UNIVERSIDADE FEDERAL DE VIÇOSA

Leveraging Relationship Information for Assessing Genotype Stability and Performance in Multi-Environment trials

Túlio Botelho Carvalho
Magister Scientiae

VIÇOSA - MINAS GERAIS
2025

TÚLIO BOTELHO CARVALHO

Leveraging Relationship Information for Assessing Genotype Stability and Performance in Multi-Environment trials

Dissertation submitted to the Genetics and Breeding Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

Adviser: Kaio O. das Gracas Dias

Co-adviser: Rodrigo Oliveira de Lima

VIÇOSA - MINAS GERAIS
2025

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T

C331L
2025

Carvalho, Túlio Botelho, 1998-

Leveraging relationship information for assessing genotype stability and performance in multi-environment trials / Túlio Botelho Carvalho. – Viçosa, MG, 2025.

1 dissertação eletrônica (48 f.): il. (algumas color.).

Texto em inglês.

Orientador: Kaio Olimpio das Graças Dias.

Dissertação (mestrado) - Universidade Federal de Viçosa, Departamento de Biologia Geral, 2025.

Referências bibliográficas: f. 40-48.

DOI: https://doi.org/10.47328/ufvbbt.2025.292

Modo de acesso: World Wide Web.

1. Interação genótipo-ambiente. 2. Milho - Melhoramento genético. 3. Modelos multiníveis (Estatísticas). 4. BLUP. I. Dias, Kaio Olimpio das Graças, 1987-. II. Universidade Federal de Viçosa. Departamento de Biologia Geral. Programa de Pós-Graduação em Genética e Melhoramento. III. Título.

CDD 22. ed. 633.152

Bibliotecário(a) responsável: Bruna Silva CRB-6/2552

TÚLIO BOTELHO CARVALHO

Leveraging Relationship Information for Assessing Genotype Stability and Performance in Multi-Environment trials

Dissertation submitted to the Genetics and Breeding Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

APPROVED: March 31, 2025.

Assent:

Túlio Botelho Carvalho
Author

Kaio Olimpio das Gracas Dias
Adviser

ACKNOWLEDGMENTS

This work has been sponsored by the following Brazilian research agencies: Coordination for the Improvement of Higher Education Personnel (CAPES; Financing code 001), Minas Gerais State Foundation for Research Aid (FAPEMIG) and National Council of Scientific and Technological Development (CNPq).

To God, for guiding and illuminating my paths.

To my parents, Sueli Fraiz Botelho de Carvalho, Adelson Mariano de Carvalho, and my sister, Sarah Botelho Carvalho, for their unconditional love, constant support, and encouragement in all moments of my life. I extend my gratitude to all my family members for their strength, affection, and support.

To the Federal University of Viçosa for my academic formation and significant contributions to my personal and professional growth.

To my advisors, professors Kaio Olímpio and Rodrigo Oliveira de Lima, for their guidance, patience, dedication, and valuable teachings.

To professor Saulo Fabrício, for his invaluable assistance with the data analysis of this work.

To Wemerson Rezende and Luiz Luz, for their important contribution to the evaluation of the experiments.

To the entire Programa Milho- UFV team, who over five years became my family in Viçosa, for the friendships created, exchanges of experiences, and fundamental support throughout the project’s development, significantly contributing to my personal and professional growth.

To the Quantitative Genetics and Statistical Learning LAB team, for their part nership, support, and essential collaboration in developing this work.

To the staff of the Experimental Field in Coimbra, Horta Nova, and Vale da Agronomia, for their dedication, competence, and indispensable technical support during the conduction of the experiments.

To all the friends who accompanied me on this journey in Vi¸cosa, for their friendship, companionship, and constant encouragement.

None of this would have been possible without you. Thank you very much!

ABSTRACT

CARVALHO, Túlio Botelho, M.Sc., Universidade Federal de Viçosa, March, 2025. **Leveraging Relationship Information for Assessing Genotype Stability and Performance in Multi-Environment trials**. Adviser: Kaio Olimpio das Gracas Dias. Co-adviser: Rodrigo Oliveira de Lima.

