



**ERIC VINICIUS VIEIRA SILVA**

**GENOMIC PREDICTION STRATEGIES FOR GRAIN YIELD  
STABILITY IN SECOND SEASON MAIZE HYBRIDS**

**LAVRAS – MG  
2023**

**ERIC VINICIUS VIEIRA SILVA**

**GENOMIC PREDICTION STRATEGIES FOR GRAIN YIELD STABILITY IN  
SECOND SEASON MAIZE HYBRIDS**

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

Prof. Dr. Renzo Garcia Von Pinho  
Orientador

**LAVRAS – MG  
2023**

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

Silva, Eric Vinicius Vieira.

Genomic prediction strategies for grain yield stability in  
second season maize hybrids / Eric Vinicius Vieira Silva. – 2023.  
71 p. : il.

Orientador: Renzo Garcia Von Pinho

Tese (Doutorado) - Universidade Federal de Lavras, 2023  
Bibliografia.

1. *Zea mays*. 2. Interação genótipos por ambientes. 3. Seleção  
genômica. I. Von Pinho, Renzo Garcia. II. Título.

**ERIC VINICIUS VIEIRA SILVA**

**GENOMIC PREDICTION STRATEGIES FOR GRAIN YIELD STABILITY IN  
SECOND SEASON MAIZE HYBRIDS**

**ESTRATÉGIAS DE PREDIÇÃO GENÔMICA PARA A ESTABILIDADE  
PRODUTIVA DE HÍBRIDOS DE MILHO DE SEGUNDA SAFRA**

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

APROVADO em 10 de julho de 2023.

Dr. José Maria Villela Pádua	UFLA
Dr. Carlos Pereira da Silva	UFLA
Dr. André Humberto de Brito	Corteva Agriscience
Dr. Vander Fillipe de Souza	FTS Sementes S.A.

Prof. Dr. Renzo Garcia Von Pinho  
Orientador

**LAVRAS – MG  
2023**

*À minha família, sem a qual este sonho não seria possível.  
Dedico.*

## AGRADECIMENTOS

Primeiramente a Deus, pelo dom da vida, pelas oportunidades a mim concedidas e por todas as pessoas especiais que encontrei durante esta jornada.

Aos meus pais, Glécio e Ediléia, por todos os valores ensinados, por acreditarem e me apoiarem em todas as etapas.

Às minhas irmãs, Júlia e Amanda, por estarem ao meu lado, em todos os momentos de alegrias e dificuldades, seu apoio foi fundamental.

À minha futura esposa, Juliana, por todo o amor, carinho, compreensão e apoio incondicional sem o qual eu não teria chegado até aqui. Te amo!

À Universidade Federal de Lavras e ao programa de pós-graduação em Genética e Melhoramento de Plantas, pela oportunidade concedida para a realização do doutorado. Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), CAPES e à FAPEMIG.

Ao meu orientador, Renzo Garcia Von Pinho, pelos ensinamentos, conhecimentos, oportunidades, dedicação e confiança.

Ao meu coorientador, Márcio Balestre (*in memoriam*), que quem, embora o curto período, pude aprender bastante e foi fundamental para a idealização deste trabalho.

A todos os professores do PPGGM, por todo o conhecimento a mim transmitido, sua contribuição foi fundamental para a minha formação.

Ao GEN, com o qual contribuí e do qual pude fazer parte, durante minha jornada na UFLA, pelas discussões e momentos de aprendizado.

Às grandes amizades que fiz, durante toda a minha jornada na UFLA, em especial, aos amigos Fábio (Fabim), Alan, Rafael (Botinha), Marco Renan (Marcola), Heloisa, Brena, Maiara, Roxane, Rodolfo, Paula, Thiagueira, Aline, Gerald (Paraguai), Thiago Tristão, Luiz (Pato), Carlos, Ewerton (Peixe), Guilherme, Jessica, Frederico. Cada um de vocês contribuiu de forma ímpar nesta jornada!

A todos os meus mais sinceros agradecimentos, muito obrigado!

O presente trabalho foi realizado com o apoio da Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) e da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

## RESUMO

A interação genótipos por ambientes (GxA) é um fator complicador do melhoramento de milho. Selecionar híbridos que apresentem comportamento estável, ao longo dos locais e anos, é essencial para o programa. Essa característica ganha ainda mais importância para os programas voltados para a segunda safra. A inclusão de ferramentas de predição genômica, nos programas de melhoramento de milho, tem sido cada vez mais frequente. A seleção genômica não somente permite a redução do tempo necessário, para a realização de um ciclo de melhoramento, mas também aumentar a quantidade de genótipos em estudo, sem aumentar significativamente os custos com fenotipagem. Há diversos relatos das vantagens da inclusão da GxA nos modelos de predição. Entretanto, há poucos relatos de estudos visando prever a estabilidade produtiva em híbridos de milho. Diante do exposto, o presente trabalho foi realizado com o intuito de verificar a viabilidade de prever a estabilidade em híbridos de segunda safra e definir estratégias simples e eficientes para lidar com os cenários reais de ensaios Multiambientais. Foram realizados dois trabalhos de pesquisa. Para este estudo, foi utilizado um conjunto de dados de um programa privado de melhoramento de milho. Mais de 1300 híbridos foram avaliados, em 12 ambientes distintos, durante as safrinhas de 2012/13 e 2013/14. O conjunto de dados foi subdividido em três: 1) 185 híbridos comuns ao longo dos seis ambientes em 2012/13; 2) 309 híbridos avaliados nos seis ambientes de 2012/13; 3) 710 híbridos avaliados nos seis ambientes de 2013/14. O conjunto 1 foi utilizado no primeiro trabalho, enquanto os conjuntos 2 e 3 no segundo. No primeiro trabalho, as predições de nove índices de adaptabilidade e estabilidade foram comparadas com uma abordagem multiambiental, em um cenário de balanceamento genético, ao longo dos ambientes. Foi utilizado o modelo BRR (Bayesian Ridge Regression), e as capacidades preditivas foram aferidas via validação cruzada (10-fold). Os índices da distância euclidiana e MHPRVG (Média Harmônica da Performance Relativa dos Valores Genéticos) se mostraram superiores à abordagem multiambiental. No segundo trabalho, o modelo BRR foi mantido, entretanto foram considerados quatro cenários de predição: *i*) ambiente-único; *ii*) índices de estabilidade; *iii*) multiambiental (ME) desconsiderando GxA; e *iv*) ME incluindo a GxA. Além disso, foram considerados dois esquemas de validação cruzada: CV1 (10-fold) e CV2 (predição de ambientes inteiros, utilizado para a abordagem ME). O índice da distância euclidiana não se mostrou viável, já a utilização da MHPRVG mostrou-se consistente para ambos os conjuntos de dados. Considerando CV1, a inclusão da interação GxA não foi vantajosa, aumentando o tempo necessário para as predições, mas sem ganhos (2012/13) ou com ganhos marginais (2013/14) na capacidade preditiva. Considerando CV2, *iii* e *iv* apresentaram capacidades preditivas muito baixas. De forma geral, no segundo trabalho, a abordagem ME (*iii*) foi superior aos índices de estabilidade (*ii*), entretanto a utilização do índice MHPRVG, como estratégia para prever a estabilidade de híbridos de milho de segunda safra, mostrou-se viável nos dois trabalhos.

**Palavras-chave:** *Zea mays*. Interação genótipos por ambientes. Seleção genômica. Índices de estabilidade. MHPRVG.

## ABSTRACT

The genotype by environment interaction (GxE) is a major factor in maize breeding. Therefore, it is essential to select genotypes that are stable across locations and over the years. Genotype stability is even more important for second season maize breeding programs. The use of genomic prediction tools in maize breeding has been frequent. Genomic selection not only reduces the time required per breeding cycle but also allows to study of a higher number of genotypes without significantly increasing the phenotyping costs. Several reports highlight the advantages of including the GxE in the prediction models. However, to predict maize stability has been scarcely reported. Given the above, the present work was carried out aiming to verify the feasibility of predicting second season maize stability and to define simple and efficient strategies to deal with real scenarios of multi-environment trials. Two research works were performed. For this study, a private maize breeding dataset was used. Over 1300 maize hybrids were assessed across 12 environments during the 2012/13 and 2013/14 second seasons. The dataset was split into three: 1) 128 hybrids that were common across the six 2012/13's environments; 2) all 309 hybrids assessed in the six 2012/13's environments; 3) all 710 hybrids assessed in the six 2013/14's environments. Dataset 1 was used in the first research work, while datasets 2 and 3 in the second one. In the first work, the predictions of nine adaptability and stability indices were compared to a multi-environmental approach under a genetically balanced (across the environments) scenario. A BRR (Bayesian Ridge Regression) model was used, and the predictive abilities were measured via cross-validation (10-fold). The Euclidean Distance and MHPRVG (harmonic mean of the relative performance of the breeding values) indices outperformed the multi-environmental approach. In the second work, the BRR model was maintained, however, four prediction scenarios were considered: *i*) single-environment; *ii*) stability indices; *iii*) Multi-environmental (ME) without including GxE, and *iv*) ME including GxE effects. In addition, it was considered two cross-validation schemes: CV1 (10-fold), and CV2 (whole environment predictions, ME only). The Euclidian distance index did not prove feasible, on the other hand, the MHPRVG results were consistent for both datasets. Considering CV1, the inclusion of GxE was not advantageous, increasing the time required for predictions either without (2012/13) or with marginal gains (2013/14) in predictive ability. Considering CV2, *iii* e *iv* presented very poor predictability. In general, considering the second work, the ME approach (*iii*) outperformed the stability indices (*ii*), however, in both works, the use of MHPRVG index has proved feasible as a strategy to predict the stability of second season maize hybrids.

**Keywords:** *Zea mays*. Genotypes-by-environments interaction. Genomic selection. Stability index. MHPRVG.



## SUMÁRIO

<b>PRIMEIRA PARTE .....</b>	<b>9</b>
<b>1 GENERAL INTRODUCTION.....</b>	<b>9</b>
<b>REFERENCES.....</b>	<b>11</b>
<b>SEGUNDA PARTE .....</b>	<b>13</b>
<b>ARTIGO 1 – Directly predicting maize hybrid’s grain yield stability: Is it an efficient strategy? .....</b>	<b>13</b>
<b>ARTIGO 2 – Genomic prediction for maize grain yield stability under multi- environment scenarios.....</b>	<b>42</b>

## PRIMEIRA PARTE

### 1 GENERAL INTRODUCTION

The importance of maize for Brazilian agriculture is undeniable. In the last 20 years, Brazil's maize productivity has increased by over 145 kg ha<sup>-1</sup> per year (Von Pinho; Silva; Oliveira, 2022). As Brazil is a tropical country, Brazilian maize farmers face intense challenges, such as pests and disease pressures, and above all drought conditions, especially in the second season (ANDREA et al., 2018). The genotype by environment interaction (GxE) is one of the major challenges faced by maize breeders. The GxE plays a crucial role in the genotype selection process (VON PINHO; SILVA; OLIVEIRA, 2022).

The study and better understanding of this interaction aid the breeder to overcome the GxE challenges. Breeders may minimize or even capitalize on the GxE effects. In this scenario, it is essential to select genotypes that are stable across locations and over the years. Several methodologies have been developed and reported aiming to identify stable genotypes.

Wricke (1965) proposed to measure the contribution of a given genotype to the GxE, in a way that stable genotypes are those whose least contribute to the interaction. Eberhart and Russel (1966) presented a regression approach to study the stability of the genotypes, in which the stable genotypes present a predictable behavior over the environment. Annicchiarico (1992) suggests a "risk of adoption" index, in which the higher the risk the lower the stability of the genotype. In addition, other methodologies have been proposed over the years, such as AMMI (GOLLOB, 1968; MANDEL, 1969, 1971) and GGEBiplot (YAN et al., 2000).

Since its first proposal (MEUWISSEN; HAYES, B. J.; GODDARD, 2001), genomic prediction (GP) methods have become an important tool in plant breeding pipelines. Heslot et al. (2014) highlight that GP shortens the time required to complete a breeding cycle, increasing the breeding program efficiency. In addition, the implementation of GP tools allows to increase the number of genotypes under studies without significantly increasing phenotyping costs. Beyene et al. (2019) reported cost reductions of up to 32% by including GP tools in tropical maize breeding.

Many reports can be found related to how to improve GP approaches. Dias et al. (2018) achieve predictive gains by including non-additive effects in the prediction models. The advantages of modeling the GxE in the GP approach have been widely reported (ACOSTA-PECH et al., 2017; BURGUEÑO et al., 2012; DIAS et al., 2018, MAGETO et al., 2020; WANG et al., 2020). The opportunity of using environmental covariables has been highlighted by

Costa-Neto, Crossa e Fritsche-Neto (2021), Gillberg et al. (2019) and Jarquín et al. (2021). In addition, Montesinos-López et al. (2021) present the use of deep learning in GP, while Atanda et al. (2021) and Fritsche-Neto et al. (2021) published strategies and guidelines to optimize training populations. However, predicting maize stability has been scarcely reported

Given the above, the present work was carried out aiming to verify the feasibility of predicting second season maize stability and to define simple and efficient strategies to deal with real scenarios of multi-environment trials.

In the first chapter, we investigated the prediction performance of several stability indices using a real second season maize dataset. We evaluated nine stability indices and compared them to a Bayesian multi-environmental approach under a genetically balanced scenario. In the second chapter, two unbalanced datasets were used to verify the feasibility of maize stability predictions and to identify the best strategies to lead with multi-environmental scenarios. Four strategies were evaluated under two different cross-validation schemes.

## REFERENCES

- ACOSTA-PECH, R. et al. Genomic models with genotype x environment interaction for predicting hybrid performance: an application in maize hybrids. **Theoretical and Applied Genetics**, Berlin, v. 130, n. 7, p. 1431-1440, July 2017.
- ANDREA, M. C. da S. et al. Variability and limitations of maize production in Brazil: Potential yield, waterlimited yield and yield gaps. **Agricultural Systems**, Essex, v. 165, p. 264-273, Sept. 2018.
- ANNICCHIARICO, P. Cultivar adaptation and recommendation from alfalfa trials in Northern Italy. **Journal of Genetics and Breeding**, Rome, v. 46, p. 269-269, 1992.
- ATANDA, S. A. et al. Maximizing efficiency of genomic selection in CIMMYT's tropical maize breeding program. **Theoretical and Applied Genetics**, Berlin, v. 134, n. 1, p. 279-294, Jan. 2021.
- BEYENE, Y. et al. Empirical comparison of tropical maize hybrids selected through genomic and phenotypic selections. **Frontiers in Plant Science**, Lausanne, v. 10, n. 1502, p. 1–11, Nov. 2019.
- BURGUEÑO, J. et al. Genomic prediction of breeding values when modeling genotype x environment interaction using pedigree and dense molecular markers. **Crop Science**, Madison, v. 52, N. 2, p. 707-719, Mar. 2012.
- COSTA-NETO, G.; CROSSA, J.; FRITSCHE-NETO, R. Enviromic assembly increases accuracy and reduces costs of the genomic prediction for yield plasticity in maize. **Frontiers in Plant Science**, Lausanne, v. 12, n. 717552, p. 1-21, Oct. 2021.
- DIAS, K. O. das G. et al. Estimating genotype × environment interaction for and genetic correlations among drought tolerance traits in maize via factor analytic multiplicative mixed models. **Crop Science**, Madison, v. 58, n. 1, p. 72-83, Jan./Feb. 2018.
- EBERHART, S. A.; RUSSEL, W. A. Stability parameters for comparing varieties. **Crop Science**, Madison, v. 6, n. 1, p. 36-40, Jan./Feb. 1966.
- FRISTCHE-NETO, R. et al. Optimizing genomic-enabled prediction in small-scale maize hybrids breeding programs: a roadmap review. **Frontiers in Plant Science**, Lausanne, v. 12, n. 658267, p. 1-16, July 2021.
- GILLBERG, J. et al. Modeling G x E with historical weather information improves genomic prediction in new environments. **Bioinformatics**, London, v. 35, n. 20, p. 4045-4052, Oct. 2019.
- GOLLOB, H. F. A statistical model which combines features of factor analytic and analysis of variance techniques. **Psychometrika**, New York, v. 33, n. 1, p. 73–115, Mar. 1968.

- HESLOT, N. et al. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. **Theoretical and Applied Genetics**, Berlin, v. 127, n. 2, p. 463–480, 2014.
- JARQUÍN, D. et al. Utility of climatic information via combining ability models to improve genomic prediction for yield within the genomes to fields maize project. **Frontiers in Genetics**, Lausanne, v. 11, p. 1-11, Mar. 2021.
- MANDEL, J. A new analysis of variance model for non-additive data. **Technometrics**, Washington, v. 13, n 1, p. 1-18, Feb. 1971.
- MANDEL, J. The partitioning of interactions in analysis of variance. **Journal of Research of the National Bureau of Standards**, Gaithersburg, v. 73B, n. 4, p. 309-328, Oct./Dec. 1969.
- MAGETO, E. K. et al. Genomic predictions with genotype by environment interaction analysis for kernel zinc concentration in tropical maize germplasm. **G3: genes-genomes-genetics**, Bethesda, v. 10, n. 8, p. 2629-2639, Aug. 2020.
- MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genomewide dense marker maps. **Genetics**, New York, v. 157, n. 4, p. 1819-1829, Apr. 2001.
- MONTESINOS-LÓPEZ, O. A. et al. A review of deep learning applications for genomic selection. **BMC Genomics**, London, v. 22, n. 19, p. 1-23, 2021.
- VON PINHO, R. G.; SILVA, E. V. V.; OLIVEIRA, T. L. de. Challenges of maize breeding under tropical conditions of Brazil. **Revista Brasileira de Milho e Sorgo**, Sete Alagoas, v. 21, nesp., 1258, p. 1-25, 2022.
- WANG, N. et al. Genomic prediction across years in a maize double haploid breeding program to accelerate early-stage testcross testing. **Theoretical and Applied Genetics**, Berlin, v. 133, p. 2869-2879, 2020.
- WRICKE, G. Zur berechnung der okovalenz bei sommerweizen und hafer. Z. **Pflanzenzuchtg**, [S.l.], v. 52, p. 127-138, 1965.
- YAN, W. et al. Cultivar evaluation and megaenvironment investigation based on GGE biplot. **Crop Science**, Madison, v. 40, n. 3, p. 597-605, May/June 2000.

