do cenário 1.

Para todos os níveis de desbalanceamento, percebe-se que ocorre um aumento ou redução do efeito da interação conforme se varia as coordenadas de latitude. Por exemplo, para o genótipo G1, fica claro o aumento do efeito da interação após aproximadamente a latitude de -10.00, com declínio próximo à latitude de -4.64 (Figuras 7 a 10). As curvas suavizadas vermelhas, referentes aos valores preditos, captaram esse padrão de comportamento e acompanharam as curvas dos valores simulados. Isto pode ser observado para todos os genótipos em todos os níveis de desbalanceamento em todos os cenários (Figuras 7 a 18).

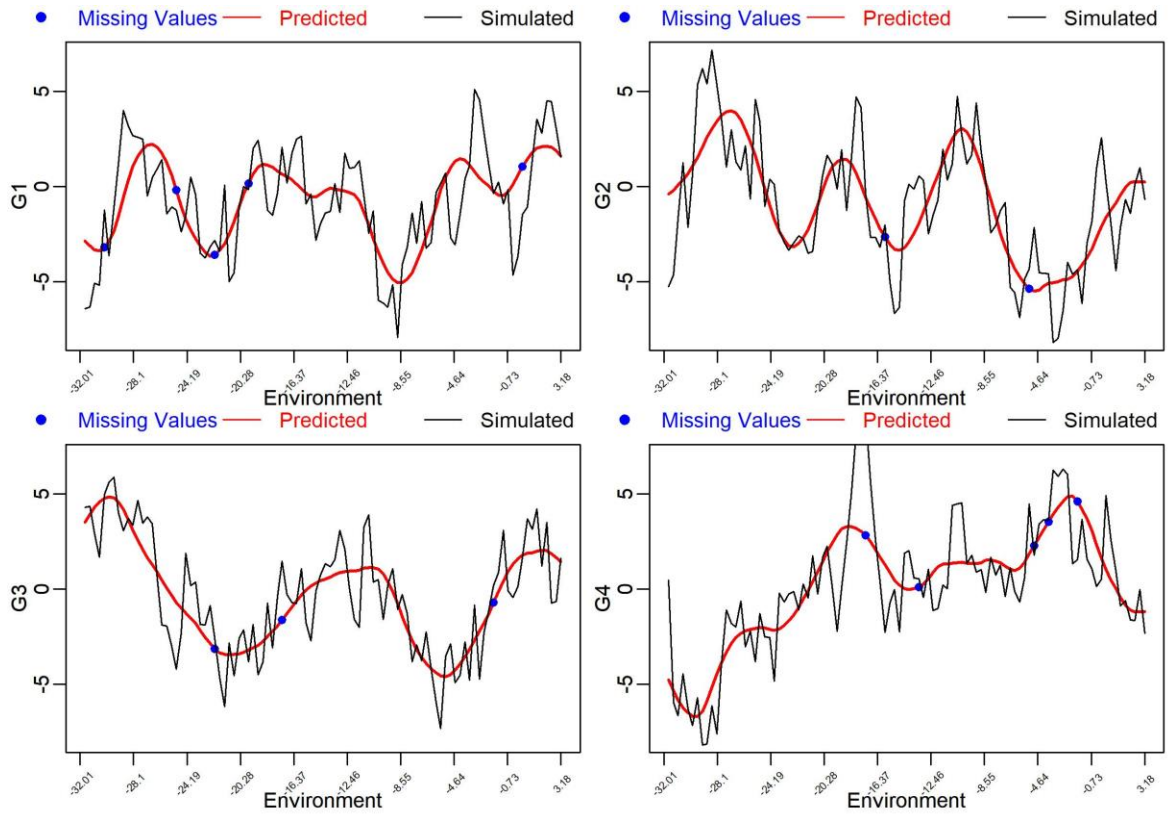


Figura 11. Padrão de interação dos genótipos G1, G2, G3 e G4 ao se variar a latitude dos ambientes, referente ao primeiro fold da análise considerando um desbalanceamento de 5% dos dados do cenário 5.

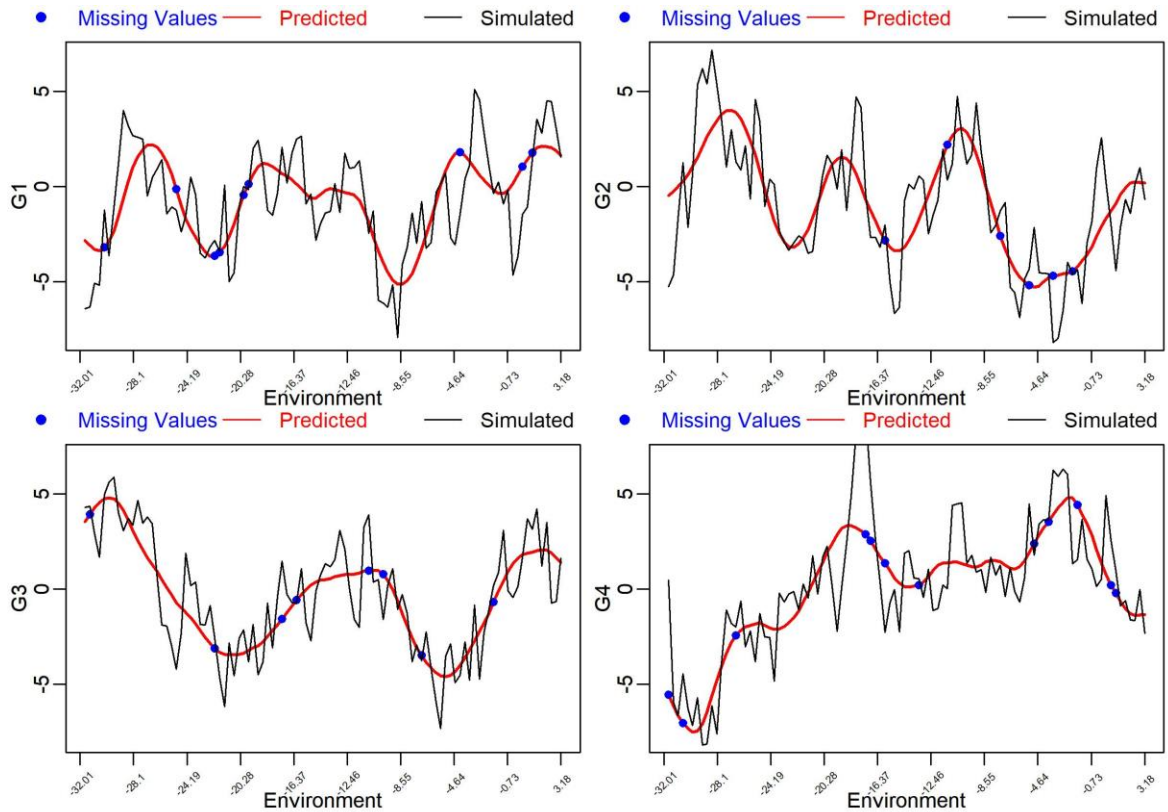


Figura 12. Padrão de interação dos genótipos G1, G2, G3 e G4 ao se variar a latitude dos ambientes, referente ao primeiro fold da análise considerando um desbalanceamento de 10% dos dados do cenário 5.