Genotype-by-environment interaction (G×E) is a central challenge in plant breeding programs, particularly in countries with high agroclimatic diversity. Traditional methods, such as analysis of variance and the AMMI model, are widely used in multi-environment trials (METs), but they exhibit limitations when dealing with unbalanced data and the modeling of covariances. In this context, mixed models provide a more robust alternative by integrating fixed and random effects, as well as enabling the use of molecular or pedigree-based genetic information, thereby improving prediction accuracy. This thesis proposes an extension of the WAASB method (Weighted Average of Absolute Scores of BLUPs), a stability index based on mixed models originally developed by Olivoto et al. (2019). The proposed extension incorporates a pedigree-based relationship matrix into the model. This addition addresses two key limitations of the original WAASB: the requirement for a complete G×E effect matrix to perform singular value decomposition (SVD), and the inability to estimate the stability of parental inbred lines that were not directly evaluated in METs. Phenotypic data from 160 three-way maize hybrids and seven commercial checks were used, evaluated across eight distinct environments. A pedigree-based relationship matrix was incorporated into a compound symmetry model to estimate G×E effects, from which WAASB values were calculated. A multi-trait selection index was applied to jointly assess performance and stability for grain yield, plant height, and ear height. Additionally, four simulation scenarios were constructed with varying levels of heritability and G×E interaction intensity. These were used to compare the performance of pedigree-based covariance imputation with the commonly used EM-SVD method. The results showed that the proposed approach reliably estimated parental line stability based on progeny performance and consistently outperformed the EM-SVD method. The methodology proved to be efficient and promising for selecting superior hybrids and stable parental lines in maize breeding programs.

Keywords: Genotype-by-environment interaction; Mixed models; WAASB; Maize breeding

RESUMO

CARVALHO, Túlio Botelho, M.Sc., Universidade Federal de Viçosa, março de 2025. **Incorporação de informações de parentesco para avaliar a estabilidade e desempenho de genótipos em ensaios de múltiplos ambientes**. Orientador: Kaio Olimpio das Gracas Dias. Coorientador: Rodrigo Oliveira de Lima.

A interação genótipos por ambientes (G×A) é um desafio central em programas de melhoramento genético, sobretudo em países com grande diversidade agroclimática. Métodos tradicionais, como a análise de variância e o modelo AMMI, são amplamente utilizados em ensaios multiambientes (METs), mas apresentam limitações frente a dados desbalanceados e à modelagem das covariâncias. Nesse contexto, modelos mistos oferecem uma alternativa mais robusta, ao integrar efeitos fixos e aleatórios, além de possibilitar o uso de informações genéticas moleculares ou de pedigree, melhorando as predições. Esta dissertação propõe uma extensão do método WAASB (Média Ponderada dos Escores Absolutos dos BLUPs), um índice de estabilidade baseado em modelos mistos originalmente desenvolvido por Olivoto et al. (2019). A extensão proposta incorpora ao modelo uma matriz de relacionamento baseada em pedigree. Essa adição busca superar duas limitações centrais do WAASB original: a necessidade de uma matriz completa de efeitos G×A para a realização da decomposição em valores singulares (SVD) e a impossibilidade de estimar a estabilidade de linhagens parentais que não foram avaliadas diretamente em METs. Foram utilizados dados fenotípicos de 160 híbridos triplos de milho e sete testemunhas comerciais, avaliados em oito ambientes distintos. Uma matriz de relacionamento baseada em pedigree foi incorporada a um modelo de simetria composta para estimar os efeitos G×A, a partir dos quais os valores de WAASB foram calculados. Um índice de seleção multicaracter foi aplicado para avaliar conjuntamente o desempenho e a estabilidade das características produtividade de grãos, altura de planta e altura de espiga. Adicionalmente, quatro cenários de simulação foram construídos com diferentes níveis de herdabilidade e intensidade de interação G×A. Esses cenários foram utilizados para comparar o desempenho da imputação baseada em covariância genética com o método EM-SVD, amplamente utilizado. Os resultados indicaram que a abordagem proposta estimou com precisão a estabilidade de linhagens parentais com base nas progênies e superou o método EM-SVD em todos os cenários. A metodologia mostrou-se eficiente e promissora para selecionar híbridos superiores e linhagens estáveis em programas de melhoramento de milho