**SEGUNDA PARTE****ARTIGO 1 – DIRECTLY PREDICTING MAIZE HYBRID’S GRAIN YIELD  
STABILITY: IS IT AN EFFICIENT STRATEGY?**

Artigo redigido conforme as normas da revista *Euphytica*

(International Journal of Plant Breeding).

## Directly predicting maize hybrid's grain yield stability: Is it an efficient strategy?

Eric Vinicius Vieira Silva

**Abstract** – Tropical agriculture provides intense climatic challenges. Although its relevance, the second season presents greater crop risks primarily due to lower water availability. Furthermore, the genotypes-by-environments (GxE) interaction is even more challenging to maize breeding in tropical conditions. In this scenario, the identification of highly stable maize hybrids is a powerful strategy to minimize the GxE effects. The genomic prediction (GP) approach has been increasingly employed in plant breeding, and the inclusion of GxE effects in the GP models is an efficient strategy to improve prediction accuracy. However, multi-environmental GP is computationally demanding and time-consuming. Therefore, simpler, and easier strategies could be applied to perform GP to select highly phenotypic stable genotypes. The objective of this work was to identify simple and efficient strategies to perform genomic selection by directly predicting single-cross maize hybrids' stability. Nine indices were estimated using a balanced dataset of 128 maize hybrids evaluated over six environments. Genomic predictions were performed by applying a Bayesian Ridge Regression model. Ten rounds of 10-fold cross-validation were performed to evaluate predictive ability (PA) and accuracy (ACC). Considering the nature of the dataset, seven out of nine indices present satisfactory PA and ACC, fluctuating from 0.17 to 0.31, and from 0.33 to 0.61, respectively. On average, indices predicting outperform the Bayesian Multi-Environment approach by 29.7%, considering the seven well-predicted indices. Wricke's ecovalence index and the main effect of genotypes present better performance, however, considering larger datasets, the Euclidian Distance and the Harmonic Mean of the Relative Performance of the Breeding Values (MHPRVG) indices should be more adequate.

**Keywords:** *Zea mays*; Genotypes-by-environments interaction; Genomic selection; Stability index; Bayesian Multi-Environment approach.

## **Introduction**

The great importance of the second season, or safrinha, maize for Brazilian agriculture is undeniable. However, in tropical agriculture, maize farmers face more intense challenges, such as greater pests and disease pressures. Furthermore, according to Andrea et al. (2018), water deficit stress is the major limiting factor for tropical maize yielding.

Even though its relevance, the second growing season presents greater crop risks requiring accurate planning in terms of cultivar choice, planting date, and crop management. The genotypes-by-environments interaction (GxE) is one of the major challenges faced by maize breeders. A better understanding of the GxE provides valuable information for hybrids selection, breeding zones, or mega-environments definition, as well as specific genotypes recommendation to target environments (Von Pinho; Silva; Oliveira, 2022).

In this scenario, the identification of highly stable maize hybrids is critical for maize farmers and breeding companies regarding the allocation, recommendation, and adoption of the hybrid.

Several methodologies have been developed to help plant breeders identify and select stable genotypes, such as Wricke's ecovalence (Wricke, 1965), Annicchiarico's index (Annicchiarico, 1992), AMMI (Gollob, 1968; Mandel, 1969, 1971), GGEBiplot (Yan et al., 2000), Eberhart and Russel's stability parameters (Eberhart and Russel, 1966), and others. However, to study the GxE, it is required the evaluation of the genotypes in multi-environmental trials (MET), which represents high costs for the breeding company.

With the constant reduction in genotyping costs and the increase in phenotyping costs, genomic prediction (GP) and selection (GS) approaches have been increasingly employed in



plant breeding programs (Volpato; Bernardeli; Gomez, 2021). Using this powerful prediction tool, it is possible to reduce the number of hybrids to be phenotypically evaluated, reducing phenotyping costs. Furthermore, the main benefit of GP is to select individuals based only on their genotypes, increasing the rate of genetic gain per unit of time.

Recent studies have demonstrated the possibility of incorporating the GxE into the genomic prediction models. Burgueño et al. (2012) showed that prediction accuracy could be improved by modeling GxE. Later, Acosta-Pech et al. (2017) achieved up to 22% more predictive ability when including GxE into specific (SCA) and general (GCA) combining ability models, resulting in higher accuracy in the prediction of untested maize hybrids. In its turn, Dias et al. (2018) demonstrated the benefits of including GxE in additive and additive-dominance models to predict drought tolerance in maize single-cross hybrids. Furthermore, Wang et al. (2020) highlight the benefits of using a training population evaluated over multiple years to increase prediction accuracy.

Modeling GxE is particularly useful in unbalanced trials, a common scenario in most breeding programs. The GP models might be used to predict non-evaluated genotypes in target environments, and even to predict complete environments using genetic-environmental covariates (Jarquin et al., 2021).

Although the inclusion of GxE effects in GP models is of utmost useful, in real situation scenarios, such predictions are quite computational demanding and time-consuming when using a Bayesian framework (Dos Santos et al., 2020; Hamazaki and Iwata, 2022; Silveira et al., 2019; Wang et al., 2015; Wang et al., 2017; Montesinos-López et al., 2019a). Therefore, simplified, and efficient strategies to deal with GxE might be necessary.

The objective of this work was to identify simple and efficient strategies to study the genotypes by environment interaction by directly predicting the stability of single-cross maize hybrids.

## Materials and methods

### Phenotypic data

A dataset of 128 maize hybrids were evaluated for grain yield ( $\text{t ha}^{-1}$ ) in six different environments during the safrinha of the 2012/2013 crop season (Table 1). The maize hybrids were evaluated under incomplete block designs, with a variable number of replicates, in 5-m-wide four rows spaced 0.5 m (between rows) plots.

**Table 1.** Grain yield means and coefficient of variation of maize single-cross hybrids cultivated across six environments during the safrinha of the 2012/2013 crop season.

Environment	GY mean ( $\text{t ha}^{-1}$ )	CVe (%)
E1 – Paraúna/GO	6.70	27.88
E2 – Riverlândia/GO	4.41	32.38
E3 – Chapadão do Céu/GO	4.84	28.03
E4 – Sorriso/MT	5.70	38.57
E5 – Rio Verde/GO	4.29	30.12
E6 – Campo Novo do Parecis/MT	5.20	35.62

CVe: coefficient of variation.

These maize hybrids were obtained from crosses of conventional inbred and double haploid (DH) lines from different backgrounds (Tropical, subtropical, and temperate regions). The lines characterization and heterotic groups were described by Cantelmo et al. (2017a).

### Genotypic data

The parental maize lines were genotyped by Diversity Arrays Technology company. 23153 DArTs markers were generated. The markers were filtered by adopting a threshold of 5% of Minor Allele Frequency (MAF). The imputation of missing markers data was performed

based on the markers' mean using the rrBLUP package (Endelman, 2011), however, markers that present more than 50% of missing data were removed.

After filtering, a panel of 21683 DArTs markers was used to build the genomic relationship (G) matrix using VanRaden's method (VanRaden, 2008). The hybrids markers panel was obtained from the combination of the parental lines, then the G matrix genomic was built.

### **Analysis of phenotypic data**

First, individual analyses, by the environment, were performed, aiming for a better understanding of the dataset. The components of variance were estimated using a mixed model approach. For this purpose, the following model was used:

$$Y = X\beta + Z_b u_b + Z_g u_g + \varepsilon$$

where  $Y$  is the vector of the phenotypic data;  $\beta$  is the vector of the fixed effect of the replicates and  $X$  is the incidence matrix associated with the fixed effect of the replicates;  $u_b$  is the vector of the random effect of the blocks nested within the replicates, with  $u_b \sim N(0, I_b \sigma_{u_b}^2)$ ;  $u_g$  is the vector of random effects of the genotypes, with  $u_g \sim N(0, I_g \sigma_{u_g}^2)$ ;  $Z_b$  and  $Z_g$  are the incidence matrix associated with the vectors  $u_b$  and  $u_g$ , respectively; and  $\varepsilon$  is the vector of the residuals, with  $\varepsilon \sim N(0, I_\varepsilon \sigma_\varepsilon^2)$ ;  $\sigma_{u_b}^2$ ,  $\sigma_{u_g}^2$ , and  $\sigma_\varepsilon^2$  are the components of variance of the blocks within replicates, genotypes, and residuals, respectively, while  $I_b$ ,  $I_g$ , and  $I_\varepsilon$  are the identity matrices associated with  $u_b$ ,  $u_g$ , and  $\varepsilon$ , respectively.

The significance of the components of variance was verified using the likelihood ratio (LR) tests:

$$LR = -2 * [\log L_{\text{Reduced}} - \log L_{\text{Full}}] \sim \chi_{f_{\text{Full}} - f_{\text{Reduced}}}^2$$

where:  $\log L_{\text{Reduced}}$  is the log of the reduced model likelihood (without a given random effect);  $\log L_{\text{Full}}$  is the log of the full model likelihood;  $\chi^2_{f_{\text{Full}} - f_{\text{Reduced}}}$  is a Chi-squared ( $\chi^2$ ) distribution with  $f_{\text{Full}} - f_{\text{Reduced}}$  degrees of freedom, in which  $f_{\text{Full}}$  and  $f_{\text{Reduced}}$  are the number of parameters in the full and reduced models, respectively.

In addition, the estimates of heritability were obtained using the methodology proposed by Cullis et al. (2006):

$$H^2 = 1 - \frac{\bar{V}_{\text{BLUP}}}{2\sigma_{u_g}^2}$$

where  $\bar{V}_{\text{BLUP}}$  is the mean-variance of a difference between two BLUPs (Best Linear Unbiased Prediction);  $\sigma_{u_g}^2$  is the component of variance of the genotypes.

In a second moment, the genotype effect was taken as fixed, and a joint analysis, considering the six environments was performed. For this purpose, the following model was used:

$$Y = X\beta + Z_b u_b + Z_a u_a + Z_{ga} u_{ga} + \varepsilon$$

where  $Y$  is the vector of the phenotypic data;  $\beta$  is the vector of fixed effects of the genotypes, and the replicates nested within the environments;  $X$  is the incidence matrix associated with the fixed effects;  $u_b$  is the vector of the random effect of the blocks nested within the replicates nested within the environments, with  $u_b \sim N(0, I_b \sigma_{u_b}^2)$ ;  $u_a$  is the vector of random effect of the environments, with  $u_a \sim N(0, I_a \sigma_{u_a}^2)$ ;  $u_{ga}$  is the vector of random effect of the genotypes x environments interaction, with  $u_{ga} \sim N(0, I_{ga} \sigma_{u_{ga}}^2)$ ;  $Z_b$ ,  $Z_a$ , and  $Z_{ga}$  are the incidence matrix associated with the vectors  $u_b$ ,  $u_a$ , and  $u_{ga}$ , respectively; and  $\varepsilon$  is the vector of residuals, with  $\varepsilon \sim N(0, I_\varepsilon \sigma_\varepsilon^2)$ ;  $\sigma_{u_b}^2$ ,  $\sigma_{u_a}^2$ ,  $\sigma_{u_{ga}}^2$ , and  $\sigma_\varepsilon^2$  are the components of variance of the blocks nested within the replicates nested within the environments, environments, genotypes x

environments interaction, and residuals, respectively, while  $I_b$ ,  $I_a$ ,  $I_{ga}$ , and  $I_\varepsilon$  are the identity matrices associated with  $u_b$ ,  $u_a$ ,  $u_{ga}$ , and  $\varepsilon$ , respectively.

The grain yield stability indices were estimated using the GxE matrix. These indices were used for genomic predictions.

### Stability indices

Wricke's ecovalence index ( $W_i^2$ ) were obtained from the adjusted phenotypic means. This methodology aims to determine the contribution of each genotype to the GxE interaction, in which the more stable genotypes are those whose present the smaller contribution to the genotypes by environments interaction. The  $W_i^2$  estimates were obtained from the following estimator (Ramalho et al., 2012):

$$W_i^2 = \sum_{j=1}^k (\bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} - \bar{Y}_{..})^2 = \sum_{j=1}^k \hat{g}\hat{e}_{ij}^2$$

where:  $\bar{Y}_{ij}$  is the mean of the  $i$ th maize hybrid in the  $j$ th environment;  $\bar{Y}_{i.}$  e  $\bar{Y}_{.j}$  represents the overall means of the  $i$ th maize hybrid and the  $j$ th environment, respectively; and  $\bar{Y}_{..}$  is the overall mean.

In addition, a principal component analysis (PCA) was performed to simplify the total observed GxE variation, in just a few explanatory principal components. The Euclidian distances (ED) of each maize hybrid were calculated using their scores. The ED was calculated using the norm of the score vectors of each maize hybrid:

$$ED_i = \sqrt{\sum_{k=1}^n PC_{ki}^2}$$

where:  $ED_i$  is the Euclidian distance of the  $i$ th maize hybrid;  $PC_{ki}$  is the score of the principal component  $k$  for the  $i$ th maize hybrid, being  $k = 1, e n = \min(\text{row}, \text{column})$  correspondent to the rank of the GxE matrix.

The observed ED of each hybrid was used as an alternative stability index, in which the smaller hybrid's ED, the greater its stability.

The indices proposed by Resende (2002; 2004) were also used. These indices consider not only the GxE matrix but also the main effect of the genotypes (G). Therefore, the indices of the MHVG (harmonic mean of the breeding values), PRVG (relative performance of the breeding values), and MHPRVG (harmonic mean of the relative performance of the breeding values) were adopted. The MHVG, PRVG, and MHPRVG estimates were obtained by:

$$MHVG_i = \frac{n}{\sum_{j=1}^n Vg_{ij}}; \quad PRVG_i = \frac{\sum_{j=1}^n \frac{Vg_{ij}}{l_j}}{n}; \quad MHPRVG_i = \frac{n}{\sum_{j=1}^n \frac{Vg_{ij}}{l_i}}$$

where:  $n$  is the number of environments where the  $i$ th maize hybrid was evaluated;  $Vg_{ij}$  is the breeding value of the  $i$ th maize hybrid in the  $j$ th environment; and  $l_j$  is the overall mean of the  $j$ th environment.

In addition, the indices proposed by Finlay and Wilkinson (1963) were also estimated. The Finlay-Wilkinson model consists of a regression on the environmental mean, as follows:

$$\bar{y}_{ij} = \mu + G_i + \beta_i E_j + \bar{\varepsilon}_{ij}$$

where  $\bar{y}_{ij}$  is the mean of the  $i$ th maize hybrid in the  $j$ th environment;  $\mu$  is the overall mean,  $G_i$  is the main effect of the  $i$ th maize hybrid, given by the average grain yield of the  $i$ th maize hybrid across environments;  $\beta_i$  is the slope of the  $i$ th maize hybrid;  $E_j$  is the main effect of the  $j$ th environment, given by the average grain yield of the  $j$ th environment across the maize hybrids;  $\bar{\varepsilon}_{ij}$  is the residual associated to  $\bar{y}_{ij}$ .

According to the Finlay-Wilkinson methodology, the stable genotypes present lower residual variance ( $\text{Var } E$ ), while the intercept ( $\mu_i = \mu + G_i$ ) and slope ( $\beta_i$ ) of each genotype indicate its general adaptation (G. Adaptation) and adaptability, respectively.

The estimates of the stability indices and the BLUEs of each hybrid were used for the training of the genomic prediction models, aiming to predict the grain yield stability of the maize hybrids under study.

### **Genomic prediction model**

The genomic predictions of the stability indices were performed via the R environment (R Core Team, 2021) using the BGLR package (Perez and De Los Campos, 2014). The Bayesian Ridge Regression (BRR) model was chosen, which assumes a Gaussian distribution with  $\mu = 0$  and  $\sigma_{\beta}^2 \sim X^{-2}$  as a prior of the model. Further details of the BRR model can be found in Perez and De Los Campos (2014).

The predictive abilities were measured using 10-fold cross-validation (cv) scheme. For each strategy, ten rounds of 10-fold cv were performed. For each round, the predictive ability of the model was calculated by the Pearson correlation of predicted and observed values in the testing populations.

The heritability ( $h^2$ ) and predictive accuracy (ACC) were estimated using the estimators proposed by Gezan et al. (2017):

$$\text{ACC} = \frac{\text{Corr}(y_{\text{obs}}, \hat{a})}{\sqrt{h^2}}$$

where:  $\text{corr}(y_{\text{obs}}, \hat{a})$  is the Pearson correlation of predicted and observed values;  $h^2$  is the heritability of the trait under study, being  $h^2 = 1 - \frac{\sigma_e^2}{\sigma_y^2}$ , in which  $\sigma_e^2$  e  $\sigma_y^2$  correspond to residual and phenotypic components of variance, respectively.