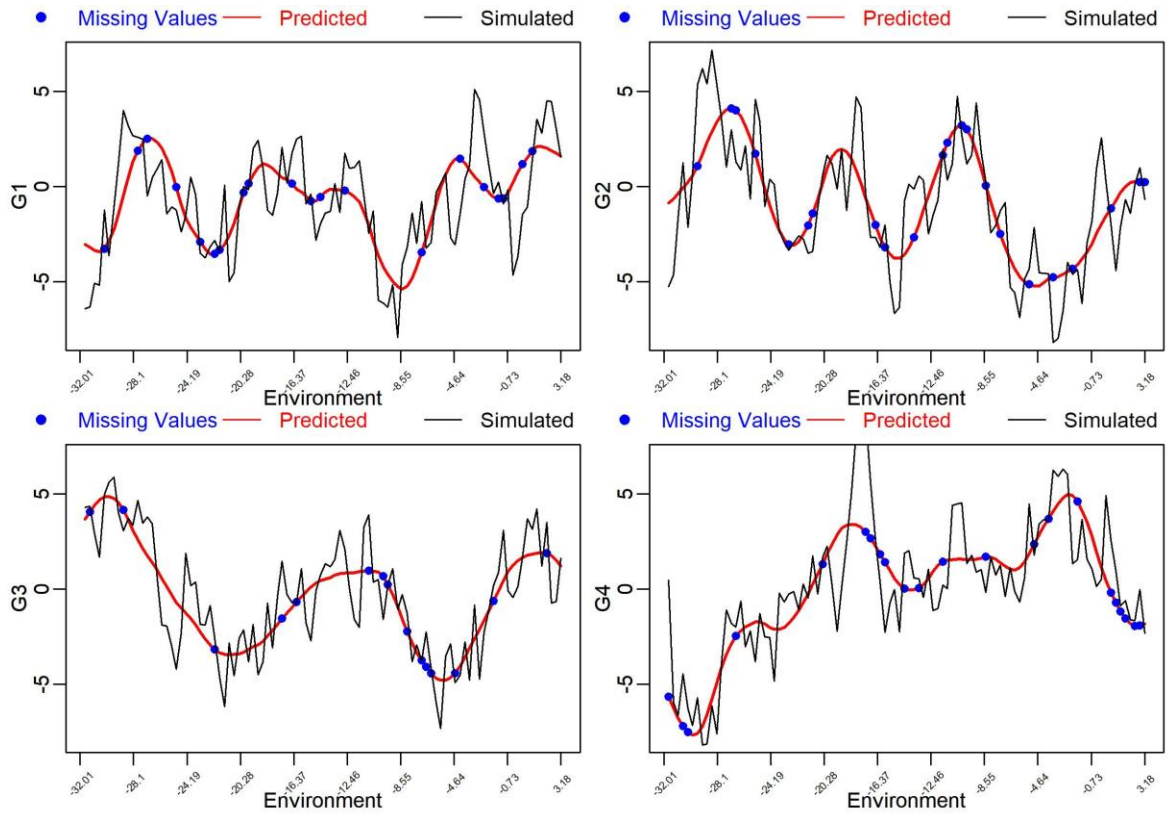


Figura 13. Padrão de interação dos genótipos G1, G2, G3 e G4 ao se variar a latitude dos ambientes, referente ao primeiro fold da análise considerando um desbalanceamento de 20% dos dados do cenário 5.

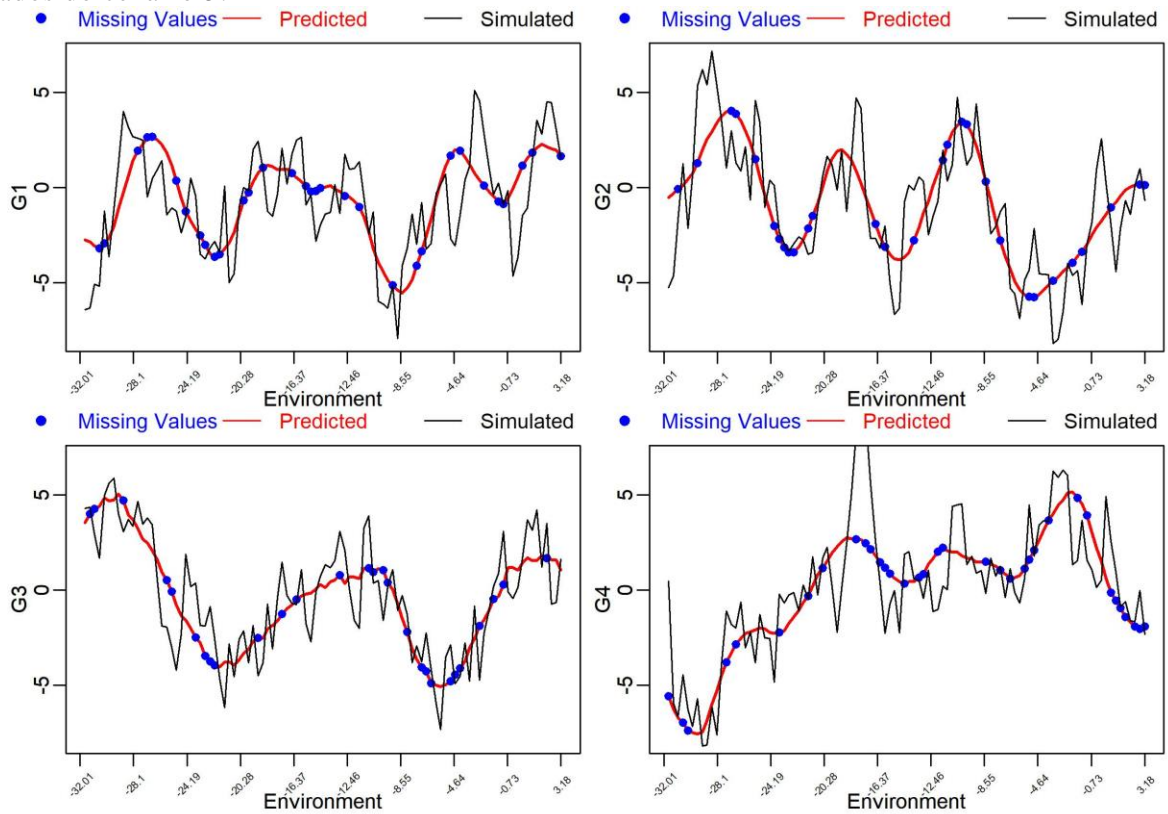


Figura 14. Padrão de interação dos genótipos G1, G2, G3 e G4 ao se variar a latitude dos ambientes, referente ao primeiro fold da análise considerando um desbalanceamento de 30% dos dados do cenário 5.