Palavras-chave: Interação genótipos por ambientes; Modelos mistos; WAASB; Melhoramento de milho

List of Figures

List of Tables

Table of Contents

1 Introduction

One of the main challenges in plant breeding, particularly in a country with vast agroclimatic diversity like Brazil, is managing the genotype-by-environment interaction (G×E). This phenomenon, characterized by the differential performance of genotypes across environmental conditions, limits the predictability of genotype responses and complicates selection decisions (CRUZ; REGAZZI; CARNEIRO, 2012). G×E can be classified as either non-crossover, when genotype rankings remain consistent across environments, or crossover, when changes in rank occur, which is particularly detrimental for selecting stable cultivars (MONTEVERDE et al., 2018). In tropical environments where factors such as high solar radiation, elevated temperatures, water stress, and nutrient limitations are prevalent, complex G×E is especially prominent and requires more refined analytical approaches to identify resilient, high-yielding genotypes (CRUZ; REGAZZI; CARNEIRO, 2012).

To overcome these challenges, plant breeding programs have widely adopted statistical methodologies to quantify and interpret G×E interaction, as well as to assess the stability and adaptability of genotypes across environments. Traditional models such as analysis of variance (ANOVA), regression-based methods, and AMMI (Additive Main Effects and Multiplicative Interaction) (GAUCH; ZOBEL, 1988) have provided valuable tools for understanding phenotypic responses under multi-environment trials (METs). These methods, however, often fall short in handling unbalanced datasets and incorporating biological prior information.

In recent years, mixed models have emerged as a powerful alternative for G×E analysis. They allow the inclusion of both fixed and random effects, accommodate heterogeneity of variances, and provide reliable predictions of genetic merit through the estimation of Best Linear Unbiased Predictors (BLUPs) (HENDERSON, 1975; PIEPHO et al., 2008). Moreover, the use of Residual Maximum Likelihood (REML) to estimate variance components increases the robustness of these models (PATTERSON; THOMPSON, 1971). When combined with pedigree or genomic relationship matrices, mixed models offer an additional advantage, the ability to infer performance and stability of non-tested genotypes based on the performance of related individuals. This is particularly valuable in breeding pipelines where the direct evaluation of parental lines in METs is constrained by cost or logistical limitations (VELAZCO et al., 2019).

Maize (Zea mays L.) represents an ideal model for these advancements. As one of Brazil's most important crops for food, feed, biofuel, and industrial applications, maize has benefited immensely from hybrid breeding and the incorporation of stability metrics into selection strategies (PAES, 2006; MôRO; FRITSCHE-NETO, 2017). The dominance of hybrid varieties, especially triple-cross hybrids, reflects ongoing efforts to balance productivity with resilience in diverse growing conditions (SOUZA, 2018). Still, much of the evaluation in METs is centered on hybrid performance, with limited attention to the stability of their parental lines, an important limitation when aiming for long-term genetic gain.

To address these gaps, the present study introduces a methodological innovation in the Weighted Average of Absolute Scores (WAASB), a stability index derived from the singular value decomposition of BLUPs for G×E effects. Originally proposed by Olivoto et al. (2019a), WAASB combines the interpretability of AMMI with the predictive power of mixed models. However, its application was limited by its reliance on complete G×E matrices and its inability to assess non-evaluated genotypes, such as parental lines. In this study, we enhance WAASB by incorporating pedigree-based genetic relationship information into the mixed model framework. This integration enables the imputation of missing values in the G×E matrix based on genetic covariance and inference of parental line stability through progeny performance. These improvements render WAASB more robust and biologically meaningful, especially for breeding programs operating under data imbalance and limited resources.