In addition, Bayesian Multi-environment (BME) genomic prediction methodology, presented in Montesinos-López et al. (2019b), was also performed to compare to indices prediction results. The BME analyzes were performed using the BME function from BMTME R-package (Luna-Vazquez et al., 2020). Once again, the BRR model was chosen and 10-fold cv was applied.

All genomic prediction analyzes were performed considering a total of 30,000 iterations and burn-in of the first 20,000 iterations. The prediction analyses were performed using an Intel Core i5 2.20GHz processor and 8.00 GB of RAM.

## Results

The components of genetic variance were significant in all environments, except E4: Sorriso – MT (Table 2). The non-significance of genetic variance in this environment can be explained by the high CVe (38.57%) and the low magnitude of the heritability (0.12) of the grain yield trait. The other environments presented acceptable CVe (27.88% to 35.62%) and heritability (0.36 to 0.52) estimates for a complex trait such as grain yield.

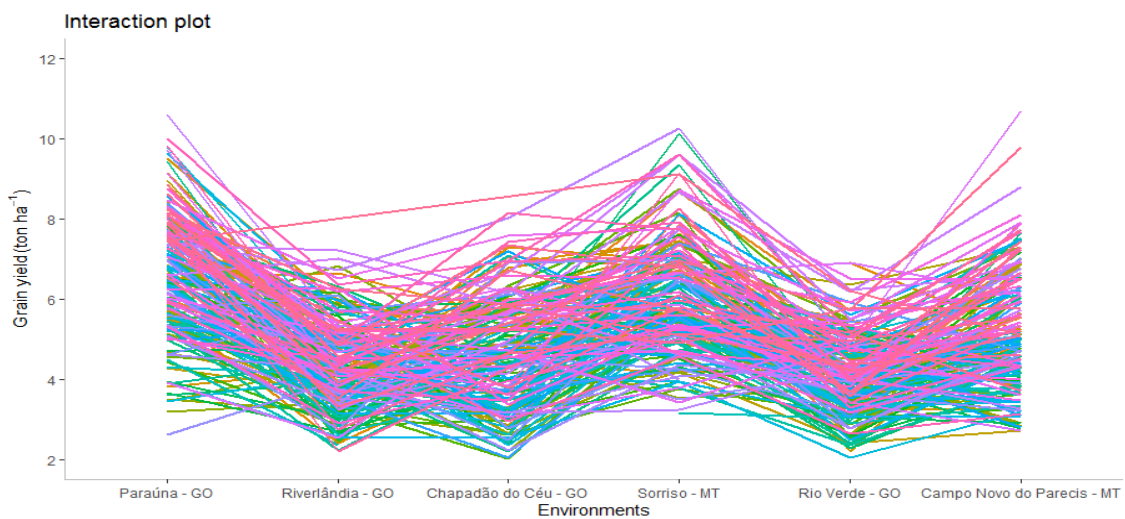
**Table 2.** Summary of the individual analysis of variance of the 128 maize hybrids evaluated across six environments during the safrinha of the 2012/2013 crop season.

Parameters	Environments					
	E1	E2	E3	E4	E5	E6
$\sigma_G^2$	0.90**	0.42**	1.02**	0.24 <sup>ns</sup>	0.42**	0.66**
$\sigma_e^2$	3.49	2.04	1.84	4.84	1.67	3.43
$H^2$ Cullis	0.43	0.37	0.52	0.12	0.41	0.36
CVe (%)	27.88	32.38	28.03	38.57	30.12	35.62



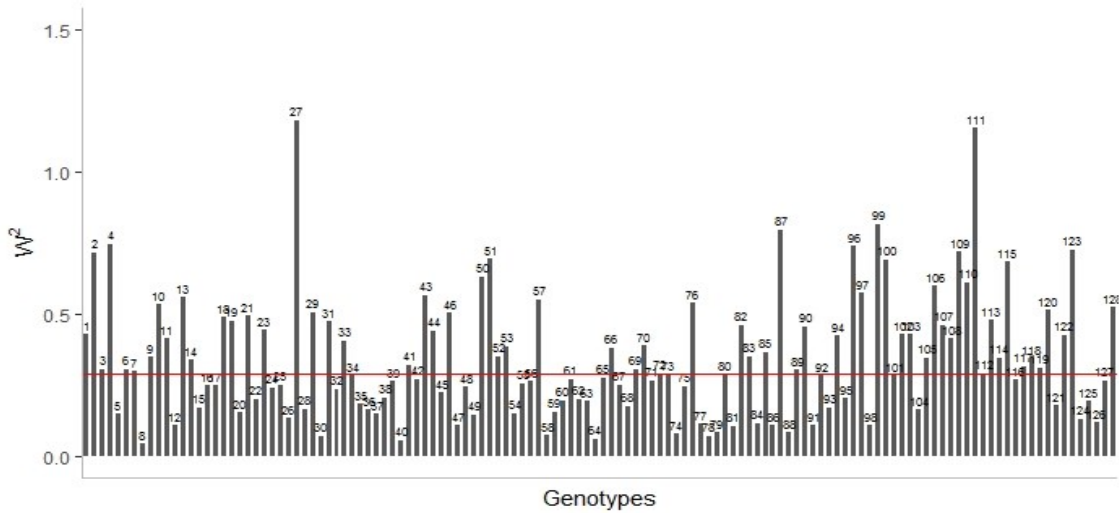
$\sigma_G^2$ : genetic variance;  $\sigma_e^2$ : residual variance;  $H^2$ : heritability; CVe: coefficient of variation

Using the GxE matrix, obtained through joint analysis, it was performed an interaction plot considering the six environments (Figure 1). The GxE interaction was quite present. Preliminarily, it was observed that Riverlândia, Chapadão do Céu, and Rio Verde were the unfavorable environments for most evaluated maize hybrids.



**Figure 1.** Genotype x Environment interaction plot.

The  $W^2$  estimates of each of the 128 hybrids were plotted in Figure 2. As expected, a large variation, in terms of grain yield stability, was observed. Considering the median of  $W^2$  estimates as a reference (red line), it is possible to identify the most stable (below the reference line) and less stable (above the reference line) genotypes (Figure 2).



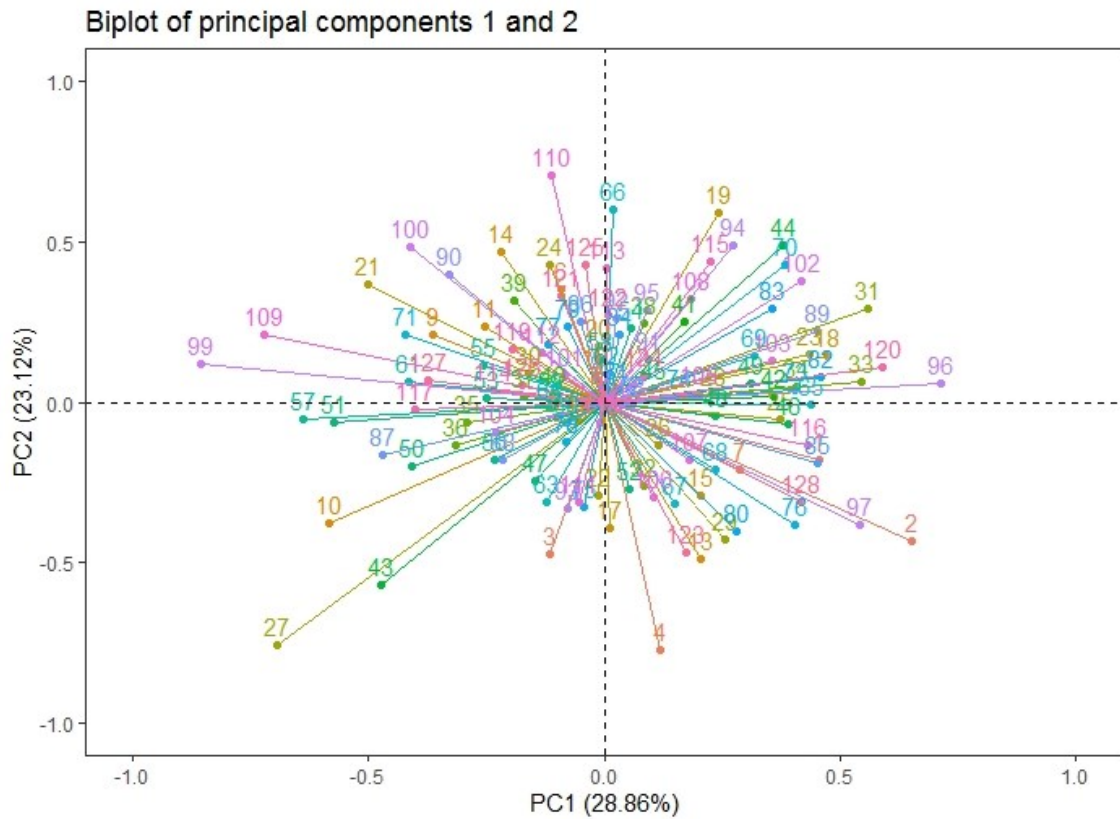
**Figure 2.** Estimates of Wricke's ecovalence ( $W^2$ ) index for each of the 128 evaluated maize hybrids.

Aiming to simplify the study of GxE, the dataset was submitted to Principal Component Analysis (PCA). The first four principal components explained 84.69% of the total variation presented in the dataset (Table 3).

**Table 3.** Magnitudes of singular values, percentage of contribution (%), and cumulative contribution of the singular values, obtained through Principal Component Analysis.

	Singular values	(%)	Cumulative (%)
$\lambda_1$	0.097948	28.86	28.86
$\lambda_2$	0.078457	23.12	51.98
$\lambda_3$	0.064027	18.87	70.85
$\lambda_4$	0.046960	13.84	84.69
$\lambda_5$	0.029811	8.78	93.47
$\lambda_6$	0.022190	6.53	100

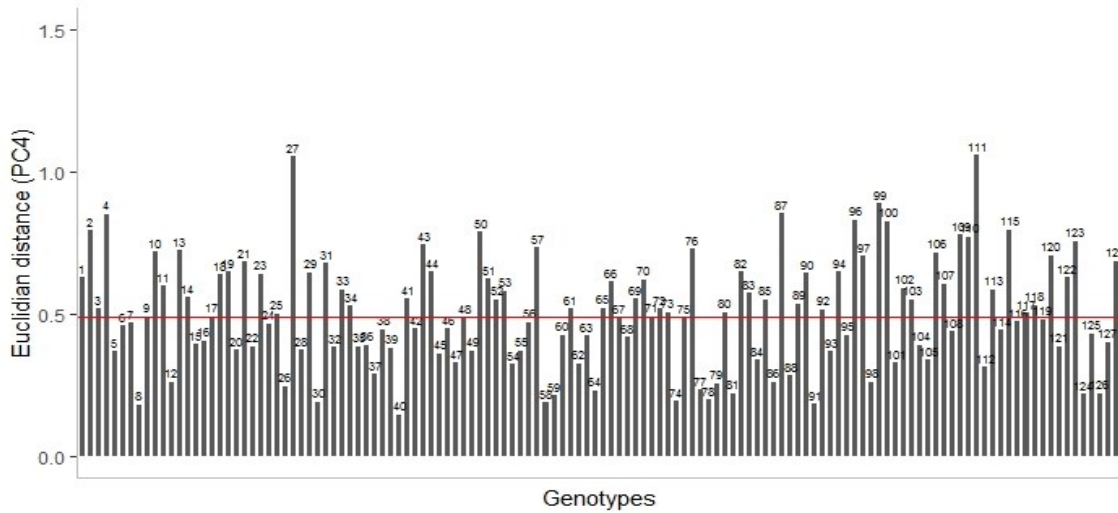
Although PC1 and PC2, explain, respectively, 28.86% and 23.12% of the observed variation, the scores of these components were plotted on the X and Y axes of a biplot, to simplify the interpretation of the results (Figure 3). The maize hybrids near the center of the biplot are classified as more stable than those that are far from the center of the biplot.



**Figure 3.** Biplot representation of principal components 1 and 2.

For the final estimates of the Euclidian distance of each hybrid, the norms of the vectors were calculated considering the scores of the first four principal components. The Euclidian distances of each hybrid are plotted in Figure 4.

The results were quite similar to Wricke's ecovalence. Again, considering the median as a reference, those genotypes that were below the red line are identified as more stable.



**Figure 4.** Estimates of the Euclidian distance, considering the first four principal components of the 128 evaluated maize hybrids.

Considering the 128 hybrids, the BRR model was applied to estimate the phenotypic and genotypic parameters. The component of phenotypic variance ranged from 0.002 to 0.257, for Var E and MHVG indices, respectively. The component of residual variance fluctuated from 0.002 (Var E) to 0.198 (MHVG) (Table 4).

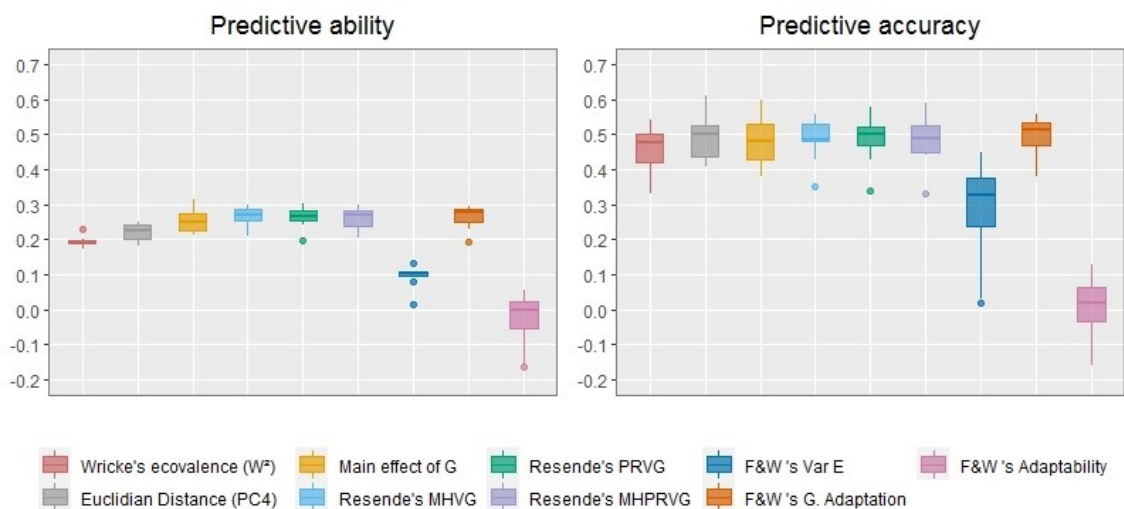
**Table 4.** Phenotypic and genotypic parameters associated with the several indices considering the 128 evaluated maize hybrids.

Index	Parameter				
	$\sigma_F^2$	$\sigma_e^2$	$H^2$	Corr.	Interval
Wricke's ecovalence ( $W^2$ )	0.045	0.037	0.18	0.63	0.51 – 0.72
Euclidian Distance (ED PC4)	0.036	0.029	0.19	0.65	0.54 – 0.74
The main effect of genotype (G)	0.186	0.141	0.25	0.69	0.59 – 0.77
Resendes' MHVG	0.257	0.198	0.23	0.67	0.56 – 0.76
Resendes' PRVG	0.010	0.007	0.25	0.69	0.59 – 0.77
Resendes' MHPRVG	0.010	0.007	0.25	0.69	0.59 – 0.77
F&W's Var E	0.002	0.002	0.12	0.60	0.48 – 0.70
F&W's Adaptability	0.013	0.011	0.14	0.65	0.53 – 0.74
F&W's General Adaptation	0.245	0.188	0.24	0.68	0.57 – 0.76

$\sigma_F^2$ : phenotypic variance;  $\sigma_e^2$ : residual variance;  $H^2$ : heritability; Corr: The Pearson correlation between observed and predicted values; Interval: Interval of the Pearson correlation.

The heritability estimates were low, with a peak of 0.25. In general, the heritability estimates of the stability indices ( $W^2$ , ED, and Var E) were the lowest, ranging from 0.12 to 0.19, for Var E and  $W^2$ , respectively. The Pearson correlations, considering the whole dataset were used as references for the cross-validation inferences (Table 4).

The results of the 10 rounds of 10-fold cross-validation were plotted in figure 5. The result of each cross-validation round and its average can be found in the supplementary material (Table 1 SM). Considering the references heritability and Pearson correlation presented in table 4, it is possible to consider that most of the indices presented a low to medium performance. In general, the predictive ability (PA) of the chosen indices fluctuated from 0.17 to 0.31, except for Var E and Adaptability indices, which presented poor performance.

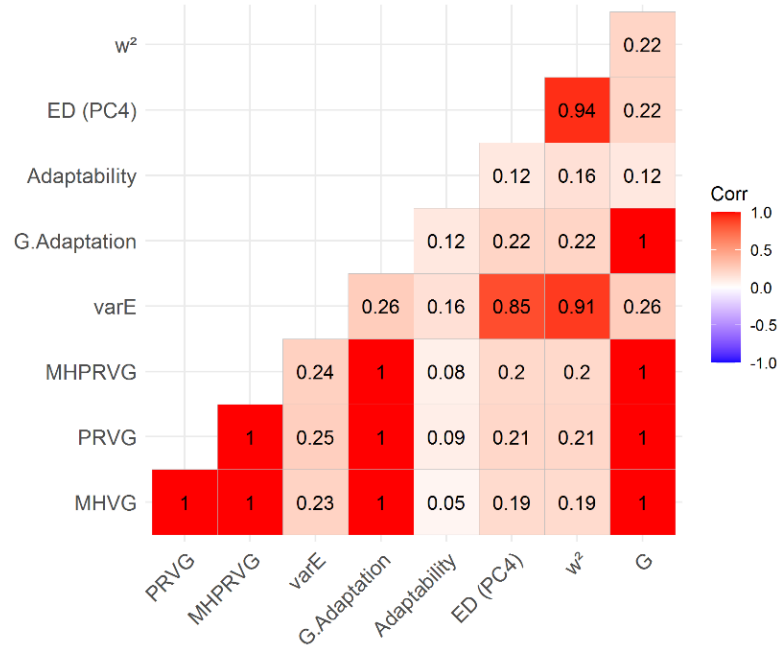


**Figure 5.** Boxplot representation of the predictive ability and accuracy of the indices, after 10 rounds of 10-fold cross-validation predictions.