As linhas que representam as previsões do cenário 1 (Figuras 7 a 10) apresentaram um melhor ajuste do que as mesmas linhas para o cenário 5 (Figuras 11 a 14). Para o cenário 5, onde ocorrem maior número de genótipos, o ajustamento não é tão perfeito. Isto pode ser confirmado pelos valores das correlações obtidas entre valores simulados e preditos (Tabela 9). Porém, fica claro que as curvas referentes à valores preditos foi eficiente em captar o padrão de interação novamente. Usando o genótipo G1 como exemplo novamente, verifica-se que a curva suavizada vermelha acompanhou todos os aumentos ou reduções dos efeitos da interação (Figuras 11 a 14). É possível verificar quatro picos de valores máximos dos efeitos da interação ao longo do intervalo de coordenadas de latitude, independentemente do nível de desbalanceamento testado. O primeiro está próximo à latitude de -28.1, o segundo próximo à latitude -18.00, o terceiro próximo à latitude -4.64, e o quarto próximo à latitude 1.00.

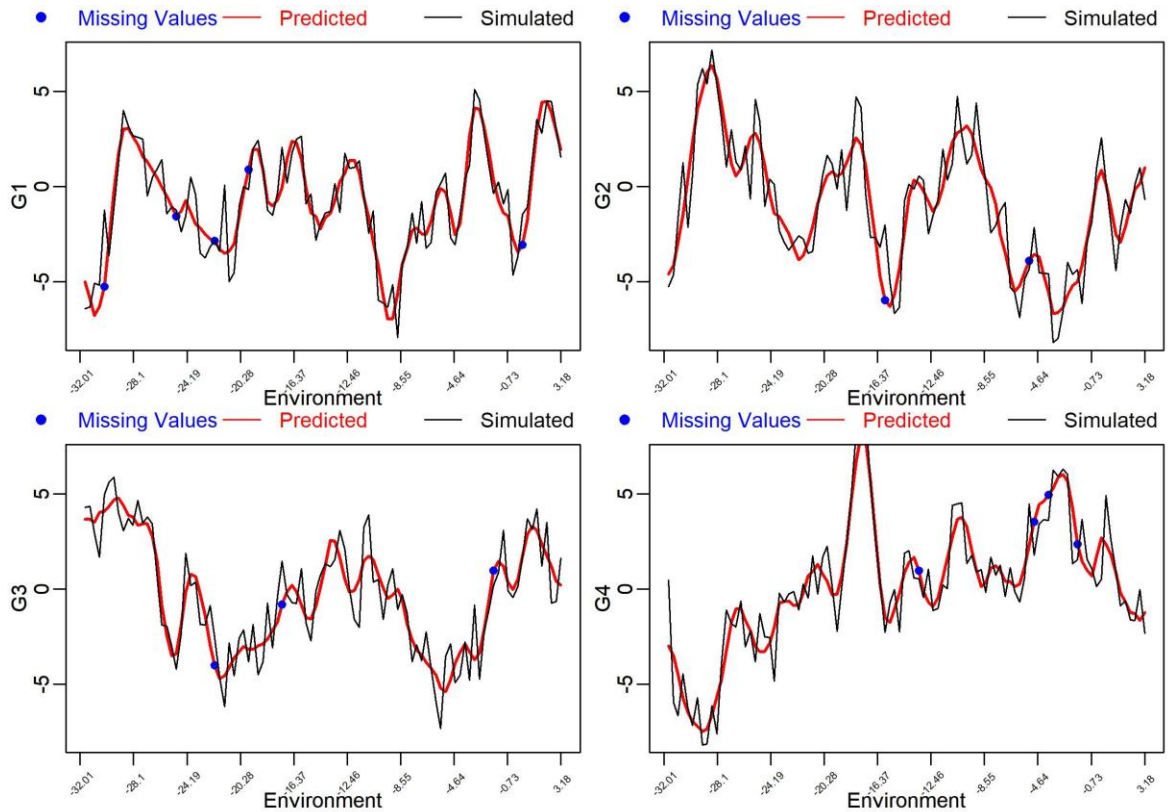


Figura 15. Padrão de interação dos genótipos G1, G2, G3 e G4 ao se variar a latitude dos ambientes, referente ao primeiro fold da análise considerando um desbalanceamento de 5% dos dados do cenário 19.

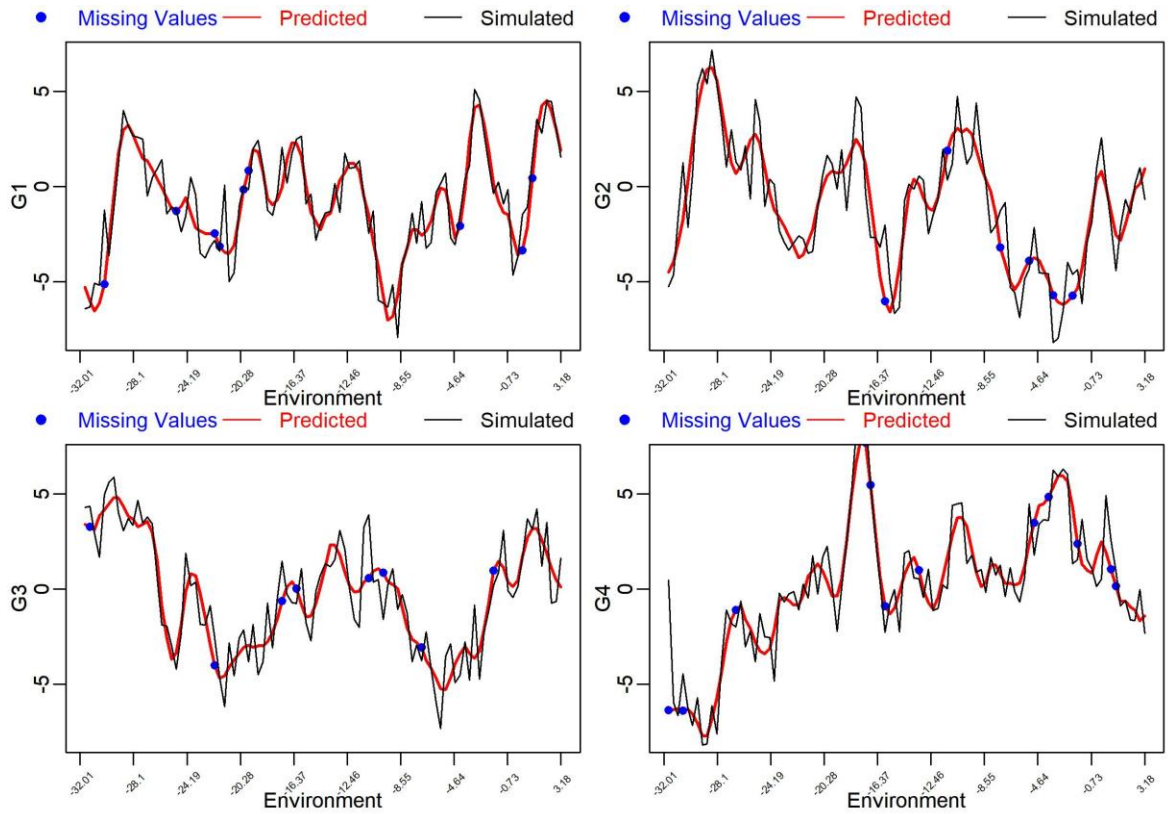


Figura 16. Padrão de interação dos genótipos G1, G2, G3 e G4 ao se variar a latitude dos ambientes, referente ao primeiro fold da análise considerando um desbalanceamento de 10% dos dados do cenário 19.