1.1 Objectives

- To propose a methodological enhancement to the WAASB model by incorporating pedigree-based genetic relationship information, thereby improving the imputation of missing data and expanding the scope of stability analyses.
- To identify superior three-way maize hybrids from the UFV breeding program that outperform commercial checks in both performance and stability across multiple target environments.
- To infer the stability of parental inbred lines indirectly, based on the performance and stability of their hybrid progenies, facilitating more informed decisions for line advancement in breeding pipelines.

1.2 Literature review

1.2.1 Genotype-by-environment interaction

In plant breeding, experiments are conducted across multiple locations, growing seasons, and years. By considering the combined effect of genetic factors and environmental conditions, these studies enable a comprehensive and robust analysis of the performance of genotypes under selection or recommendation (ELIAS et al., 2016). However, multi-environment trials (METs) reveal a major challenge for breeders, the genotype-by-environment interaction (G×E) (MALOSETTI; RIBAUT; EEUWIJK, 2013).

Understanding G×E is crucial for advancing plant breeding, as it elucidates how different genotypes adapt or respond to varying cultivation conditions (CHAVES et al., 2023). This interaction introduces inconsistencies in genotype evaluation, complicating variety recommendation, given that the performance of a single genotype can vary significantly depending on the environment (MONTEVERDE et al., 2018). The variations in gene expression modulated by environmental conditions partially explain the phenotypic differences observed among varieties. This differential response indicates that specific genes may be activated or silenced in response to environmental stimuli, directly influencing plant phenotypic traits (WANG et al., 2017).

To address the challenges posed by G×E, various strategies have been adopted. One approach is the selection of genotypes specifically adapted to each environment, however, this strategy demands substantial investment and operational complexity. Ecological zoning, which groups regions with similar conditions, is another method employed, though it does not eliminate temporal variations associated with different growing seasons. Currently, the most effective solution involves identifying genotypes based on adaptability and stability, ensuring consistent performance across diverse cultivation conditions (CRUZ; REGAZZI; CARNEIRO, 2012).

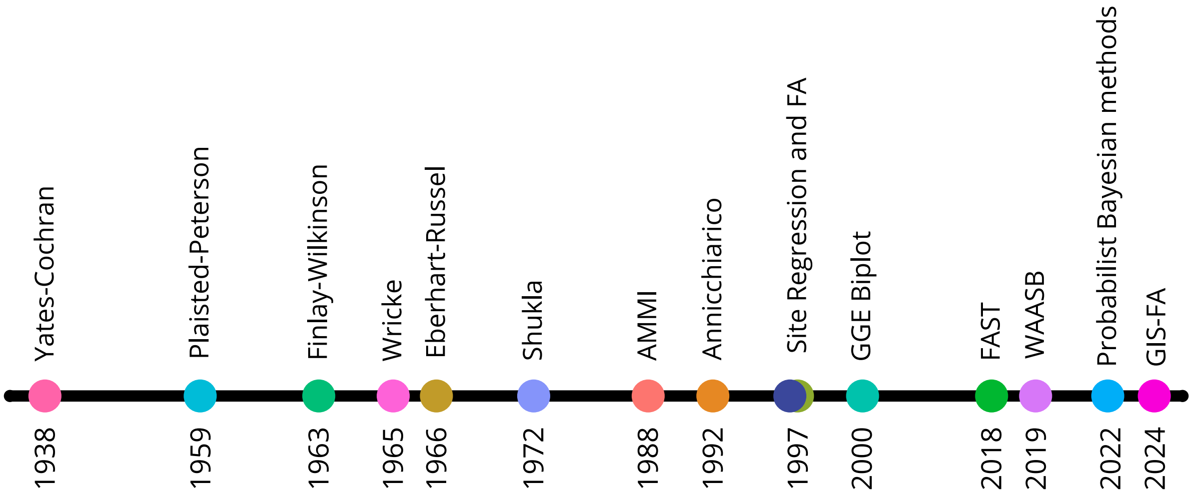
Adaptability refers to the ability of genotypes to respond advantageously to environmental variations, while the concept of stability varies depending on the evaluation method used (CRUZ; REGAZZI; CARNEIRO, 2012). Lin, Binns e Lefkovitch (1986) proposed four stability concepts: Type 1, where a genotype is considered stable if its variance across environments is minimal; Type 2, where stability is defined by a genotype's response being parallel to the mean performance of all genetic materials evaluated; Type 3, where stability is inferred from a low mean square deviation from regression, assessing environmental response consistency; and Type 4, where stability is determined by a low mean square of the genotype-by-year interaction within locations.