Although the predictive accuracy (ACC) presented a similar pattern to PA, the estimates of ACC were higher, fluctuating between 0.33 and 0.61, except for Adaptability, in which the ACC was almost null.

To better understand the relationship among the nine indices, a correlation plot between the estimates of the indices is presented in figure 6. It was possible to cluster the nine indices

into three groups: Group 1: G, MHVG, PRVG, MHPRVG, G. Adaptation; Group 2:  $W^2$ , ED (PC4), Var E.; Group 3: Adaptability.



**Figure 6.** Correlation plot among the indices.

To make additional comparisons, genomic predictions were performed considering the Bayesian Multi-Environment model (BME). The predictive abilities when considering both effects: G + GxE were greater than GxE only (Table 5). In general, the predictive abilities achieved by using the BME were outperformed by indices predictions using the BRR model. On average, indices models outperformed BME by about 3.16%, 15.79%, 33.16%, 40.53%, 38.95%, 37.37%, and 38.95% while predicting  $W^2$ , ED (PC4), G, MHVG, PRVG, MHPRVG, G. Adaptation, respectively. However, Var E. and Adaptability indices' predictive abilities were about 50% and 115% lower than BME's average.

**Table 5.** Results of the Bayesian Multi-Environment model (BME) considering the main effect of genotypes plus GxE interaction (G + GxE), and the GxE interaction only (GxE).

Environment	G + GxE	GxE
E1 – Paraúna/GO	0.19	-0.01
E2 – Riverlândia/GO	0.23	0.07
E3 – Chapadão do Céu/GO	0.21	0.07
E4 – Sorriso/MT	0.11	-0.17
E5 – Rio Verde/GO	0.21	0.11
E6 – Campo Novo do Parecis/MT	0.20	0.04
Average	0.19	0.02

## Discussion

The significances of the components of genetic variance indicate the presence of genetic variability among the maize hybrids. This variability is expected due to the diverse backgrounds of the parental lines, as reported by Cantelmo et al. (2017a, 2017b).

The low estimates of heritability are mainly explained by the complexity of grain yield traits. Pereira et al. (2022) reported similar results when evaluating maize hybrids in summer season and safrinha conditions. In addition, safrinha maize faces more intense environmental challenges, which contributes to the lower estimates of heritability and higher CVe.

The GxE interaction was quite severe, as noted in the interaction plot (Figure 1). Such a scenario represents a complicator factor for genotypes selection and recommendation. In cases of significant GxE, three strategies might be applied: *i*) to ignore; *ii*) capitalize, or *iii*) minimize the GxE.

Although exploiting GxE seems interesting, too specific genotype recommendations can excessively inflate the hybrids portfolio, which might increase seeds production costs, due to the necessity of maintaining a large portfolio; therefore, the identification of high stable genotypes, that minimizes the GxE, is a widely adopted strategy.

Over the years, several methodologies were developed to study genotypes' stability. In this work, four well-consolidated methodologies were used: Ecovalence (Shojaei et al., 2021), Euclidian Distance (Annicchiarico, 1997; Oliveira et al., 2014; Yamamoto et al., 2021), Resende's indices (Mendes et al., 2012; Oliveira et al., 2017; Yamamoto et al., 2021), Finlay-Wilkinson (Malosetti; Ribaut; Eeuwijk, 2013; Eeuwijk; Bustos-Korts; Malosetti, 2016; Lozada and Carter, 2020).

Figures 2 and 4, in addition to table 4, indicate that both Ecovalence ( $W^2$ ) and Euclidian Distances (ED) indices achieved similar results. Although some variation for both indices was observed in this dataset, the principal component analysis was not very effective. In comparison to  $W^2$ , the PCA reduced from 6 to 4 main components, and just increased the heritability from 0.18 to 0.19, and the Pearson correlation from 0.63 to 0.65. In other words, the PCA reduces the dataset noise but increases the heritability and the predictive ability by only 5% and 3%, respectively, however, the ED could outperform  $W^2$  in a larger dataset.

Due to the complexity of working only with the GxE matrix, additional indices, that also consider the genetic merit of the maize hybrids per se, were used. In such a scenario, Resende's indices were chosen. Since the harmonic mean penalizes large deviations, the harmonic mean of the breeding values (MHVG) can be used as an index of the genotype's stability, while the relative performance of the breeding values (PRVG) indicates the genotype's adaptability. Therefore, the harmonic mean of the relative performance of the breeding values represents both genotype's stability and adaptability in a single index: the MHPRVG.

Considering Resende's methodology, the results of MHVG, PRVG, and MHPRVG were similar to those achieved by using the main effect of genotypes (G). There are several reports on the benefits of Resende's indices in the literature (Resende, 2002, 2004; Oliveira et al., 2017, Mendes et al., 2012). Although predicting G could be considered an easier strategy



in this case, the focus of this work is to predict the stability, and in large datasets (in terms of genotypes and environments), the MHVG and MHPRVG indices should be a more adequate approach. Indeed, after 10 rounds of 10-fold cross-validation, MHVG, PRVG, and MHPRVG outperformed G by about 5.53%, 4.35%, and 3.16%, respectively.

The results of predicting Finlay-Wilkinson's indices were more diversified. While the general adaptation index approximated to the G results, Var E and adaptability indices were unique, being the worst in terms of heritability: 0.12 (Var E) and 0.14 (Adaptability), and predictive ability 0.10 (Var E) and -0.03 (Adaptability), on average, respectively. However, Lozada and Carter (2020) achieved good predictive ability (0.46 to 0.66) while predicting Finlay-Wilkinson regression coefficients in agronomic traits of 456 winter wheat lines. In our work, a relatively small dataset was used (only 128 maize hybrids), which might contribute to the low predictive ability of Var E and Adaptability indices.

The estimates of predictive accuracy were lower than those presented in Gezan et al. (2017), in which the BRR model was applied to predict different traits in strawberries. The same authors suggest that predictive ability and accuracies might increase as the training population size increases. In addition, Sánchez and Akdemir (2021) highlight that the larger training population sets the better prediction, and the genetic relationship between training and the testing population is required to obtain high accuracies.

Clustering the indices into the three groups was efficient to better visualize the indices relationship, and to optimize the genotypes stability prediction approach by reducing the number of indices to be predicted or estimated. In this dataset, all indices that consider the performance of the hybrid per se could be summarized by predicting G. However, as highlighted by Resende (2004), the MHPRVG index accounts for both genotype's stability and adaptability, then, the MHPRVG could outperform G in other scenarios, such as larger datasets.

Therefore, the MHPRVG index could be prioritized in future works involving the whole dataset: 2770 maize hybrids evaluated in 50 locations in Brazil (Pereira et al., 2022).

Although there are no alternative indices for adaptability, the Var E. index might be replaced by predicting  $W^2$  or ED (PC4) indices, which presented more satisfactory prediction abilities. The  $W^2$  and ED present similar performances considering the stability indices (group 2). However, in larger datasets PCA followed by ED could surpass Wricke's methodology. In addition, considering our dataset, the Adaptability index (group 3) should be disregarded due to its poor and non-consistent performance.

Finally, one main question remains: Which strategy should be applied? *i*) to predict the genotype's stability index directly, or *ii*) a two-step approach (GxE predictions followed by stability index estimation)?

To answer this question, a Bayesian Multi-Environment (BME) prediction approach was applied. Once again, working only with the GxE matrix is challenging, and the results were far from acceptable. However, predictive abilities greatly increased when working with G+GxE.

Seven out of the nine indices outperformed the BME predictions, which indicates that indices' direct prediction might be more efficient than a two-step approach.

It is important to highlight that the computational cost of performing BME was up to five hours, considering 128 maize hybrids, six environments, and 10-fold cross-validation. On the other hand, for each index, the predictions took less than 30 min for each cross-validation round.

## **Conclusion**

Among the indices, Wricke's ecovalence ( $W^2$ ), and the main effect of genotypes (G) were enough to predict maize genotypes' stability and agronomic performance, respectively. However, in larger datasets, the Euclidian Distance (ED) and the MHPRVG indices should be more adequate.

The prediction abilities of the  $W^2$  and ED indices were computationally efficient and outperformed the results of the Bayesian Multi-Environment (BME) model.

Although additional studies, using a larger and unbalanced dataset are required, the results demonstrated that predicting the stability index directly could be a simpler and more efficient strategy than a two-step approach: performing GxE genomic predictions followed by the estimation of the stability indices.

## References

- 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. [10.1007/s00122-017-2898-0](https://doi.org/10.1007/s00122-017-2898-0)
- Andrea MCdaS, Boote KJ, Sentelhas PC, Romanelli TL (2018) Variability and limitations of maize production in Brazil: Potential yield, water-limited yield and yield gaps. *Agricultural Systems*, 165: 264-273. [10.1016/j.agsy.2018.07.004](https://doi.org/10.1016/j.agsy.2018.07.004)
- Annicchiarico P (1992) Cultivar adaptation and recommendation from alfalfa trials in Northern Italy. *Journal of Genetics and Breeding*, 46: 269-269.
- Annicchiarico P (1997) Joint regression vs AMMI analysis of genotype-environment interactions for cereals in Italy. *Euphytica*, 94: 53-62. <https://doi.org/10.1023/A:1002954824178>

- Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic prediction of breeding values when modeling genotype x environment interaction using pedigree and dense molecular markers. *Crop Science*, 55: 707-719. [10.2135/cropsci2011.06.0299](https://doi.org/10.2135/cropsci2011.06.0299)
- Cantelmo NF, Von Pinho RG, Balestre M (2017a) Genomic analysis of maize lines introduced in the early stages of a breeding programme. *Plant Breeding*, 136(6): 845–860. [10.1111/pbr.12543](https://doi.org/10.1111/pbr.12543)
- Cantelmo NF, Von Pinho RG, Balestre M (2017b) Genome-wide prediction for maize single-cross hybrids using the GBLUP model and validation in different crop seasons. *Molecular Breeding*, 37: 1–13. [10.1007/s11032-017-0651-7](https://doi.org/10.1007/s11032-017-0651-7)
- Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. *Journal of Agricultural, Biological, and Environmental Statistics*, 11: 381-393. <https://doi.org/10.1198/108571106x154443>
- Dias KOG, Gezan SA, Guimarães CT, Parentoni SN, Guimarães PEDO, Carneiro NP, Portugal EAB, Cardoso MJ, Anoni CO, Magalhães JV, Souza JC, Guimarães LJM, Pastina MM (2018) Estimating genotype × environment interaction for and genetic correlations among drought tolerance traits in maize via factor analytic multiplicative mixed models. *Crop Science*, 58(1): 72-83. [10.2135/cropsci2016.07.0566](https://doi.org/10.2135/cropsci2016.07.0566)
- Dos Santos JPR, Fernandes SB, McCoy S, Lozano R, Brown PJ, Leakey ADB, Buckler ES, Garcia AAF, Gore MA (2020) Novel Bayesian networks for genomic prediction of developmental traits in biomass sorghum. *G3 Genes| Genomes| Genetics*, 10: 769-781. <https://doi.org/10.1534/g3.119.400759>
- Eberhart SA, Russel WA (1966) Stability parameters for comparing varieties. *Crop Science*, 6(1): 36-40. <https://doi.org/10.2135/cropsci1966.0011183X000600010011x>

- Eeuwijk FA, Bustos-Korts DV, Malosetti M (2016) What should students in plant breeding know about the statistical aspects of genotype x environment interactions? *Crop Science*, 56: 2119-2140. 10.2135/cropsci2015.06.0375
- Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome*, 4: 250-255. <https://doi.org/10.3835/plantgenome2011.08.0024>
- Finlay KW, Wilkinson GN (1963) The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research*, 14(6): 742-754. 10.1071/AR9630742
- Gezan SA, Osorio LF, Verma S, Whitaker VM (2017) An experimental validation of genomic selection in octoploid strawberry. *Horticulture Research*, 4(16070): 1-9. <https://doi.org/10.1038/hortres.2016.70>
- Gollob HF (1968) A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika*, 33: 73–115. 10.1007/BF02289676
- Hamazaki K, Iwata H (2022) Bayesian optimization of multivariate genomic prediction models based on secondary traits for improved accuracy grains and phenotyping costs. *Theoretical and Applied Genetics*, 135: 35-50. <https://doi.org/10.1007/s00122-021-03949-1>
- Jarquín D, de Leon N, Romay C, Bohn M, Buckler ES, Ciampittis I, Edwards J, Ertle D, Flint-García S, Gore MA, Graham C, Hirsch CN, Holland JB, Hooker D, Kaeppler SM, Knoll J, Lee EC, Lawrence-Dill CJ, Lynch JP, Moose SP, Murray SC, Nelson R, Rocheford T, Schnable JC, Schnable PS, Smith M, Springer N, Thomison P, Tuinstra M, Wisser RJ, Xu W, Yu J, Lorenzi A (2021) Utility of climatic information via combining ability models to improve genomic prediction for yield within the genomes to fields maize project. *Frontiers in Genetics*, 11: 1-11. 10.3389/fgene.2020.592769
- Lozada DN, Carter AH (2020) Insights into the genetic architecture of phenotypic stability traits in winter wheat. *Agronomy*, 10(368): 1-15. 10.3390/agronomy10030368

- Luna-Vázquez FJ, Toledo FH, Montesinos-López O, Montesinos-López A, Crossa J (2020) BMTME: Bayesian multi-trait multi-environment for genomic selection analysis.
- Malosetti M, Ribaut J-M, Eeuwijk FA (2013) The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. *Frontiers in physiology*, 4(44): 1-17. 10.3389/fphys.2013.00044
- Mandel J (1969) The partitioning of interactions in analysis of variance. *Journal of Research of the National Bureau of Standards*, 73: 309-328. 10.6028/JRES.073B.031
- Mandel J (1971) A new analysis of variance model for non-additive data. *Technometrics*, 13(1): 1-18. <https://doi.org/10.2307/1267072>
- Mendes FF, Guimarães LJM, Souza JC, Guimarães PEO, Pacheco CAP, Machado JRdeA, Meirelles WF, Silva AR, Parentoni SN (2012) Adaptability and stability of maize varieties using mixed model methodology. *Crop Breeding and Applied Biotechnology*, 12: 111-117. <https://doi.org/10.1590/S1984-70332012000200003>
- Montesinos-López O, Montesinos-López A, Crossa J, Cuevas J, Montesinos-López JC, Gutiérrez ZS, Lillemo M, Philomin J, Singh R (2019a) A Bayesian genomic multi-output regressor stacking model for predicting multi-trait multi-environment plant breeding data. *G3 Genes| Genomes| Genetics*, 9(3381): 1-13.
- Montesinos-López O, Montesinos-López A, Luna-Vázquez FJ, Toledo FH, Pérez-Rodríguez P, Lillemo M, Crossa J (2019b) An R package for Bayesian analysis of multi-environment and multi-trait multi-environment data for genome-based prediction. *G3 Genes| Genomes| Genetics*, 9(5): 1355-1369. 10.1534/g3.119.400126
- Oliveira RL de, Von Pinho RG, Ferreira DF, Pires LPM, Melo WMC (2014) Selection index in the study of adaptability and stability in maize. *The Scientific World Journal*, 2014(360570): 1-6. <http://dx.doi.org/10.1155/2014/360570>

- Oliveira IJde, Atroch AL, Dias MC, Guimarães LJ, Guimarães PEdeO (2017). Selection of corn cultivars for yield, stability and adaptability in the state of Amazonas, Brazil. *Pesquisa Agropecuária Brasileira*, 52(6): 455-463. 10.1590/S0100-204X2017000600009
- Pereira F de C, Ramalho MAP, Resende Junior MFR de, Von Pinho RG (2022) Mega-environment analysis of maize breeding data from Brazil. *Scientia Agricola*, 79(2): 1-10. <http://doi.org/10.1590/1678-992X-2020-0314>
- Pérez P, de los Campos G (2014) Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics*, 198(2): 483-1495. 10.1534/genetics.114.164442
- R Development Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ramalho MAP, Abreu FFB, Santos JB dos, Nunes JAR (2012) Aplicações da genética quantitativa no melhoramento de plantas autógamas. Lavras: Ed. UFLA. 522 p.
- Resende MDV (2002) Genética biométrica e estatística no melhoramento de plantas perenes. Brasília: Embrapa Informação Tecnológica; Colombo: Embrapa Florestas. 975p.
- Resende MDV (2004) Métodos Estatísticos Ótimos na Análise de Experimentos de Campo. Embrapa Florestas. Colombo. 65 p. (Documentos 100).
- Sánchez JI y, Akdemir D (2021) Training set optimization for sparse phenotyping in genomic selection: a conceptual overview. *Frontiers in Plant Science*, 12: 1-14. 10.3389/fpls.2021.715910
- Shojaei SH, Mostafavi K, Lak A, Omrani A, Omrani S, Mousavi SMN, Illés A, Bojtor C, Nagy J (2021) Evaluation of stability in maize hybrids using univariate parametric methods. *Journal of Crop Science and Biotechnology*, 1-8. <https://doi.org/10.1007/s12892-021-00129-x>