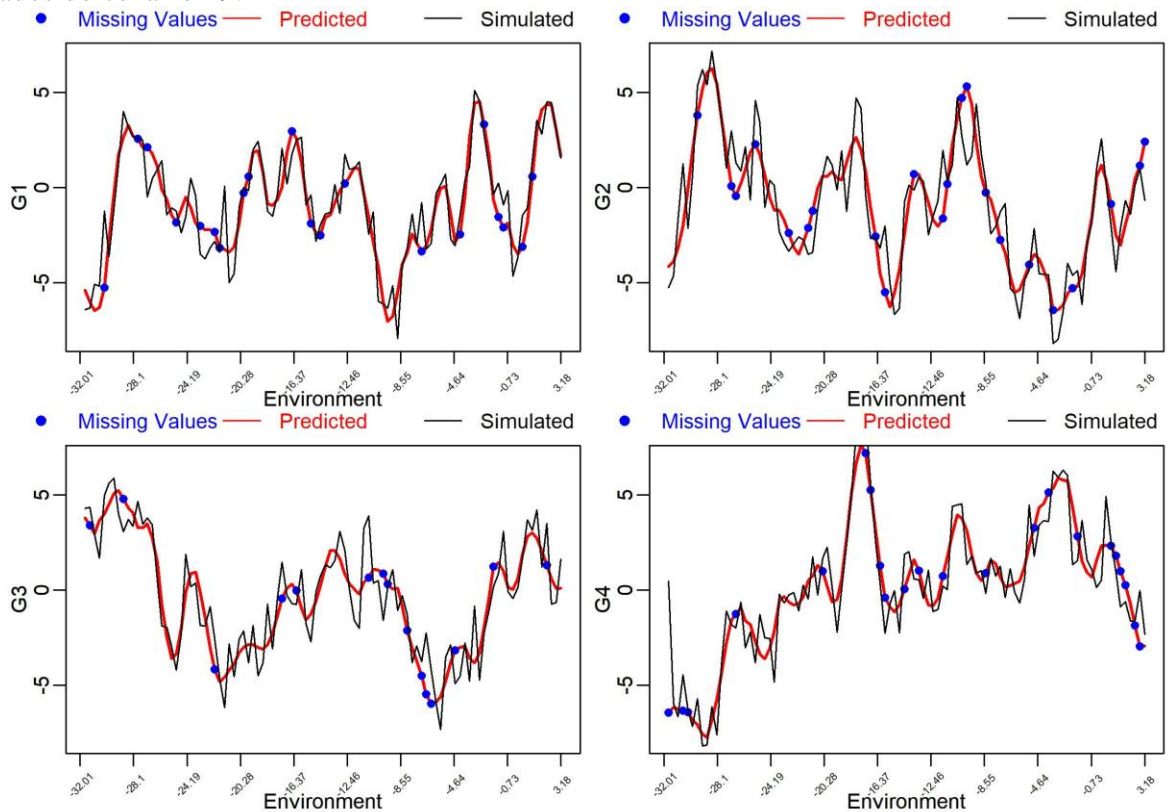


Figura 17. Padrão de interação dos genótipos G1, G2, G3 e G4 ao se variar a latitude dos ambientes, referente ao primeiro fold da análise considerando um desbalanceamento de 20% dos dados do cenário 19.

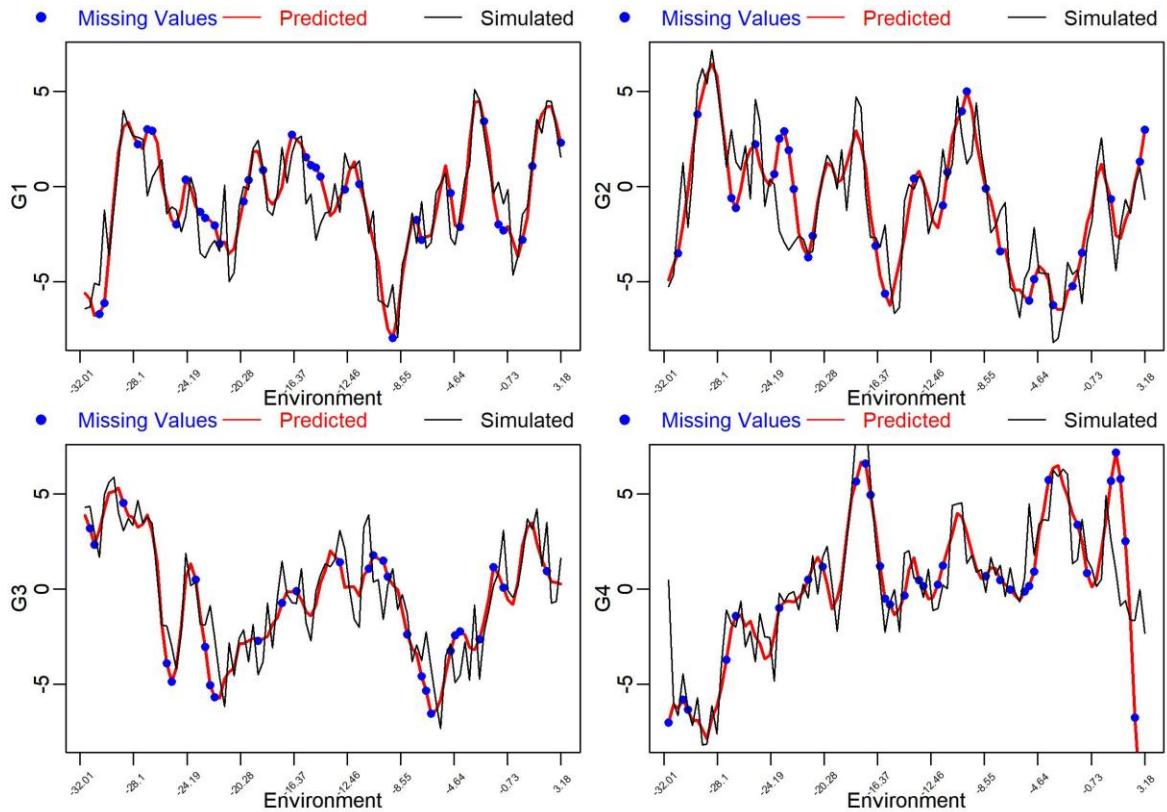


Figura 18. Padrão de interação dos genótipos G1, G2, G3 e G4 ao se variar a latitude dos ambientes, referente ao primeiro fold da análise considerando um desbalanceamento de 30% dos dados do cenário 19.

Com o aumento do número de componentes principais para 30, verifica-se um ajustamento quase perfeito das curvas plotadas para todos os níveis de desbalanceamento (Figuras 15 a 18). É possível verificar que as curvas vermelhas são quase idênticas às curvas pretas. O maior número de componentes principais permite obter uma maior explicação do padrão de interação e, portanto, um melhor ajustamento das curvas de valores preditos.