- Silveira LS, Martins Filho S, Azevedo CF, Barbosa EC, Resende MDV, Takahashi EK (2019) Bayesian models applied to genomic selection for categorical traits. *Genetics and Molecular Research*, 18(4): 1-10. <http://dx.doi.org/10.4238/gmr18490>
- VanRaden PM (2008) Efficient methods to compute genomic predictions. *Journal of dairy science*, 91(11): 4414-4423. [10.3168/jds.2007-0980](https://doi.org/10.3168/jds.2007-0980)
- Volpato L, Bernardeli A, Gomez F (2021) Genomic selection with rapid cycling: Current insights and future prospects. *Crop Breeding and Applied Biotechnology*, 21: 1-8. <http://dx.doi.org/10.1590/1984-70332021v21Sa27>
- Von Pinho RG, Silva EVV, Oliveira TL (2022) Challenges of maize breeding under tropical conditions of Brazil. *Revista Brasileira de Milho e Sorgo*, 21(e1258): 1-25. <https://doi.org/10.18512/rbms2022vol21e1258>
- Wang T, Chen Y-PP, Goddard ME, Meuwissen THE, Kemper KE, Hayes BJ (2015) A computationally efficient algorithm for genomic prediction using a Bayesian model. *Genetics Selection Evolution*, 47(34): 1-16. [10.1186/s12711-014-0082-4](https://doi.org/10.1186/s12711-014-0082-4)
- Wang T, Chen Y-PP, MacLeod IM, Pryce JE, Goddard ME, Hayes BJ (2017) Application of a Bayesian non-linear model hybrid scheme to sequence data for genomic prediction and QTL mapping. *BMC Genomics*, 18(618): 1-23. [10.1186/s12864-017-4030-x](https://doi.org/10.1186/s12864-017-4030-x)
- Wang N, Wang H, Zhang A, Liu Y, Yu D, Hao Z, Ilut D, Glaubitz JC, Gao Y, Jones E, Olsen M, Li X, Vicente FS, Prasanna BM, Crossa J, Pérez-Rodríguez P, Zhang X (2020) Genomic prediction across years in a maize double haploid breeding program to accelerate early-stage testcross testing. *Theoretical and Applied Genetics*, 133: 2869-2879. <https://doi.org/10.1007/s00122-020-03638-5>
- Wricke G (1965) Zur berechnung der okovalenz bei sommerweizen und hafer. *Z. Pflanzenzuchtg*, 52: 127-138.



- Yamamoto ELM, Gonçalves MC, Davide LMC, Santos A dos, Candido LS (2021) Adaptability and stability of maize genotypes in growing regions of central Brazil. *Revista Ceres*, 68(3): 201-211. 10.1590/0034-737X202168030006
- Yan W, Hunt LA, Sheng Q, Szlavnic Z (2000) Cultivar evaluation and megaenvironment investigation based on GGE biplot. *Crop Science*, 40(3): 597-605. <https://doi.org/10.2135/cropsci2000.403597x>

## Supplementary material

Table 1 SM. Results of the ten rounds of 10-fold cross-validation of indices genomic predictions using the BRR model.

	W <sup>2</sup>	ED	G	MHVG	PRVG	MHPRVG	VAR E	ADAP	G. ADAP
Round 1									
Corr.	<b>0.23</b>	<b>0.25</b>	<b>0.28</b>	<b>0.29</b>	<b>0.28</b>	<b>0.28</b>	0.13	0.00	<b>0.29</b>
ACC	0.54	0.61	0.60	0.54	0.54	0.53	0.45	0.00	0.54
Round 2									
Corr.	<b>0.19</b>	<b>0.19</b>	<b>0.25</b>	<b>0.25</b>	<b>0.26</b>	<b>0.24</b>	0.10	0.03	<b>0.27</b>
ACC	0.50	0.41	0.45	0.48	0.49	0.45	0.40	0.06	0.52
Round 3									
Corr.	<b>0.20</b>	<b>0.21</b>	<b>0.21</b>	<b>0.30</b>	<b>0.30</b>	<b>0.30</b>	0.11	-0.05	<b>0.30</b>
ACC	0.54	0.49	0.39	0.56	0.58	0.59	0.33	-0.14	0.56
Round 4									
Corr.	<b>0.17</b>	<b>0.18</b>	<b>0.22</b>	<b>0.26</b>	<b>0.26</b>	<b>0.26</b>	0.10	0.04	<b>0.25</b>
ACC	0.41	0.41	0.38	0.43	0.43	0.44	0.32	0.08	0.43
Round 5									
Corr.	<b>0.20</b>	<b>0.24</b>	<b>0.31</b>	<b>0.28</b>	<b>0.28</b>	<b>0.28</b>	0.11	<b>-0.17</b>	<b>0.29</b>
ACC	0.45	0.53	0.60	0.48	0.47	0.48	0.26	-0.48	0.51
Round 6									
Corr.	<b>0.20</b>	<b>0.24</b>	<b>0.28</b>	<b>0.29</b>	<b>0.29</b>	<b>0.29</b>	0.10	-0.16	<b>0.28</b>
ACC	0.50	0.57	0.54	0.49	0.51	0.50	0.23	-0.48	0.47
Round 7									
Corr.	<b>0.20</b>	<b>0.23</b>	<b>0.25</b>	<b>0.24</b>	<b>0.24</b>	<b>0.24</b>	0.08	0.00	<b>0.23</b>
ACC	0.46	0.49	0.50	0.53	0.52	0.53	0.18	0.00	0.54
Round 8									
Corr.	<b>0.19</b>	<b>0.20</b>	<b>0.25</b>	<b>0.29</b>	<b>0.28</b>	<b>0.28</b>	0.01	0.06	<b>0.29</b>
ACC	0.49	0.51	0.48	0.53	0.52	0.52	0.03	0.13	0.52
Round 9									
Corr.	<b>0.18</b>	<b>0.24</b>	<b>0.22</b>	<b>0.26</b>	<b>0.25</b>	<b>0.24</b>	0.11	-0.05	<b>0.25</b>
ACC	0.39	0.51	0.42	0.48	0.47	0.45	0.39	-0.16	0.47
Round 10									
Corr.	<b>0.20</b>	<b>0.22</b>	<b>0.26</b>	<b>0.21</b>	<b>0.20</b>	<b>0.20</b>	0.10	0.01	<b>0.19</b>
ACC	0.33	0.42	0.48	0.35	0.34	0.33	0.34	0.04	0.38
Average									
Corr.	0.20	0.22	0.25	0.27	0.26	0.26	0.10	-0.03	0.26
ACC	0.46	0.50	0.48	0.49	0.49	0.48	0.33	-0.10	0.49

**Bold** values indicate significance

**ARTIGO 2 – GENOMIC PREDICTION FOR MAIZE GRAIN YIELD STABILITY  
UNDER MULTI-ENVIRONMENT SCENARIOS**

Artigo redigido conforme as normas da revista *Euphytica*  
(International Journal of Plant Breeding).

## Genomic prediction for maize grain yield stability under multi-environment scenarios

Eric Vinicius Vieira Silva

**Abstract** – The genotype by environment interaction (GxE) plays an important role in maize breeding, and its effects are even more complex considering the second season in Brazil, therefore, selecting highly stable hybrids is crucial to overcome the challenges of tropical agriculture. Genomic prediction (GP) has become an important tool for plant breeding, and it was gradually integrated into many breeding pipelines. Reports highlight that GP accuracies can be improved by modeling the GxE in a multi-environment (ME) approach. The objective of this work was to study different genomic prediction strategies considering maize yield stability and multi-environment scenarios. 1180 hybrids were evaluated for grain yield in 12 environments during the second season of the 2012/2013 and 2013/2014 crop years, however, both crop years were used separately. Bayesian Ridge Regression genomic predictions were performed considering: *i*) single-environment (SE) predictions; *ii*) stability index predictions; *iii*) ME prediction without including GxE, and *iv*) including GxE effects. Two cross-validations were used: CV1 (10-fold), and CV2 (whole environment predictions, ME only). On average, the predictive ability (PA) for SE was 0.25. For the Euclidian Distance index, the PAs were - 0.079 (2012/2013) and 0.16 (2013/2014), while the MHPRVG index achieved 0.53 (2012/2013) and 0.51 (2013/2014) of predictive abilities. Considering CV1, no gain was observed by including the GxE in the 2012/2013 ME approach (PA 0.56), for 2013/2014 marginal gain was observed (from 0.75 to 0.76). CV2 present very poor predictability. Predicting the MHPRVG might be efficient. The inclusion of GxE did not justify its higher computational cost. Additional studies using better-designed datasets are required.

**Keywords:** *Zea mays*; Genotypes x environments interaction; multi-environment genomic prediction; Stability index.

## Introduction

Brazil's tropical condition is very challenging in terms of biotic and abiotic stress. It is well-known that second season (or safrinha) maize often faces drought conditions and higher insect pressures, resulting in yield losses (Andrea et al., 2018). The genotype by environment interaction (GxE) plays a crucial role in maize breeding, and due to the low predictability over the environments and crop years, the GxE is even more complex considering second season maize (Von Pinho; Silva; Oliveira, 2022).

Since its first proposal (Meuwissen et al., 2001), genomic prediction (GP) methods have become an important tool in plant breeding pipelines. The use of GP does not only shorten the time required to complete a breeding cycle (Heslot et al., 2014) but also allows to increase the number of genotypes under studies without significantly increasing phenotyping costs, in other words, GP tools support the breeder with useful information regarding the populations or hybrids that should be synthesized and tested. Indeed, Beyene et al. (2019) reported up to 32% of cost reductions by including genomic selection in tropical maize breeding pipeline without reducing the genetic gains.

Due to its relevance, many researchers have studied different GP approaches, such as the inclusion of non-additive effects (Dias et al., 2018), the optimization of training populations (Atanda et al., 2021; Fristche-Neto et al., 2021), the GxE modeling (Burgueño et al., 2012; Acosta-Pech et al., 2017; Dias et al., 2018, Wang et al., 2020; Magento et al., 2020), the inclusion of environmental covariables (Gillberg et al., 2019; Costa-Neto et al., 2021b; Jarquín et al., 2021), deep learning (Montesinos-López et al., 2021). However, none of them have tried to directly predict maize grain yield stability.

Rezende et al. (2021) define stability as the genotype's ability to have highly predictable behavior regardless of environmental changes. In other words, stable genotypes have a similar

performance across locations and over the years, which is a highly desirable characteristic considering maize hybrid development for second season environments (Von Pinho; Silva; Oliveira, 2022).

To better study the GxE and the stability, multi-environmental trials (MET) are required. The METs represent a high cost in the breeding pipeline budget. Considering traditional least square methods (i.e., traditional analysis of variance), all the genotypes must be evaluated in all the environments, otherwise, the analysis may be affected. Such a scenario increases the phenotyping costs. In this sense, the mixed model approach is flexible, and better handling unbalanced trials, among other advantages (Eeuwijk; Bustos-Korts; Malosetti, 2016). In addition, the Bayesian approach can also be used for GxE studies, which is a more parsimonious, refined, informative, and powerful approach, allowing the incorporation of prior information (Montesinos-López et al., 2022).

In this scenario, multi-environment GP is an efficient approach for genotypes selection as well as to better understand the GxE and the target environments. However, well-designed training/testing populations (Atanda et al., 2021; Fristche-Neto et al., 2021), in addition to well-planned METs are essential to achieve good prediction accuracies (Jarquín et al., 2020).

The objective of this work was to study different genomic prediction strategies considering maize grain yield stability and multi-environment scenarios.

## **Materials and methods**

### **Phenotypic data**

A dataset of 1180 maize hybrids were evaluated for grain yield ( $\text{t ha}^{-1}$ ) in 12 different environments during the second season of the 2012/2013 and 2013/2014 crop years (Table 1).

The maize hybrids were evaluated under incomplete block designs, with a variable number of replicates, in 5-m-wide four rows spaced 0.5 m (between rows) plots.

**Table 1.** Grain yield means and coefficient of variation of maize single-cross hybrids cultivated across 12 environments during the second season of the 2012/2013 and 2013/14 crop years.

Env.	Location	Year	GY mean (t ha <sup>-1</sup> )	N. of genotypes
E1	Campo Novo do Parecis – MT	2012/13	5.08	308
E2	Chapadão do Céu – GO	2012/13	4.26	227
E3	Paraúna – GO	2012/13	6.74	309
E4	Rio Verde – GO	2012/13	4.14	294
E5	Riverlândia – GO	2012/13	4.59	292
E6	Sorriso – MT	2012/13	6.10	307
E7	Campo Novo do Parecis – MT	2013/14	5.60	709
E8	Lucas do Rio Verde – MT	2013/14	4.13	709
E9	Primavera do Leste – MT	2013/14	6.71	709
E10	Rio Verde – GO	2013/14	2.63	709
E11	Sapezal – MT	2013/14	7.58	710
E12	Sorriso – MT	2013/14	5.04	710

Env.: Environment; GY: Grain Yield

It is important to highlight that the dataset accounts for a lines-introgression maize program, and it was kindly provided by a private breeding company. Further information and genotypes' characterization can be found in Cantelmo et al. (2017).

### Genotypic data

23153 DArTs (Diversity Arrays Technology company) markers were generated for each parental line. The maize hybrids DArTs panel was constructed by the combination of the parental lines. The markers were filtered considering 5% of MAF (Minor Allele Frequency) and marker imputation was performed considering the markers' mean. Markers with more than 50% of missing data were removed. The genomic matrix was constructed using VanRaden's method (VanRaden, 2008) presented in the AGHmatrix R-package (Amadeu et al., 2016). Only hybrids that possessed both phenotypic and genotypic data were considered.

## Phenotypic data analysis

The statistical analyses were performed in two stages. First: to adjust field sources of variation, single-environment analyses were performed, then second: multi-environmental analyses were performed to obtain genotype by environment matrix. In the first stage, the following model was used to obtain the BLUEs (Best Linear Unbiased Estimators) of each hybrid in each environment:

$$Y = X\beta + Z_b u_b + \varepsilon$$

where  $Y$  is the vector of the phenotypic data;  $\beta$  is the vector of fixed effects of the genotypes and the replicates;  $X$  is the incidence matrix associated with the fixed of the genotype and the replicates;  $u_b$  is the vector of the random effect of the blocks nested within the replicates, with  $u_b \sim N(0, I_{u_b} \sigma_{u_b}^2)$ ;  $Z_b$  is the incidence matrix associated with the vector  $u_b$ ; and  $\varepsilon$  is the vector of residuals, with  $\varepsilon \sim N(0, I_\varepsilon \sigma_\varepsilon^2)$ ;  $\sigma_{u_b}^2$ , and  $\sigma_\varepsilon^2$  are the components of variance of the blocks within replicates and the residuals, respectively, while  $I_{u_b}$ , and  $I_\varepsilon$  are the identity matrices associated with  $u_b$ , and  $\varepsilon$ , respectively.

In the second stage, the following model was used to obtain the BLUEs-by-Env matrix:

$$y_{ij} = \mu + g_i + e_j + ge_{ij} + \varepsilon_{ij}$$

Where  $y_{ij}$  is the BLUEs-by-env data, which were obtained in the first stage;  $\mu$  is the constant associated with all observations;  $g_i$  is the fixed effect of genotypes;  $e_j$  is the fixed effect of environments;  $ge_{ij}$  is the fixed effect of the genotype x environment interaction; and  $\varepsilon_{ij}$  the random effect of the residuals, with  $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ , and is confounded with  $ge_{ij}$ .

Using the BLUEs-by-Environments data frame, principal component analyses (PCA) were performed using the `factorextra` R-package (Kassambara and Mundt, 2020). Then, the



Euclidian distance (ED) stability index was calculated from the norm of each maize hybrid's scores:

$$ED_i = \sqrt{\sum_{k=1}^n PC_{ki}^2}$$

where:  $ED_i$  is the Euclidian distance of the  $i$ th maize hybrid;  $PC_{ki}$  is the score of the principal component  $k$  for the  $i$ th maize hybrid, being  $k = 1, e n = \min(\text{row}, \text{column})$  correspondent to the rank of the GxE matrix.

Using the methodology proposed by Resende (2002; 2004), the MHPRVG index (harmonic mean of the relative performance of the breeding values) was also estimated. MHPRVG is an alternative index that accounts for both genotype's stability and adaptability. This index can be estimated as follows:

$$MHPRVG_i = \frac{n}{\sum_{j=1}^n \frac{Vg_{ij}}{l_j}}$$

where:  $n$  is the number of environments where the  $i$ th maize hybrid was evaluated;  $Vg_{ij}$  is the breeding value of the  $i$ th maize hybrid in the  $j$ th environment; and  $l_j$  is the overall mean of the  $j$ th environment.

Finally, the BLUEs-by-environments as well as the stability indices of each maize hybrid were used for genomic prediction purposes.

### **Genomic prediction models**

The genomic predictions were performed via software R (R Core Team, 2022). The Bayesian Ridge Regression (BRR) model was used from the BGLR package (Perez and De Los Campos, 2014). The BRR model assumes Gaussian distribution with  $\mu = 0$  and  $\sigma_{\beta}^2 \sim X^{-2}$  as a

prior of the model, and any further model details are presented in Perez and De Los Campos (2014).