O método EM-AMMI Funcional também apresentou alta capacidade preditiva ao lidar com o desbalanceamento de ambientes inteiros (Tabela 10). As mesmas tendências verificadas nas análises anteriores são observadas aqui. O aumento do número de genótipos de 50 para 3000 praticamente não teve impacto na qualidade das predições nos cenários com correlação entre ambientes de 0.8 (cenários 1, 3, 5 e 7). Porém, quando a correlação entre ambientes de 0.2 foi testada, houve grande redução da qualidade das predições (cenários 2, 4, 6 e 8). Resultados semelhantes foram obtidos com o aumento do número de ambientes (cenários 1, 2, 9, 10, 11 e 12). No caso dos cenários 13 e 14, novamente o aumento simultâneo de genótipos e ambientes, em situação de maior número de ambientes, levou à uma menor precisão das predições.

Quando foram testados os cenários com números iguais de genótipos e de ambientes, mais uma vez o cenário 15 se destacou com maior precisão da predição. Por fim, o aumento do número de componentes principais permitiu um aumento da acurácia preditiva, com destaque para os cenários 6 e 20, em que o valor da correlação praticamente dobrou.

Tabela 10. Médias e variâncias da correlação média obtidas entre valores simulados e preditos dos genótipos em cada ambiente em cada subconjunto de dados, considerando uma situação de desbalanceamento de 10% dos ambientes (*k*-fold de 10).

Cenário	NG	NA	r	NPC	Correlação Média	
					Média	σ^2
1	50	100	0.8	10	0.7481	0.0012
2	50	100	0.2	10	0.1788	0.0008
3	1000	100	0.8	10	0.6804	0.0004
4	1000	100	0.2	10	0.0896	0.0002
5	2000	100	0.8	10	0.6910	0.0002
6	2000	100	0.2	10	0.0880	0.0001
7	3000	100	0.8	10	0.7001	0.0002
8	3000	100	0.2	10	0.0838	0.0001
9	50	200	0.8	10	0.6230	0.0025
10	50	200	0.2	10	0.1417	0.0022
11	50	300	0.8	10	0.6346	0.0012
12	50	300	0.2	10	0.1148	0.0027
13	300	500	0.8	10	0.3233	0.0012
14	300	500	0.2	10	0.0616	0.0003
15	100	100	0.8	10	0.7185	0.0002
16	100	100	0.2	10	0.1349	0.0006
17	300	300	0.8	10	0.4515	0.0003
18	300	300	0.2	10	0.0863	0.0004
19	2000	100	0.8	30	0.8398	0.0001
20	2000	100	0.2	30	0.1620	0.0001

NG: Número de Genótipos; NA: Número de Ambientes; r: Correlação entre Ambientes; NPC: Número de Componentes Principais. σ^2 : variância média.

DISCUSSÃO

De modo geral, as análises realizadas levaram a conclusões semelhantes. Os cenários simulados com correlação entre ambientes de 0.8 apresentaram maiores correlações entre valores simulados e preditos em relação aos cenários simulados com correlação de 0.2, em todos os cenários e análises testados. Essa é uma situação esperada, e demanda conhecimento do melhorista para selecionar corretamente os locais de condução dos ensaios. Essa decisão está intimamente ligada à qualidade das predições e, conseqüentemente, com a correta tomada de decisão na seleção e recomendação de genótipos. As Figuras 1 a 6 evidenciam como a estrutura

de interação influencia a capacidade preditiva do modelo. As predições são baseadas no padrão e na porcentagem de explicação da interação pelos componentes principais, indicadas pelas curvas obtidas para cada gráfico. Os gráficos deixam claro como o modelo é capaz de captar o padrão de interação existente, até mesmo em uma situação de um grande conjunto de dados, no caso com 3000 genótipos. Portanto, o método é eficiente em lidar com um grande volume de genótipos, situação comum em ensaios multiambientais.

Modelos de estudo de GEI estão sujeitos à estrutura da matriz de covariâncias e, portanto, a correlação entre ambientes deve ser positiva e constante. Nesse caso, haverá um melhor ajuste do modelo (Lopez-Cruz et al., 2015; Acosta-Pech et al., 2017). Os resultados de Burgeño et al. (2012) evidenciam a importância de ser ter informação oriunda de ambientes correlacionados. As correlações obtidas pelos autores mostram que prever a performance de genótipos avaliados em locais diferentes, porém correlacionados, é mais simples do que, por exemplo, prever o comportamento de genótipos nunca testados em campo. Portanto, a escolha de ambientes correlacionados é uma alternativa para lidar com o desbalanceamento. No presente estudo, as altas correlações obtidas entre valores simulados e preditos nos cenários com correlação entre ambientes de 0.8 evidenciam a eficiência do método em prever o padrão de interação e o comportamento genotípico, até mesmo em situações com alto grau de desbalanceamento. Vale ressaltar que uma situação de correlação entre ambientes de 0.2 prejudica as predições de qualquer método preditivo. Portanto, considerando esse fator limitante, ainda sim a maioria das predições foram relativamente altas.

A análise de dados funcionais também foi eficiente quando foi testada sob situações de aumento do número de genótipos ou de ambientes. Os resultados demonstraram pouca redução dos valores das correlações obtidas quando se aumentou o número de genótipos para 3000 ou o número de ambientes para 300 (cenários 7, 8, 11 e 12). Esses cenários são bem representativos de situações reais, onde o melhorista trabalha com um grande número de genótipos em uma grande rede de ensaios. Além disso, essa tendência também é observada nos diferentes níveis de desbalanceamento. Esta também é uma situação comum em programas de melhoramento, os quais são normalmente desbalanceados devido à inclusão e descarte de genótipos ao longo dos ciclos de seleção ou devido a perdas à nível de campo ocasionadas por fatores bióticos ou abióticos (de Figueiredo et al., 2015). Novamente, isso evidencia a aplicabilidade do método em situações reais de programas de melhoramento. Por outro lado, os cenários 13 e 14 mostram que o aumento simultâneo de genótipos e de ambientes, em situação de maior número de ambientes do que de genótipos, reduz as correlações entre valores simulados e preditos.

Portanto, apesar do método ser eficiente nas mais diversas situações, o melhorista deve ter cautela ao se determinar o número de genótipos e de ambientes avaliados.

Nem sempre o melhorista vai ter à sua disposição ambientes altamente correlacionados. Muitas vezes, os genótipos são avaliados em condições completamente distintas. Portanto, a baixa correlação entre ambientes prejudicará a qualidade das predições obtidas. Os resultados deste trabalho demonstram como a inclusão de mais componentes principais pode ser viável para se aumentar as correlações entre valores simulados e preditos. Em todas as análises realizadas, o aumento do número de componentes principais de 10 para 30 levou à um grande aumento das correlações obtidas.