Four strategies were adopted for genomic prediction purposes: *i*) single-environment predictions; *ii*) directly predicting the stability indices (ED and MHPRVG); *iii*) predicting maize hybrids BLUEs by environments without the GxE interaction; and *iv*) considering the GxE information in the prediction models.

For the regression function, three linear predictor models were used:

***M1*** – G: The main effect of genotypes (**G**) only, used for single-environment and stability indices predictions.

***M2*** – G + E: The main effect of genotypes (**G**) plus the main effect of environments (**E**), being E assumed as fixed.

***M3*** – G + E + GxE: The main effect of genotypes (**G**) plus the main effect of environments (**E**) plus the **GxE** effect, being E assumed as fixed.

***M1*** was used for single-environment and stability index predictions, while ***M2*** and ***M3*** were used for multi-environments predictions.

A simple 10-fold cross-validation (cv) was used for single-environment and stability indices, however, for strategies *iii* and *iv*, two scenarios were simulated: **a**) CV1: random 10-fold cv considering all environments; and **b**) CV2: predicting whole environments. Ten rounds of cross-validation were performed. For each round, the predictive ability was estimated by the Pearson correlation between predicted and observed values. The cross-validation schemes are represented in Figure 1.

Random 10-fold cv – round 1			CV1 – round 1						CV2 – round 1					
Hybrid	Stability Index		Hybrid	Env. 1	Env. 1	...	Env. 6	Env. 6	Hybrid	Env. 1	Env. 1	...	Env. 6	Env. 6
1	Observed	Observed	1	Observed	Observed	...	Predicted	Observed	1	Predicted	Observed	...	Observed	Observed
2	Observed	Observed	2	Observed	Observed	...	Predicted	Observed	2	Predicted	Observed	...	Observed	Observed
3	Predicted	Observed	3	Predicted	Observed	...	Observed	Observed	3	Predicted	Observed	...	Observed	Observed
4	Predicted	Observed	4	Predicted	Observed	...	Predicted	Observed	4	Predicted	Observed	...	Observed	Observed
5	Observed	Observed	5	Observed	Observed	...	Observed	Observed	5	Predicted	Observed	...	Observed	Observed
6	Predicted	Observed	6	Observed	Observed	...	Observed	Observed	6	Predicted	Observed	...	Observed	Observed
7	Observed	Observed	7	Predicted	Observed	...	Predicted	Observed	7	Predicted	Observed	...	Observed	Observed
8	Predicted	Observed	8	Observed	Observed	...	Predicted	Observed	8	Predicted	Observed	...	Observed	Observed
9	Observed	Observed	9	Predicted	Observed	...	Observed	Observed	9	Predicted	Observed	...	Observed	Observed
10	Predicted	Observed	10	Predicted	Observed	...	Predicted	Observed	10	Predicted	Observed	...	Observed	Observed
11	Predicted	Observed	11	Observed	Observed	...	Observed	Observed	11	Predicted	Observed	...	Observed	Observed
...	Observed	Observed	...	...	...	...	...	...	...	...	...	...	...	...
n	Observed	Observed	n	Predicted	Observed	...	Observed	Observed	n	Predicted	Observed	...	Observed	Observed

Correlation
Correlation
Correlation

**Figure 1.** Cross-validation 1 (CV1) and 2 (CV2) schemes.

The predictive accuracy (ACC) was estimated using Gezan et al. (2017) estimators:

$$ACC = \frac{\text{Corr}(y_{\text{obs}}, \hat{a})}{\sqrt{h^2}}$$

where:  $\text{Corr}(y_{\text{obs}}, \hat{a})$  is the Pearson correlation of predicted and observed values;  $h^2$  is the heritability of the trait under study, being  $h^2 = 1 - \frac{\sigma_e^2}{\sigma_y^2}$ , in which  $\sigma_e^2$  e  $\sigma_y^2$  correspond to residual and phenotypic components of variance, respectively.

For the genomic predictions, it was considered 30,000 iterations and burn-in of the first 20,000 iterations. The prediction analyses were performed using an Intel Core i5 2.20GHz processor and 8.00 GB of RAM.

## Results

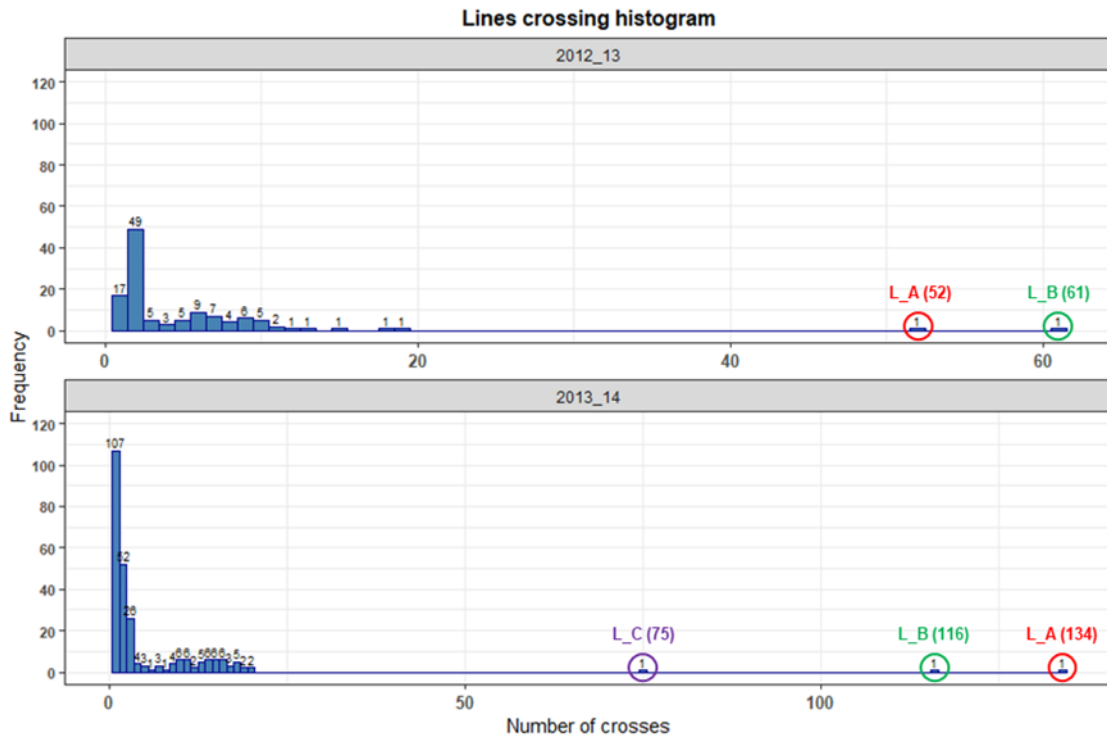
A preliminary analysis was performed to better understand the dataset. Despite the great number of hybrids, only 35 are common between the crop years (Figure 2). Due to the lack of hybrids overlapping over the 12 environments, the data was split up into two datasets: 2012/13 and 2013/14 collection.

**Coincident hybrids across environments**

E12	43	35	43	43	41	43	709	709	709	709	710	710
E11	43	35	43	43	41	43	709	709	709	709	710	710
E10	43	35	43	43	41	43	709	709	708	709	709	709
E9	43	35	43	43	41	43	708	708	709	708	709	709
E8	43	35	43	43	41	43	709	709	708	709	709	709
E7	43	35	43	43	41	43	709	709	708	709	709	709
E6	306	226	307	292	290	307	43	43	43	43	43	43
E5	291	223	292	280	292	290	41	41	41	41	41	41
E4	293	215	294	294	280	292	43	43	43	43	43	43
E3	308	227	309	294	292	307	43	43	43	43	43	43
E2	226	227	227	215	223	226	35	35	35	35	35	35
E1	308	226	308	293	291	306	43	43	43	43	43	43
	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12

**Figure 2.** Number of coincident maize hybrids across the 2012/2013 and 2013/2014 environments.

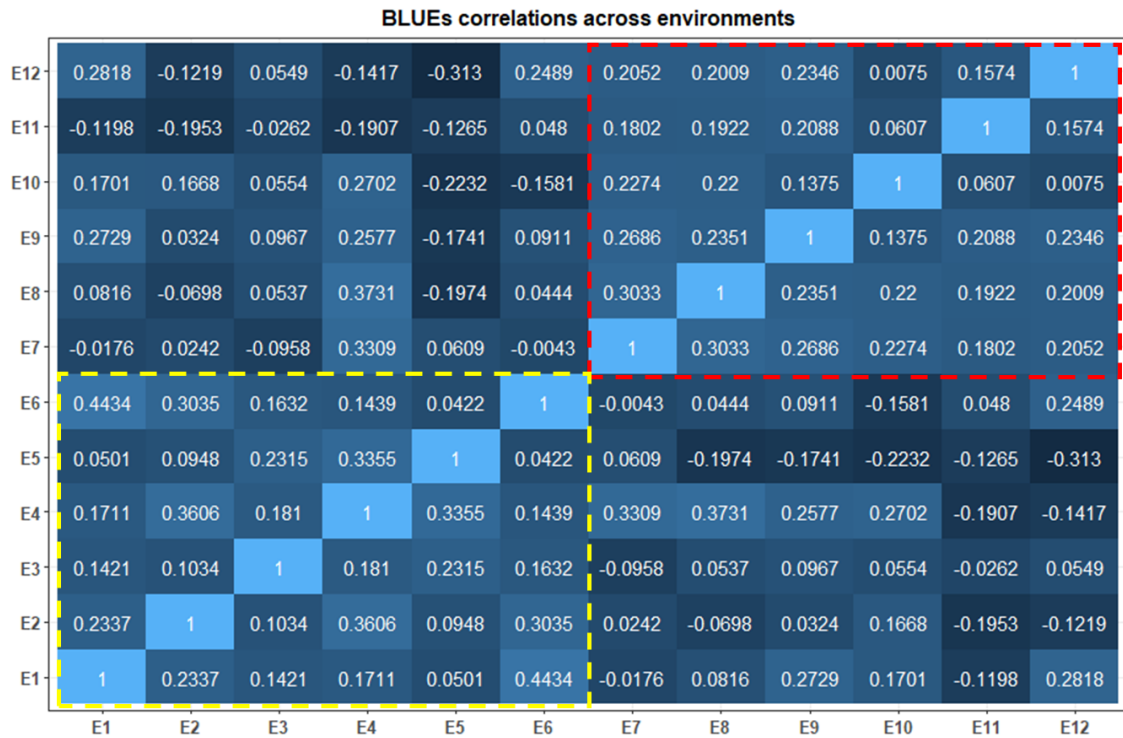
Although up to 309 (2012/2013) and 710 (2013/2014) have been synthesized, only 3 lines have been involved in most of the crosses. For the 2012/2013 dataset, Line A (L\_A) and Line B (L\_B) were involved in 52 and 61 crosses, respectively. The same lines were used to synthesized 134 (L\_A) and 116 (L\_B) hybrids in 2013/2014 as well. Furthermore, Line C (L\_C) was a parental for 75 hybrids considering the 2013/2014 dataset (Figure 3).



**Figure 3.** Histogram of the number of crosses that each maize line has been involved in, considering both 2012/2013 and 2013/2014 datasets.

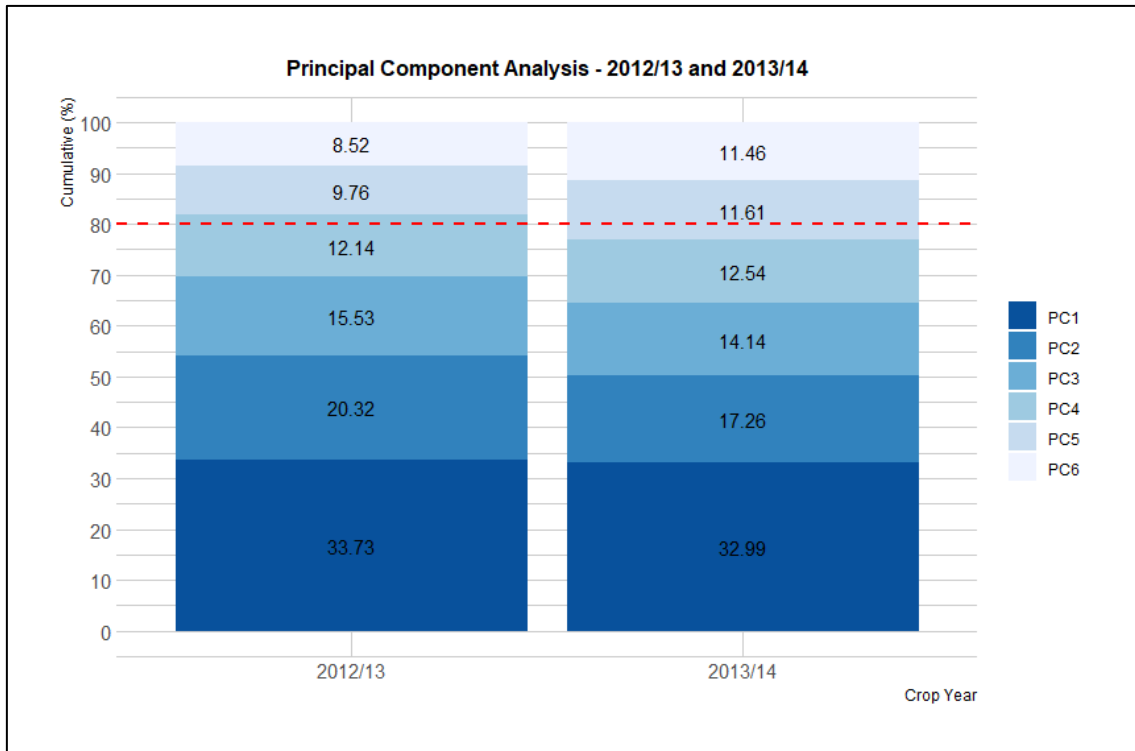
The BLUEs correlations across environments are presented in Figure 4. As expected, the low magnitude of the correlations indicates that hybrids' behavior varied across environments. Considering the 2012/2013 dataset (in yellow), the correlations ranged from 0.0422 (E5-E6) to 0.4434 (E1-E6), being 0.2 the average considering the six 2012/2013 environments.

On the other hand, for the 2013/2014 dataset (in red) the lower correlation was observed between E10 and E12 (0.0075) and the higher between E7-E8 (0.3033), the average correlation for this dataset was 0.189. The inconsistencies of the correlations across both datasets (0.023 on average) corroborate the decision of working with their data separately.



**Figure 4.** Maize hybrids BLUEs correlation across environments. The 2012/2013 dataset is highlighted in Yellow, while the 2013/2014 dataset is in red.

Figure 5 presents the summary of the Principal Component Analysis (PCA). The first four principal components explained a large percentage of the total variation present in both datasets. PC1 to PC4 explained about 82% of the 2012/2013 variation, while for 2013/2014 the first four PC explained 77% of the variation.

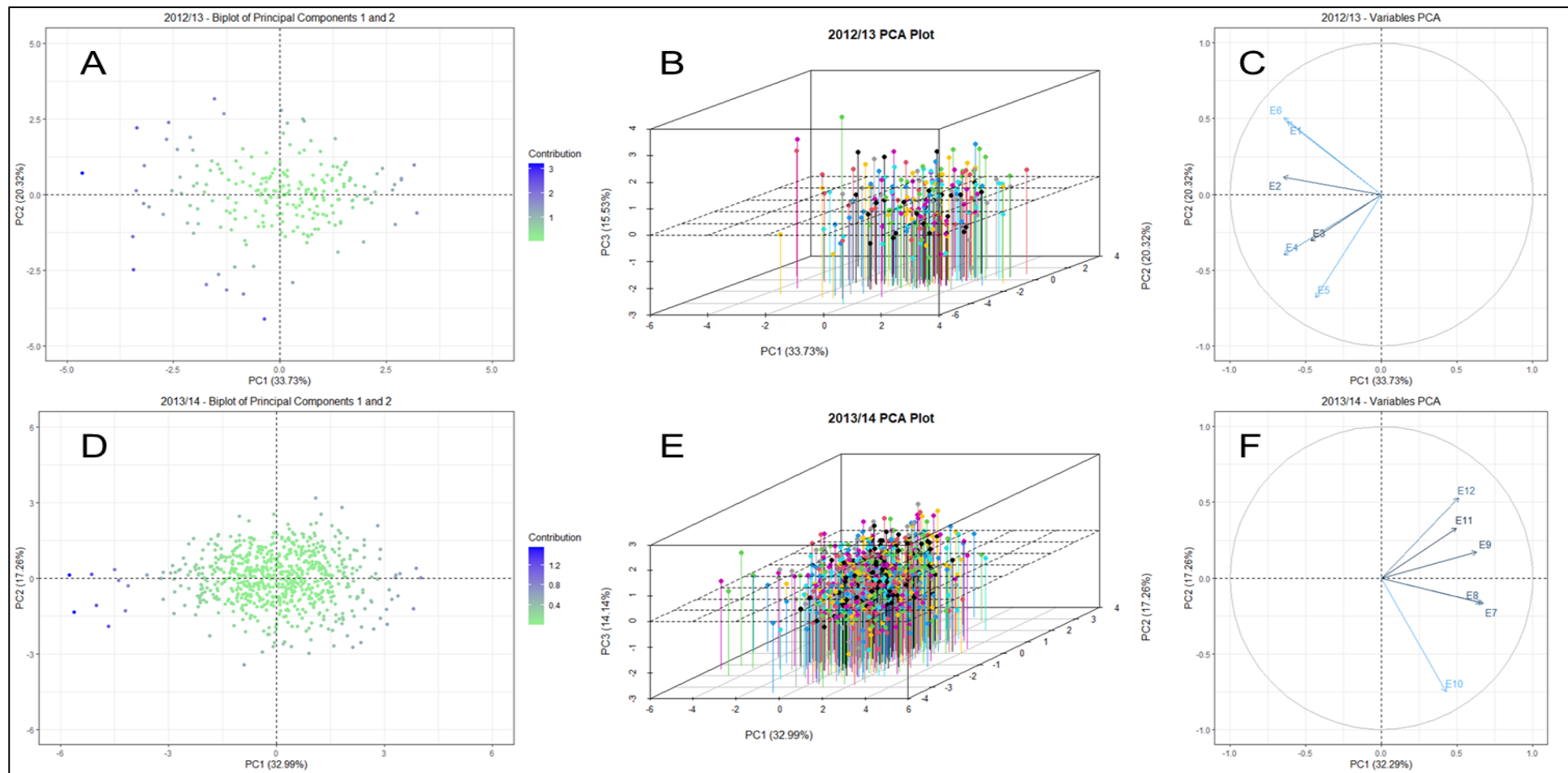


**Figure 5.** Summary of Principal Component Analysis (PCA) of the 2012/2013 and 2013/2014 datasets.

To better visualize and understand the results, the maize hybrids scores considering the principal components 1, 2, and 3 were plotted in Figure 6. Considering the biplots since the hybrids in green are near the center of the biplots, their contributions to GE interaction are lower than the hybrids represented by the blue dots. Genotypes near the center are more stable.