Uma limitação do método AMMI convencional é a dificuldade de se realizar a análise gráfica ao se reter muitos componentes principais. A análise de dados funcionais utilizada no presente estudo não faz uso de biplots, o que permite a utilização de um maior número de componentes principais sem prejudicar a análise visual dos gráficos. As Figuras 1 a 6 demonstram a necessidade de se utilizar um maior número de componentes principais para se obter melhores predições. Portanto, a FPCA fornece um novo enfoque para a análise gráfica realizada rotineiramente, onde são utilizados biplots com dois componentes principais para o estudo de GEI. Poderia ser argumentado que a retenção de muitos componentes principais teria como consequência uma alta captura de ruído. Porém, considerando o número de componentes principais possíveis, os modelos dos cenários simulados foram parcimoniosos. Os cenários que apresentaram menor número de componentes principais possíveis foram 1, 2, 9, 10, 11 e 12 em que, dos 49 possíveis, foram retidos apenas 10. Nos cenários 19 e 20, os únicos em que foram retidos 30 componentes principais, existiam 99 possíveis. Além disso, a retirada de ruído dos dados no método EM-AMMI Funcional ocorre duas vezes. A primeira ao se realizar a suavização através de splines, e a segunda através do próprio método AMMI ao se selecionar um determinado número de componentes principais a serem retidos. Portanto, o método realizou o descarte de grande parte do ruído em todos os cenários de todas as análises. Consequentemente, a análise gráfica do método proposto é superior em captar o padrão de interação e permite realizar melhores predições. Contudo, o presente estudo buscou apenas comparar as correlações entre valores simulados e preditos de cenários simulados com diferentes números de componentes principais retidos, sem a realização de testes estatísticos para determinação do número ideal de componentes a serem retidos pelo modelo. Em situações práticas, esses testes devem ser levados em consideração para que o modelo possa captar seletivamente sinal e descartar ruído de maneira adequada.

As correlações obtidas para ambientes não eliminados foram maiores do que as obtidas para ambientes eliminados, tanto para as análises onde foi realizado o desbalanceamento sistemático quanto para as análises em que foi realizado o desbalanceamento aleatório. Os ambientes eliminados têm as suas predições prejudicadas, e as correlações entre valores simulados e preditos são reduzidas. Em geral, os cenários simulados com correlação entre ambientes de 0.8 amenizaram este problema e apresentaram correlações entre valores simulados e preditos de alta magnitude. Mais uma vez, isto evidencia a importância de se realizar uma escolha criteriosa dos ambientes de condução dos ensaios. O uso de ambientes altamente correlacionados permite realizar a predição com alta precisão até mesmo de ambientes não testados. Dessa forma, é possível reduzir drasticamente o número de ambientes para condução de experimentos e, conseqüentemente, reduzir os custos do programa de melhoramento. Os resultados deste trabalho são promissores, e indicam a possibilidade de se realizar a seleção e recomendação de genótipos sem a necessidade de se realizar enormes séries de experimentos.

A abordagem funcional também foi muito eficiente ao lidar com altos níveis de desbalanceamento. Os resultados pouco se alteraram com o aumento da porcentagem de desbalanceamento. Nos cenários de correlação entre ambientes de 0.8, até mesmo quando foi realizado o desbalanceamento de ambientes inteiros as predições foram próximas dos valores simulados. Os métodos existentes para estudo da interação genótipos por ambientes não permitem lidar com tal grau de desbalanceamento. Por exemplo, a abordagem de modelos mistos usando uma estrutura de fatorial analítico (Piepho, 1998; Smith, Cullis, & Gilmour, 2001) é capaz de lidar com altos níveis de desbalanceamento, porém trata-se de um método complexo com alta demanda computacional (de Figueiredo et al., 2015). Por outro lado, o método EM-AMMI é simples e de fácil implementação. Sob uma abordagem funcional, o método foi eficiente em lidar com o desbalanceamento e foi capaz de realizar predições com altas correlações entre valores simulados e preditos.

Por meio da análise de dados funcionais, é possível captar informações que vão além de dados discretos que representam genótipos e ambientes. Os dados se tornam espaciais, e permitem captar o padrão de interação e o comportamento dos genótipos ao longo dos ambientes. Isto fica claro nos gráficos representados nas Figuras 7 a 18. A obtenção de curvas suavizadas é eficiente em captar o padrão de interação e comportamento do desempenho genotípico. Ao ordenar os ambientes, foi possível identificar tendências de comportamento ao se aumentar gradativamente as coordenadas de latitude presentes no eixo x. A visualização dos ambientes em forma de coordenadas e a identificação do padrão de interação podem ser aplicados em diversas variáveis, como por exemplo latitude e longitude. Pode-se utilizar

coordenadas geográficas para estudar as interações genótipos x latitude e genótipos x longitude. Dessa forma, pode-se avaliar se existe um padrão de mudança de comportamento com o aumento ou redução destas coordenadas. Em um país com dimensões continentais como o Brasil, onde ocorrem os mais variados climas ao longo de suas regiões, este tipo de medida se torna uma estratégia interessante. Isso abre a possibilidade de uma nova perspectiva na seleção e recomendação de cultivares. No Brasil, é comum os programas de melhoramento de milho realizarem a estratificação ambiental por meio de regiões geográficas. Dessa forma, os programas podem ser separados e conduzidos em regiões tropicais (Sudeste e Centro-Oeste) e subtropicais (Sul) (de Figueiredo et al., 2015). Ao se observar esta estratégia do ponto de vista funcional, é possível especificar ainda mais esta estratificação e realizar e recomendação de maneira mais precisa. Dentro dessas regiões podem haver diversas diferenças devido a interações genótipos x latitude e genótipos x longitude, as quais podem ser mais exploradas pelo método EM-AMMI Funcional. Isso torna o método uma ferramenta extremamente promissora para o melhorista.

O método spline também permite inferir a respeito da estabilidade genotípica através do ajustamento da curva de valores preditos à curva de valores simulados. Essa é a estabilidade do tipo 3, a qual considera um genótipo estável quando o quadrado médio dos desvios de regressão é de pequena magnitude (Lin, Binns, & Lefkovitch, 1986). Métodos de regressão como o de Eberhart e Russel (1966) utilizam um índice ambiental calculado a partir de dados que apresentam grande variação entre anos e entre safras, ou seja, que são não previsíveis. Por exemplo, ao se usar dados de precipitação, uma safra em que houve uma grande incidência de chuvas irá fornecer resultados completamente diferentes de uma safra em que as chuvas foram escassas. Portanto, essa variação desmonta todo o padrão de interação identificado em determinado ano ou safra. O método EM-AMMI Funcional é uma adaptação de métodos de regressão como os de Eberhart e Russel (1966) e de Toler e Burrows (1998), os quais realizam um ajuste de segundo ou terceiro grau. Porém, o ajuste realizado pelo método EM-AMMI Funcional permite a suavização das curvas por meio de splines, ou seja, o PCA Funcional. Dessa forma, o método permite prever GEI e medir a qualidade das curvas via desbalanceamento.