For both datasets, most of the hybrids are concentrated near the biplot center (Figures 6A and 6D), in other words, only a few (not irrelevant) hybrids are very unpredictable across the environments.

The 3D plots aid to understand the maize hybrids considering not only PC1 and 2 but also PC3. As closer to the center of the central grid (X, Y, and Z axis equals 0, 0, 0 respectively), more stable is the hybrid.

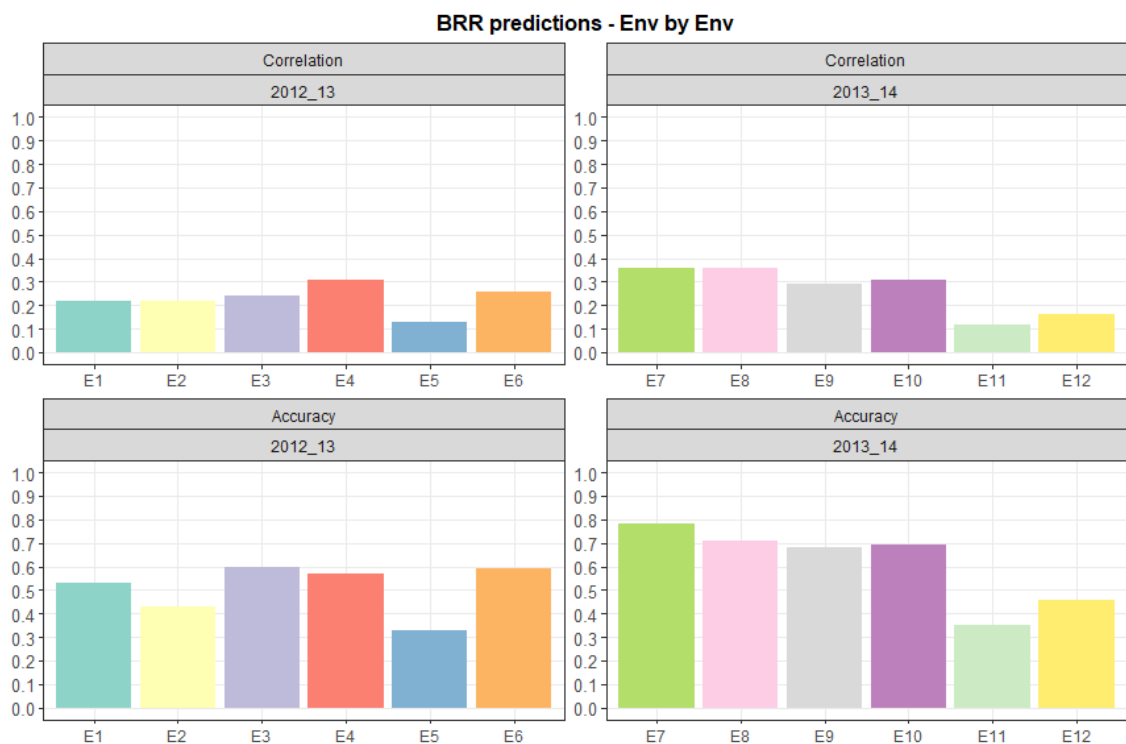


**Figure 6.** Principal Components Analysis biplots (A and D), 3D plots (B and E), and Environments biplots (C and F) of 2012/2013 and 2013/2014 datasets.



Figure 6 (C and F) presents the PCA results for the environments. Considering the 2012/2013 dataset, environments E1 and E6 have similar directions and magnitudes, a pattern also observed for E3 and E4. This plot indicates that the dataset can be reduced to 4 PCs without losing much information. For 2013/2014, the direction and magnitudes of E7 and E8 are quite similar as well.

The results of the single-Environment predictions are presented in Figure 7. The average Pearson's correlation (predictive ability) for the 2012/2013 dataset fluctuated from 0.13 (E5) to 0.31 (E4). The average considering all the six 2012/2013 environments was 0.23. The correlations for the 2013/2014 dataset were slightly higher, the lower correlations were observed in E11 (0.12) and E12 (0.16), while E7 and E8 presented higher correlations (0.36 for both). On average, the correlation for the 2013/2014 dataset was 0.27.



**Figure 7.** Average predictive abilities (correlation) and accuracies for 2012/2013 and 2013/2014 single-Environment genomic predictions. The estimates were obtained after 10 rounds of 10-fold cross-validation.

In terms of accuracy, a similar pattern was observed (Figure 7). The lower accuracies were observed in E5 (0.33), E11 (0.35), and E12 (0.46). On average, the accuracies observed were 0.51 and 0.61 for 2012/2013 and 2013/2014, respectively.

Figure 8 presents the results of the multi-Environment predictions. In general, the results for 2013/2014 were more promising than those obtained for 2012/2013. Even though the predictions for ED (PC4) and MHPRVG were computationally efficient, only 1.5 min (2012/2013) and 6 min (2013/2014) to run a round of 10-fold CV prediction, the *M1* linear predictor model (G effect only) presented poor predictive ability and accuracy considering ED (PC4) for 2012/2013. On average, the correlations for ED (PC4) were -0.079 (nil) and 0.16 for 2012/2013 and 2013/2014, respectively, while the estimates of accuracy were -0.247 (2012/2013) and 0.48 (2013/2014).

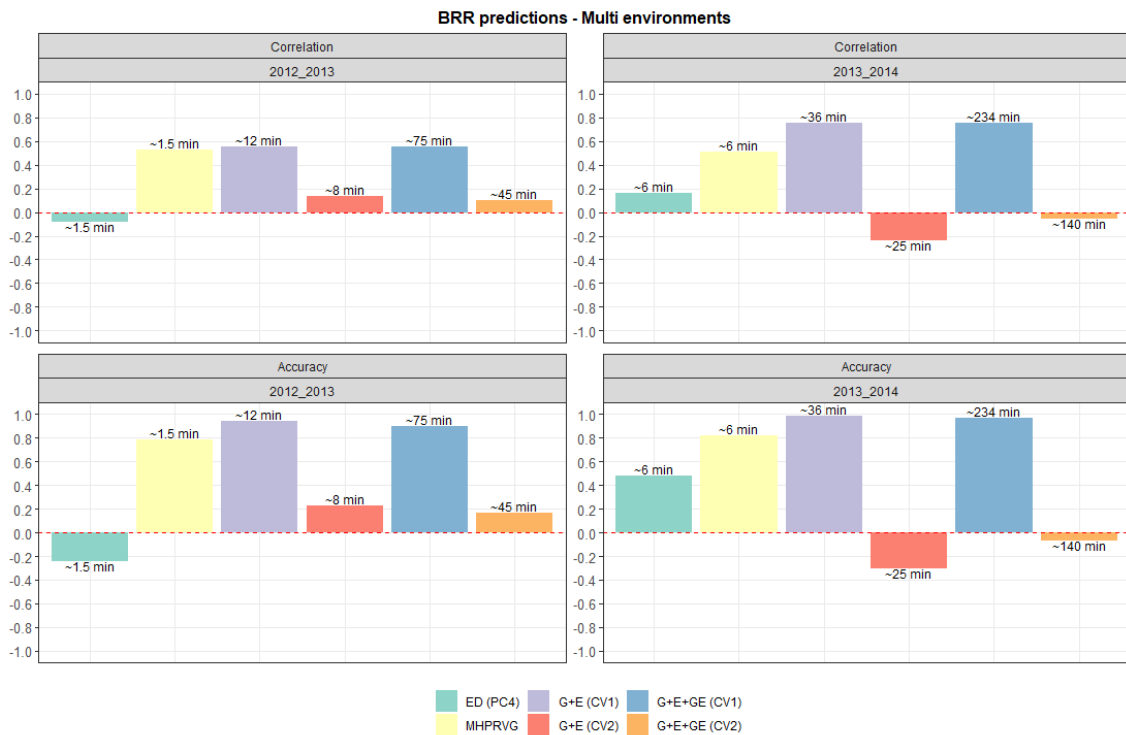
On the other hand, the correlations for MHPRVG were encouraging: 0.53 (2012/2013) and 0.51 (2013/2014) on average (Figure 8). The predictive accuracies were high: 0.78 and 0.82 for 2012/2013 and 2013/2014, respectively.

The results for *M2* (G+E) and *M3* (G+E+G\*E) were similar considering CV1 for both datasets. For 2012/2013, the estimates of correlation were 0.56, on average, for both *M2* and *M3*. While for 2013/2014 the correlations were equal to 0.75 and 0.76 considering *M2* and *M3*, respectively. In terms of accuracy, similar results were obtained: 0.94 (*M2* - 2012/2013), 0.90 (*M3* - 2012/2013), 0.98 (*M2* - 2013/2014), and 0.97 (*M3* - 2013/2014), considering the average of the 10 rounds of cross-validation (Figure 8).

A different scenario was observed for CV2, the correlations were drastically reduced, as well as the accuracies. For 2012/2013, the correlations were 0.14, for *M2*, and 0.10 for *M3*, while the accuracies were 0.23 and 0.16, for *M2* and *M3*, respectively. Considering 2013/2014,

these reductions were more severe, the correlations were equal to -0.23 and -0.05 (nil), and the accuracies were equal to -0.30 and -0.06 for *M2* and *M3*, respectively (Figure 8).

Considering the computational cost, the *M3* was time-consuming for both datasets, requiring, on average, 6x more time than *M2*, approximately (Figure 8).



**Figure 8.** Average predictive abilities and accuracies for 2012/2013 and 2013/2014 multi-Environment genomic predictions. The estimates were obtained after 10 rounds of 10-fold cross-validation. The time required to perform each round of cross-validation is presented near the bars.

## Discussion

The GxE interaction is one of the most important factors that affect plant breeding. This interaction plays a crucial role in the genotype selection process (Von Pinho; Silva; Oliveira, 2022). Tropical environments provide great challenges in terms of biotic and abiotic stresses. Considering Brazil's tropical condition, the safrinha is harshly affected, due to a smaller

window for maize planting, a higher risk of drought and insects and disease pressure in comparison to the summer season.

As presented in Figure 4, the low estimates of BLUEs correlations indicate the presence of GxE interaction, which can be split up into Genotypes x Locations (G x L) considering within the same crop year (2012/13 and 2013/14), and Genotypes x Crop (G x C) years considering only coincident hybrids between the 2012/13 and 2013/14 crop years. G x L can be easily observed in Figure 4, however, the poor genotypes overlapping across 2012/13 and 2013/14 do not allow a good and accurate inference about G x C, even though the great effects of the G x Years interaction are well-known (Pereira et al., 2022; Silva et al., 2022). This scenario corroborates the decision of working with both datasets separately.

In terms of the Principal Components Analysis (PCA), it was possible to reduce from 6 to 4 PCs without losing much information (Figure 5). Although reducing from 6 to 4 does not represent a great reduction, in larger datasets (larger number of environments), the PCA strategy might be more efficient. The Euclidian Distance has been used to measure genotypes' stabilities (Nzuve et al., 2013; Yamamoto et al., 2021; Yue et al., 2022a; 2022b). This index considers the genotype's contribution to the total variation presented in the dataset. The smaller its contribution, the higher the genotype's stability (Nzuve et al., 2013).

Figure 6 is very informative for both datasets. The blue dots, in Figures 6A and 6D, represent the "unpredictable" (low stability) hybrids. Stability is crucial for genotype selections and can be a decisive factor in terms of the advancement or not a given hybrid due to its associated risks (Santos et al., 2019; Von Pinho; Silva; Oliveira, 2022). In addition, Figures 6C and 6F represent the relationship between the environment, as previously mentioned E1 and E6 have the same directions and similar magnitudes (Figure 6C) which corroborate with the results presented in Figure 4 (E1-E6 correlation equals 0.44). The same pattern was observed for E7

and E8 which presented the highest correlation (0.30) across the 2013/2014 environments (Figure 4).

The single-Environment predictions (Figure 7) were low to medium magnitudes, which is okay considering the low heritability of yield trait and the complexity of the dataset. Good genotype representation associated with good phenotypic data is crucial to achieving good prediction results, in other words, a more representative population, in terms of the number of crosses per line, is better for predictions. Atanda et al. (2021) and Fristche-Neto et al. (2021) have shown the implications of the population structure in maize hybrids' genomic predictions. Considering cross-validation schemes, maximizing the relatedness between training and validation sets is essential. Briefly, well-design genotypes set (training and testing populations) is required to achieve good results (Windhousen et al., 2012; Riedelsheimer et al., 2013; Atanda et al., 2021; Fristche-Neto et al., 2021). Such a scenario was not achieved considering the presented dataset.

As presented in Figure 3, considering the 2013/2014 dataset, 107 lines were involved only once to synthesize the maize hybrids, therefore, most of the lines had poor representativeness. For the 2012/2013 dataset, 17 lines were used only once, while 49 lines were involved in two different crosses.

The results of the multi-Environment predictions (Figure 8) vary depending on the used approach. The Euclidian distance (ED) results were not satisfactory. For both datasets, the ED results were the lowest considering the CV1 for **M1** (ED and MHPRVG), **M2** (G+E), and **M3** (G+E+G\*E), highlighting the difficulty to work only with GxE interaction.

A different scenario was observed while predicting the MHPRVG. As this index also considers the performance of the hybrids per se, even though the GxE is presented, the best hybrids should present the best performance across the environments (Resende, 2007; Mendes

et al., 2012; Yamamoto et al., 2021). In other words, although the behavior of the hybrids varies across the environments, the best genotypes (higher MHPRVG) outperform the bad ones (lower MHPRVG). The correlation result (0.52 on average) for this approach was satisfactory, considering the complexity of the dataset and the computational costs. Therefore, the MHPRVG index outperformed both the ED index and the single-Environment approach.

Although its higher computational cost, the **M2** approach was also satisfactory (Figure 8). In comparison to MHPRVG, considering CV1, the **M2** correlation results were slightly better (+6%) for the 2012/2013 dataset, while for the 2013/2014, the **M2** approach improves the correlation results by 47%. Therefore, despite the computational cost, for both datasets, multi-environment predictions (without modeling the GxE) were consistently better than predicting a stability index (ED or MHPRVG). Additional studies, considering well-designed genotype sets (training and testing populations), are still required.

The results of **M3** indicate there was no advantage in including the GxE interaction in the prediction model (Figure 8). Still considering the CV1, the higher computational cost of **M3** (in comparison to **M2**) did not justify its use, since, in terms of predictive ability and accuracy, no gain was observed. On the other hand, the advantages of modeling the GxE interaction are widely reported in the literature, however, its positive impact highly depends on other factors such as significant GxE (Crossa et al., 2022), environmental and genetic similarities (Rogers and Holland, 2021).

By modeling the GxE, Montesinos-López et al. (2015) achieved 9-14% gains when predicting the resistance to gray leaf spot in maize lines, while Acosta-Pech et al. (2017) achieved 16.73% gains, on average, when predicting maize hybrid grain yield performance. Predicting kernel zinc concentration in tropical maize germplasm, Magento et al. (2020)

obtained better correlations when modeling the GxE either predicting newly developed lines (CV scheme 1) or predicting the performances in sparse multi-location trials (CV scheme 2).

In addition to well-designed genotypes set, to achieve satisfactory results, good genotypes-by-environments overlapping is required. The overlapped genotypes, which are the common genotypes across all the environments, work as a connection across the environments. Jarquín et al. (2020) found that 15% of genotypes overlapping was enough to reduce phenotyping costs without significant losses in prediction accuracy. The intensity of genotypes overlapping may change due to dataset characteristics, however, it is common sense that larger training population sets associated with good genetic relationships between training and testing populations are better for predictions (Sánchez and Akdemir, 2021).

The lower results were expected for CV2. However, in this study, **M2** and **M3** presented very poor results considering whole environments predictions (CV2). It is well-known that predicting whole environments is very complex and often results in lower prediction accuracies as observed in Jarquín et al. (2017), Roorkiwal et al. (2018), Persa et al. (2021), and Westhues et al. (2021). Although this approach could be very interesting, in terms of resource-saving, its effectiveness is questionable, due to the greater reductions in the predictive ability. In this sense partitioning the whole hybrids set across the environments (sparse phenotyping) is a better strategy to save resources without losing predictive ability (Jarquin et al., 2020; Sánchez and Akdemir, 2021).

In addition, environmental covariables (EC) could also be used to better connect the environments, enhancing the accuracies of whole environments predictions. The so-called envirotyping consists of the determination and measurement of all the environmental factors or ECs that can affect plant growth and yield (Xu, 2016). These factors are then processed aiming to describe the development of an organism in a target environment and can be used to create

envirotyping-based kinships (Costa Neto et al., 2021a), enhancing GxE modeling in a genomic prediction perspective. Gillberg et al. (2019) and Costa-Neto et al. (2021b) have shown the advantages of using EC. On the other hand, Jarquin et al. (2021) observed no gains with the inclusion of EC in the GxE model, therefore, additional studies are still required.

Unfortunately, the used dataset lacks on some essential environmental information such as precise planting date, soil type, coordinates, altitude, and others, which makes the dataset not appropriate for this purpose.

## Conclusion

The MHPRVG index outperformed Euclidian Distance in both datasets (2012/2013 and 2013/2014). Considering its performance and computational cost, directly predicting the MHPRVG might be efficient for genomic selection purposes.

The Multi-Environmental (*M2*) approach achieved higher correlations, outperforming the ACC of MHPRVG in 17% (2012/2013) and 16% (2013/2014) using feasible computational resources. No gain was observed by including the GxE interaction (*M3*), not justifying its higher computational cost (7fold).

Additional studies using better-designed datasets are still required.

## References

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. 10.1007/s00122-017-2898-0



- Amadeu RR, Cellon C, Olmstead JW, Garcia AAF, Resende MFR, Muñoz PR (2016) AGHmatrix: R package to construct relationship matrices for autotetraploid and diploid species: a blueberry example. *Plant Genome* 9: 1–10. [10.3835/plantgenome2016.01.0009](https://doi.org/10.3835/plantgenome2016.01.0009)
- Andrea MCdaS, Boote KJ, Sentelhas PC, Romanelli TL (2018) Variability and limitations of maize production in Brazil: Potential yield, waterlimited yield and yield gaps. *Agricultural Systems*, 165: 264-273. <https://doi.org/10.1016/j.agsy.2018.07.004>.
- Atanda SA, Olsen M, Burgueño J, Crossa J, Dzidzienyo D, Beyene Y, Gowda M, Dreher K, Zhang X, Prasanna BM, Tongoona P, Danquah EY, Olaoye G, Robbins KR (2021) Maximizing efficiency of genomic selection in CIMMYT's tropical maize breeding program. *Theoretical and Applied Genetics*, 134: 279-294. <https://doi.org/10.1007/s00122-020-03696-9>
- Beyene U, Gowda M, Olsen M, Robbins KR, Pérez-Rodríguez P, Alvarado G, Dreher K, Gao SY, Mugo S, Prasanna BM, Crossa J (2019) Empirical comparison of tropical maize hybrids selected through genomic and phenotypic selections. *Frontiers in Plant Science*, 10(1502): 1–11. <https://doi.org/10.3389/fpls.2019.01502>
- Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic prediction of breeding values when modeling genotype x environment interaction using pedigree and dense molecular markers. *Crop Science*, 55: 707-719. [10.2135/cropsci2011.06.0299](https://doi.org/10.2135/cropsci2011.06.0299)
- Cantelmo NF, Von Pinho RG, Balestre M (2017) Genomic analysis of maize lines introduced in the early stages of a breeding programme. *Plant Breeding*, 136(6): 845–860. [10.1111/pbr.12543](https://doi.org/10.1111/pbr.12543)
- Costa-Neto G, Galli G, Carvalho HF, Crossa J, Fritsche-Neto R (2021a) EnvRtype: a software to interplay enviromics and quantitative genomics in agriculture. *G3 Genes | Genomes | Genetics*, 11(4): 1-20. [10.1093/g3journal/jkab040](https://doi.org/10.1093/g3journal/jkab040)

- Costa-Neto G, Crossa J, Fritsche-Neto R (2021b) Enviromic assembly increases accuracy and reduces costs of the genomic prediction for yield plasticity in maize. *Frontiers in Plant Science*, 12(717552): 1-21. [10.3389/fpls.2021.717552](https://doi.org/10.3389/fpls.2021.717552)
- Crossa J, Montesinos-López OA, Péres-Rodríguez P, Costa-Neto G, Frische-Neto R, Ortiz R, Martini JWR, Lillemo M, Montesinos-López A, Jarquin D, Breseghello F, Cuevas J, Rincent R (2022) Chapter 9: Genome and Environment Based Prediction Models and Methods of Complex Traits Incorporating Genotype x Environment Interaction. Nourollah Ahmadi and Je'ro^me Bartholome' (eds.), *Genomic Prediction of Complex Traits: Methods and Protocols*, *Methods in Molecular Biology*, vol. 2467, [https://doi.org/10.1007/978-1-0716-2205-6\\_9](https://doi.org/10.1007/978-1-0716-2205-6_9)
- Dias KOG, Gezan SA, Guimarães CT, Parentoni SN, Guimarães PEDO, Carneiro NP, Portugal EAB, Cardoso MJ, Anoni CO, Magalhães JV, Souza JC, Guimarães LJM, Pastina MM (2018) Estimating genotype  $\times$  environment interaction for and genetic correlations among drought tolerance traits in maize via factor analytic multiplicative mixed models. *Crop Science*, 58(1): 72-83. [10.2135/cropsci2016.07.0566](https://doi.org/10.2135/cropsci2016.07.0566)
- Eeuwijk FAvan, Bustos-Korts DV, Malosetti M (2016) What should students in plant breeding know about the statistical aspects of genotypes x environment interactions? *Crop Science*, 56(5): 2119-2140. <https://doi.org/10.2135/cropsci2015.06.0375>.
- Fristche-Neto R, Galli G, Borges KLM, Costa-Neto G, Alves FC, Sabadin F, Lyra DH, Morais PPP, Andrade LRB, Granato I, Crossa J (2021) Optimizing Genomic-Enabled Prediction in Small-Scale Maize Hybrids Breeding Programs: A Roadmap Review. *Frontiers in Plant Science*, 12(658267): 1-16. [10.3389/fpls.2021.658267](https://doi.org/10.3389/fpls.2021.658267)
- Gezan SA, Osorio LF, Verma S, Whitaker VM (2017) An experimental validation of genomic selection in octoploid strawberry. *Horticulture Research*, 4(16070): 1-9. <https://doi.org/10.1038/hortres.2016.70>

- Gillber J, Marttinen P, Mamitsuka H, Kaski S (2019) Modeling G x E with historical weather information improves genomic prediction in new environments. *Bioinformatics*, 35(20): 4045-4052. [10.1093/bioinformatics/btz197](https://doi.org/10.1093/bioinformatics/btz197)
- Heslot N, Akdemir D, Sorrells ME, Jannink JL (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. <https://doi.org/10.1007/s00122-013-2231-5>.
- Jarquín D, Silva CLda, Gaynor RC, Poland J, Howard AFR, Battenfield S, Crossa J (2017) Increasing Genomic-Enabled Prediction Accuracy by Modeling Genotype x Environment Interactions in Kansas Wheats. *The Plant Genome*, 10(2): 1-15. [10.3835/plantgenome2016.12.0130](https://doi.org/10.3835/plantgenome2016.12.0130)
- Jarquín D, Howard R, Crossa J, Beyene Y, Gowda M, Martini JWR, Pazaran GC, Burgueño J, Pacheco A, Grondona M, Wimmer V, Prasanna BM (2020) Genomic Predictions Enhanced Sparse Testing for Multi-environment Trials. *G3 Genes | Genomes | Genetics*, 10: 2735-2739. <https://doi.org/10.1534/g3.120.401349>
- Jarquín D, de Leon N, Romay C, Bohn M, Buckler ES, Ciampittis I, Edwards J, Ertle D, Flint-Garcia S, Gore MA, Graham C, Hirsch CN, Holland JB, Hooker D, Kaeppler SM, Knoll J, Lee EC, Lawrence-Dill CJ, Lynch JP, Moose SP, Murray SC, Nelson R, Rocheford T, Schnable JC, Schnable PS, Smith M, Springer N, Thomison P, Tuinstra M, Wisser RJ, Xu W, Yu J, Lorenzi A (2021) Utility of climatic information via combining ability models to improve genomic prediction for yield within the genomes to fields maize project. *Frontiers in Genetics*, 11: 1-11. [10.3389/fgene.2020.592769](https://doi.org/10.3389/fgene.2020.592769)
- Kassambara A, Mundt F (2020) *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. R package version 1.0.7.

- Mageto EK, Crossa J, Pérez-Rodríguez P, Dhliwayo T, Palacios-Rojas N, Lee M, Guo R, Vicente FS, Zhang X, Hindu V (2020) Genomic Predictions with Genotype by Environment Interaction Analysis for Kernel Zinc Concentration in Tropical Maize Germplasm. *G3 Genes | Genomes | Genetics*, 10: 2629-2639. <https://doi.org/10.1534/g3.120.401172>
- Mendes FF, Guimarães LJM, Souza JC, Guimarães PEO, Pacheco CAP, Machado JRdeA, Meirelles WF, Silva AR, Parentoni SN (2012) Adaptability and stability of maize varieties using mixed model methodology. *Crop Breeding and Applied Biotechnology*, 12: 111-117. <https://doi.org/10.1590/S1984-70332012000200003>
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genomewide dense marker maps. *Genetics*, 157(4): 1819-1829. <https://doi.org/10.1093/genetics/157.4.1819>.
- Montesinos-López OA, Montesinos-López A, Pérez-Rodríguez P, de los Campos G, Eskridge K, Crossa J (2015) Threshold Models for Genomed-Enebled Prediction of Ordinal Categorical Traits in Plant Breeding. *G3 Genes | Genomes | Genetics*, 5: 291-300. [10.1534/g3.114.016188](https://doi.org/10.1534/g3.114.016188)
- Montesinos-López AO, Montesinos-López A, Pérez-Rodríguez P, Barrón-López JÁ, Martini JWR, Fajardo-Flores SB, Gaytan-Lugo LS, Santana-Mancilla PC, Crossa J (2021) A review of deep learning applications for genomic selection. *BMC Genomics*, 22(19): 1-23. <https://doi.org/10.1186/s12864-020-07319-x>
- Montesinos-López OA, Montesinos-López JC, Montesinos-López A, Ramírez-Alcaraz JM, Polland J, Singh R, Dreisigacker S, Crespo L, Mondal S, Govida V, Juliana P, Espino JH, Shrestha S, Varshney RK, Crossa J (2022) Bayesian multitrait kernel methods improve multienvironment genome-based prediction. *G3 Genes | Genomes | Genetics*, 12(2): 1–17. <https://doi.org/10.1093/g3journal/jkab406>

- Nzuve F, Githiri S, Mukunya DM, Gethi J (2013) Analysis of Genotype x Environment Interaction for Grain Yield in Maize hybrids. *Journal of Agricultural Science*, 5(11): 75-85. [10.5539/jas.v5n11p75](https://doi.org/10.5539/jas.v5n11p75)
- Pereira FdeC, Ramalho MAP, Resende Junior MFR, Von Pinho RG (2022) Mega-environment analysis of maize breeding data from Brazil. *Scientia Agricola*, 79(2): 1-10. <http://doi.org/10.1590/1678-992X-2020-0314>
- Pérez P, de los Campos G (2014) Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics*, 198(2): 483-1495. [10.1534/genetics.114.164442](https://doi.org/10.1534/genetics.114.164442)
- Persa R, Grondona M, Jarquín D (2021) Development of a genomic prediction pipeline for maintaining comparable sample sizes in training and testing sets across prediction schemes accounting for the genotype-by-environment interaction. *Agriculture*, 11(932): 1-17. <https://doi.org/10.3390/agriculture11100932>
- R Development Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Resende MDV (2002) *Genética biométrica e estatística no melhoramento de plantas perenes*. Brasília: Embrapa Informação Tecnológica; Colombo: Embrapa Florestas. 975p.
- Resende MDV (2004) *Métodos Estatísticos Ótimos na Análise de Experimentos de Campo*. Embrapa Florestas. Colombo. 65 p. (Documentos 100).
- Resende MDV (2007) *Matemática e estatística na análise de experimentos e no melhoramento genético*. Embrapa Florestas, Colombo, 362p
- Rezende WS, Cruz CD, Borém A, Rosado RDS (2021) Half a century of studying adaptability and stability in maize and soybean in Brazil. *Scientia Agricola*, 78(3): 1-9. <https://doi.org/10.1590/1678-992x-2019-0197>

- Riedelsheimer C, Endelman JB, Stange M, Sorrells ME, Jannink J-L, Melchinger AE (2013) Genomic Predictability of Interconnected Biparental Maize Populations. *Genetics*, 194: 493-503. [10.1534/genetics.113.150227](https://doi.org/10.1534/genetics.113.150227)
- Rogers AR and Holland JB (2021) Environment-specific genomic prediction ability in maize using environmental covariates depends on environmental similarity to training data. *G3 Genes | Genomes | Genetics*, 12(2): 1-14. <https://doi.org/10.1093/g3journal/jkab440>
- Roorkiwal M, Jarquín D, Sing MK, Gaur PM, Bharadwaj C, Rathore A, Howard R, Srinivasan S, Jain A, Garg V, Kale S, Chitikineni A, Tripathi S, Jones E, Robbins KR, Crossa J, Varshney RK (2018) Genomic-enabled prediction models using multi-environment trials to estimate the effect of genotype x environment interaction on prediction accuracy in chickpea. *Scientific Reports*, 8(11701): 1-11. [10.1038/s41598-018-30027-2](https://doi.org/10.1038/s41598-018-30027-2)
- Sánchez JI y, Akdemir D (2021) Training set optimization for sparse phenotyping in genomic selection: a conceptual overview. *Frontiers in Plant Science*, 12: 1-14. [10.3389/fpls.2021.715910](https://doi.org/10.3389/fpls.2021.715910)
- Santos DCdos, Pereira CH, Nunes, JAR, Lepre AL (2019) Adaptability and stability of maize hybrids in unreplicated multienvironment trials. *Revista Ciência Agronômica*, 50(1): 83-89. [10.5935/1806-6690.20190010](https://doi.org/10.5935/1806-6690.20190010)
- Silva EVV, Bruzi AT, Silva FDS, Marques FS (2022) Genotypic and phenotypic parameters associated with early maturity in soybean. *Pesquisa Agropecuária Brasileira*, 57(e02545): 1-10. <https://doi.org/10.1590/S1678-3921.pab2022.v57.02545>
- VanRaden PM (2008) Efficient methods to compute genomic predictions. *Journal of dairy science*, 91(11): 4414-4423. [10.3168/jds.2007-0980](https://doi.org/10.3168/jds.2007-0980)
- Von Pinho RG, Silva EVV, Oliveira TL (2022) Challenges of maize breeding under tropical conditions of Brazil. *Revista Brasileira de Milho e Sorgo*, 21(e1258): 1-25. <https://doi.org/10.18512/rbms2022vol21e1258>

- Wang N, Wang H, Zhang A, Liu Y, Yu D, Hao Z, Ilut D, Glaubitz JC, Gao Y, Jones E, Olsen M, Li X, Vicente FS, Prasanna BM, Crossa J, Pérez-Rodríguez P, Zhang X (2020) Genomic prediction across years in a maize double haploid breeding program to accelerate early-stage testcross testing. *Theoretical and Applied Genetics*, 133: 2869-2879. <https://doi.org/10.1007/s00122-020-03638-5>
- Westhues CC, Mahine GS, Silva Sda, Thorwarth P, Schmidt M, Richter J-C, Simianer H, Beissinger TM (2021) Prediction of maize phenotypic traits with genomic and environmental predictors using gradient boosting frameworks. *Frontiers in Plant Science*, 12(699589): 1-22. 10.3389/fpls.2021.699589
- Windhause VS, Atlin GN, Hickey JM, Crossa J, Jannik J-L, Sorrels ME, Raman B, Cairns JE, Tarekegne A, Semagn K, Beyene Y, Grudloyma P, Technow F, Riedelsheimer C, Melchinger AE (2012) Effectiveness of Genomic Prediction of Maize Hybrid Performance in Different Breeding Populations and Environments. *G3 Genes | Genomes | Genetics*, 2: 1427-1436. 10.1534/g3.112.003699
- Xu Y (2016) Envirotyping for deciphering environmental impacts on crop plants. *Theoretical and Applied Genetics*, 129: 653-673. 10.1007/s00122-016-2691-5
- Yamamoto ELM, Gonçalves MC, Davide LMC, Santos Ados, Candido LS (2021) Adaptability and stability of maize genotypes in growing regions of central Brazil. *Revista Ceres*, 68(3): 201-211. 10.1590/0034-737X202168030006
- Yue H, Gauch HG, Wei J, Xie J, Chen S, Peng H, Bu J, Jiang X (2022a) Genotype by Environment Interaction Analysis for Grain Yield and Yield Components of Summer Maize Hybrids across the Huanghuaihai Region in China. *Agriculture*, 12(602): 1-17. <https://doi.org/10.3390/agriculture12050602>
- Yue H, Olivoto T, Bu J, Li J, Wei J, Xie J, Chen S, Peng H, Nardino M, Jiang X (2022b) Multi-trait selection for mean performance and stability of maize hybrids in mega-environments

delineated using envirotyping techniques. *Frontier in Plant Science*, 3(1030521): 1-17.

10.3389/fpls.2022.1030521