O método proposto no presente estudo usa dados de coordenadas de latitude, a qual possui amplitude definida e constante. A ordenação dos genótipos de acordo com suas coordenadas de latitude permite criar uma relação funcional entre ambientes. Além disso, ambientes com coordenadas semelhantes tendem a ser altamente correlacionados. Isto pode ser verificado nos gráficos do genótipo G4 testado no cenário 5 (Figuras 11 a 14). As curvas de valores simulados

apresentam grande variação entre aproximadamente as coordenadas -12.00 e -5.00, indicada pelos picos erráticos de efeitos da interação. Porém, a curva de valores preditos mostra um comportamento estável ao longo destes ambientes, indicado pela estabilização da curva em forma próxima à uma reta. Portanto, ao se agrupar os ambientes em características comuns, no caso valores de latitude semelhantes, é possível verificar que o comportamento do genótipo G4 na verdade é estável ao longo destes ambientes. Portanto, nesse grupo de ambientes, a resposta do genótipo G4 é sempre positiva. O método proposto neste estudo permite uma reavaliação do conceito de estabilidade. A abordagem é semelhante ao agrupamento em mega ambientes, ou seja, o agrupamento de ambientes onde os efeitos da interação são nulos e que os genótipos de melhor desempenho são os mesmos (Yan et al., 2002; Yan e Rajcan, 2002). Ao se realizar a análise EM-AMMI Funcional para agrupar ambientes, os componentes principais retidos pelo modelo isolam os efeitos onde um determinado grupo de genótipos apresentam efeito nulo da interação. Portanto, ao se captar padrão de interação, aquilo que era considerado instável devido a um comportamento errático, indicado pelos picos de efeitos da interação, pode ser na verdade um genótipo de comportamento estável. O padrão de interação pode ser transformado em previsibilidade baseado em dados imutáveis através da suavização da curva de efeitos de GEI.

Vale também destacar que os métodos SSVD (Xiao et al., 2014) e EM-AMMI Funcional foram capazes de captar uma relação funcional entre genótipos e ambientes em um conjunto de dados não necessariamente funcional. A simulação realizada no presente estudo não levou em consideração uma estrutura funcional na geração dos dados. Nem todo conjunto de dados submetido a uma análise funcional possui uma estrutura funcional (Ramsay e Silverman, 2005, p. 5). Porém, os objetivos de uma análise podem ser funcionais naturalmente, como nos casos de se usar um conjunto de dados finitos para estimar uma função, suas derivadas ou os valores de outros funcionais (Ramsay e Dalzell, 1991). Isso reforça a eficiência e a aplicabilidade do método no estudo de GEI.

Apesar do método SSVD apresentar alguns problemas em aberto, como em aplicações com dados perdidos, o seu uso pode se tornar cada vez mais comum devido à sua simplicidade e robustez (Xiao et al., 2014). Os resultados obtidos no presente trabalho indicam que sua integração ao método EM-AMMI para realizar a imputação de dados e lidar com o desbalanceamento pode alterar a perspectiva de estudo da interação genótipos por ambientes. O estudo de GEI sob uma perspectiva funcional não tem a interação como foco principal, mas sim o padrão de interação. A análise não é baseada no comportamento de um genótipo em determinado ambiente, e sim em como o comportamento genotípico é influenciado pela alteração de condições ambientais. Para isso, o método não pode ser dependente de variáveis

do ambiente que não se repetem de um ano para outro. Dados como coordenadas de latitude representam uma condição geográfica, a qual é constante. Isso abre novas perspectivas de inclusão de variáveis ambientais no estudo de GEI. A relação funcional existente entre genótipos e a interação cria a possibilidade de extrapolar os dados obtidos para ambientes não testados e, assim, reduzir custos pelo menor número de ensaios sem perda de eficiência do programa de melhoramento. Além disso, o método também é eficiente em analisar um grande número de genótipos, situação comum dentro de programas de melhoramento. A identificação de padrões de efeitos da interação através do método EM-AMMI Funcional representa uma nova forma de entender GEI, e abre diversas possibilidades para aplicação no melhoramento de plantas.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

REFERÊNCIAS

- Acosta-Pech, R., Crossa, J., de los Campos, G., Teyssèdre, S., Claustres, B., Pérez-Elizalde, S., Pérez-Rodríguez, P. (2017). Genomic models with genotype x environment interaction for predicting hybrid performance: an application in maize hybrids. *Theoretical and Applied Genetics*, 130(7), 1431-1440.
- Allard, R. W., & Bradshaw, A. D. (1964). Implications of genotype-environmental interactions in applied plant breeding. *Crop science*, 4(5), 503-508.
- Bernardo Júnior, L. A. Y., da Silva, C. P., de Oliveira, L. A., Nuvunga, J. J., Pires, L. P. M., Von Pinho, R. G., & Balestre, M. (2018). AMMI bayesian models to study stability and adaptability in maize. *Agronomy Journal*, 110(5), 1765-1776.
- Burgueño, J., de los Campos, G., Weigel, K., & Crossa, J. (2012). Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. *Crop Science*, 52(2), 707-719.
- Crossa, J. (1990). Statistical analyses of multilocation trials. In *Advances in agronomy* (Vol. 44, pp. 55-85). Academic Press.
- da Silveira Neto, P. C. (2013). *Suavização não-paramétrica e análise de variância funcional* (Monograph). Retrieved from Federal University of Rio Grande do Sul Digital Repository <https://www.lume.ufrgs.br/handle/10183/66473>

- de Figueiredo, A. G., Von Pinho, R. G., Silva, H. D., & Balestre, M. (2015). Application of mixed models for evaluating stability and adaptability of maize using unbalanced data. *Euphytica*, 202(3), 393-409.
- Eberhart, S. A., & Russell, W. A. (1966). Stability parameters for comparing varieties. *Crop science*, 6(1), 36-40.
- Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58(3), 453-467.
- Gauch, H. G. (2006). Statistical analysis of yield trials by AMMI and GGE. *Crop science*, 46(4), 1488-1500.
- Gauch, H. G. (2013). A simple protocol for AMMI analysis of yield trials. *Crop Science*, 53(5), 1860-1869.
- Gauch, H. G., Piepho, H. P., & Annicchiarico, P. (2008). Statistical analysis of yield trials by AMMI and GGE: Further considerations. *Crop science*, 48(3), 866-889.
- Gauch, H. G., & Zobel, R. W. (1990). Imputing missing yield trial data. *Theoretical and Applied Genetics*, 79(6), 753-761.
- Gavrilets, S., & Scheiner, S. M. (1993). The genetics of phenotypic plasticity. V. Evolution of reaction norm shape. *Journal of evolutionary biology*, 6(1), 31-48.
- Krzyśko, M., Derejko, A., Górecki, T., & Gacek, E. (2013). Principal component analysis for functional data on grain yield of winter wheat cultivars. *Biometrical Letters*, 50(2), 81-94.
- Lin, C. S., Binns, M. R., & Lefkovich, L. P. (1986). Stability Analysis: Where Do We Stand? *Crop science*, 26(5), 894-900.
- Lopez-Cruz, M., Crossa, J., Bonnett, D., Dreisigacker, S., Poland, J., Jannink, J. L., ... & de los Campos, G. (2015). Increased prediction accuracy in wheat breeding trials using a marker \times environment interaction genomic selection model. *G3: Genes, Genomes, Genetics*, 5(4), 569-582.
- Piepho, H. P. (1998). Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. *Theoretical and Applied Genetics*, 97(1-2), 195-201.
- R Development Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2018.
- Ramalho, M. A. P., Abreu, A. F. B., Santos, J. B., & Nunes, J. A. R. (2012). *Aplicações da Genética Quantitativa no Melhoramento de Plantas Autógamas*. Lavras: Ed. UFLA.
- Ramsay, J. O., & Dalzell, C. J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3), 539-572.
- Ramsay, J. O., & Silverman, B. W. (2005). Introduction. In Ramsay, J. O., & Silverman, B. W. (Eds.), *Functional Data Analysis* (pp. 1-18). New York, NY: Springer.
- Ramsay, J. O., & Silverman, B. W. (2005). From functional data to smooth functions. In Ramsay, J. O., & Silverman, B. W. (Eds.), *Functional Data Analysis* (pp. 37-58). New York, NY: Springer.

- Ramsay, J. O., & Silverman, B. W. (2005). Principal components analysis for functional data. In Ramsay, J. O., & Silverman, B. W. (Eds.), *Functional Data Analysis* (pp. 147-172). New York, NY: Springer.
- Smith, A. B., Cullis, B. R., & Gilmour, A. (2001). The analysis of crop variety evaluation data in Australia. *Australian & New Zealand Journal of Statistics*, 43(2), 129–145.
- Toler, J. E., & Burrows, P. M. (1998). Genotypic performance over environmental arrays: a non-linear grouping protocol. *Journal of Applied Statistics*, 25(1), 131-143.
- van Eeuwijk, F. A., Bustos-Korts, D. V., & Malosetti, M. (2016). What should students in plant breeding know about the statistical aspects of genotype \times environment interactions? *Crop Science*, 56(5), 2119-2140.
- Wang, J. L., Chiou, J. M., & Müller, H. G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, 3, 257-295.
- Xiao, L., Zipunnikov, V., Ruppert, D., & Crainiceanu, C. (2014). Fast covariance estimation for high-dimensional functional data. *Statistics and computing*, 26(1-2), 409-421.
- Yan, W., Hunt, L. A., Johnson, P., Stewart, G., & Lu, X. (2002). On-farm strip trials vs. replicated performance trials for cultivar evaluation. *Crop Science*, 42(2), 385-392.
- Yan, W., & Rajcan, I. (2002). Biplot analysis of test sites and trait relations of soybean in Ontario. *Crop Science*, 42(1), 11-20.
- Yan, W., & Tinker, N. A. (2006). Biplot analysis of multi-environment trial data: Principles and applications. *Canadian journal of plant science*, 86(3), 623-645.
- Zobel, R. W., Wright, M. J., & Gauch, H. G. (1988). Statistical analysis of a yield trial. *Agronomy journal*, 80(3), 388-393.

MATERIAL SUPLEMENTAR

Tabela S1. Médias e variâncias da correlação média obtida entre valores simulados e preditos em cada amostragem. Os resultados são referentes aos ambientes avaliados sob situação de desbalanceamento, ou seja, dos ambientes não eliminados considerando quebra da estrutura pela retirada de 10% dos ambientes de maneira sistemática.

Cenário	NG	NA	r	NPC	Correlação Média	
					Média	σ^2
1	50	100	0.8	10	0.8329	0.00000
2	50	100	0.2	10	0.5347	0.00013
3	1000	100	0.8	10	0.7779	0.00000
4	1000	100	0.2	10	0.4282	0.00000
5	2000	100	0.8	10	0.7807	0.00000
6	2000	100	0.2	10	0.4119	0.00000
7	3000	100	0.8	10	0.7817	0.00000
8	3000	100	0.2	10	0.4100	0.00000
9	50	200	0.8	10	0.7470	0.00000
10	50	200	0.2	10	0.4009	0.00059
11	50	300	0.8	10	0.6922	0.00000
12	50	300	0.2	10	0.3390	0.00046
13	300	500	0.8	10	0.4711	0.00000
14	300	500	0.2	10	0.2826	0.00001
15	100	100	0.8	10	0.8073	0.00000
16	100	100	0.2	10	0.4900	0.00016
17	300	300	0.8	10	0.5592	0.00000
18	300	300	0.2	10	0.3289	0.00000
19	2000	100	0.8	30	0.9331	0.00000
20	2000	100	0.2	30	0.6833	0.00000

NG: Número de Genótipos; NA: Número de Ambientes; r: Correlação entre Ambientes; NPC: Número de Componentes Principais. σ^2 : variância.

Tabela S2. Médias e variâncias da correlação média obtida entre valores simulados e preditos em cada uma das 1000 amostragens realizadas. Os resultados são referentes aos ambientes avaliados sob situação de desbalanceamento, ou seja, dos ambientes não eliminados considerando quebra da estrutura pela retirada de 5, 10, 20 e 30 % dos ambientes de maneira aleatória, repetida 1000 vezes.

Cenário	NG	NA	r	NPC	5% de Perdas		10% de Perdas		20% de Perdas		30% de Perdas	
					Correlação Média		Correlação Média		Correlação Média		Correlação Média	
					Média	σ^2	Média	σ^2	Média	σ^2	Média	σ^2
1	50	100	0.8	10	0.8316	0.00000	0.8332	0.00001	0.8370	0.00001	0.8422	0.00002
2	50	100	0.2	10	0.5327	0.00018	0.5341	0.00027	0.5413	0.00044	0.5471	0.00065
3	1000	100	0.8	10	0.7756	0.00000	0.7776	0.00000	0.7822	0.00000	0.7881	0.00001
4	1000	100	0.2	10	0.4209	0.00000	0.4287	0.00000	0.4457	0.00000	0.4661	0.00001
5	2000	100	0.8	10	0.7788	0.00000	0.7807	0.00000	0.7851	0.00000	0.7907	0.00001
6	2000	100	0.2	10	0.4054	0.00000	0.4125	0.00000	0.4279	0.00001	0.4475	0.00001
7	3000	100	0.8	10	0.7799	0.00000	0.7818	0.00000	0.7861	0.00000	0.7915	0.00001
8	3000	100	0.2	10	0.4045	0.00000	0.4102	0.00000	0.4241	0.00001	0.4418	0.00001
9	50	200	0.8	10	0.7467	0.00000	0.7478	0.00000	0.7506	0.00001	0.7543	0.00001
10	50	200	0.2	10	0.3947	0.00034	0.3948	0.00040	0.3972	0.00052	0.3983	0.00063
11	50	300	0.8	10	0.6918	0.00000	0.6930	0.00000	0.6958	0.00001	0.6993	0.00001
12	50	300	0.2	10	0.3414	0.00032	0.3339	0.00041	0.3247	0.00057	0.3214	0.00063
13	300	500	0.8	10	0.4708	0.00000	0.4719	0.00000	0.4749	0.00000	0.4786	0.00000
14	300	500	0.2	10	0.2799	0.00000	0.2823	0.00001	0.2876	0.00002	0.2930	0.00003
15	100	100	0.8	10	0.8061	0.00000	0.8079	0.00000	0.8123	0.00000	0.8175	0.00001
16	100	100	0.2	10	0.4912	0.00014	0.4938	0.00017	0.5026	0.00026	0.5132	0.00033
17	300	300	0.8	10	0.5588	0.00000	0.5600	0.00000	0.5632	0.00000	0.5675	0.00001
18	300	300	0.2	10	0.3261	0.00000	0.3296	0.00001	0.3375	0.00002	0.3470	0.00003
19	2000	100	0.8	30	0.9311	0.00000	0.9332	0.00000	0.9382	0.00000	0.9445	0.00000
20	2000	100	0.2	30	0.6719	0.00000	0.6837	0.00000	0.7107	0.00000	0.7433	0.00000

NG: Número de Genótipos; NA: Número de Ambientes; r: Correlação entre Ambientes; NPC: Número de Componentes Principais; σ^2 : variância.