Several methodologies have been proposed for adaptability and stability studies, categorized according to their statistical approach. These include variance analysis-based models (YATES; COCHRAN, 1938; PLAISTED; PETERSON, 1959; WRICKE, 1965), linear regression-based models (FINLAY; WILKINSON, 1963; EBERHART; RUSSELL, 1966), non-parametric statistic-based models (SHUKLA, 1972; ANNICCHIARICO, 1992), multivariate analysis-based models (GAUCH; ZOBEL, 1988; CROSSA; CORNELIUS, 1997; YAN et al., 2000a), combined approaches integrating multivariate analysis with mixed models (SMITH; CULLIS, 2018; OLIVOTO et al., 2019a), more recently probabilist bayesian methods (DIAS et al., 2022; CHAVES et al., 2024) and enviromics based methods

(ARAúJO et al., 2024) (Figure 1).

A notable method in the stability analyses is the Weighted Average of Absolute Scores (WAASB) method proposed by Olivoto et al. (2019a) . WAASB integrates the principles of both AMMI (Additive Main Effects and Multiplicative Interaction) and BLUP (Best Linear Unbiased Prediction) models to improve stability analysis in METs. This index is derived from the singular value decomposition of the BLUPs for the G×E effects. Unlike conventional stability indices that rely on a limited number of interaction principal component axes (IPCA), WAASB incorporates all estimated IPCAs, ensuring a more comprehensive representation of stability. The method enhances genotype ranking by capturing the full complexity of G×E, thereby facilitating more reliable selection decisions in breeding program (OLIVOTO et al., 2019a).

Figure 1 – Timeline of the main methods for analyzing G × E interaction.



1.2.2 Mixed Models and Pedigree Information

Mixed models have become a fundamental statistical tool in plant breeding due to their ability to handle unbalanced datasets, account for genetic and environmental correlations, and improve the precision of genetic evaluations (PIEPHO et al., 2008). Unlike traditional fixed-effect models, mixed models incorporate both fixed and random effects, allowing for more accurate estimation of genetic values. The estimation of variance components in mixed models is commonly performed using the Restricted Maximum Likelihood (REML) method, which provides unbiased estimates of variance parameters by correcting for the loss of degrees of freedom associated with fixed effects (PATTERSON; THOMPSON, 1971).

A key application of mixed models in plant breeding is the estimation of Best Linear Unbiased Predictors (BLUPs) of genotypic values (HENDERSON, 1975). BLUP-based selection allows breeders to predict the genetic potential of individuals by accounting for environmental effects, improving the accuracy of selection in breeding trials. In the context of G×E interaction studies, mixed models facilitate the estimation of genotype performance across environments by incorporating heterogeneous genetic variance structures, leading to more reliable selection decisions (SMITH; CULLIS; THOMPSON, 2005).

One of the major advantages of mixed models is their ability to incorporate pedigree information through the relationship matrix, which models the genetic covariance among related individuals. By using pedigree data, breeders can improve the precision of genetic evaluations, particularly when phenotypic records are sparse or unbalanced (VELAZCO et al., 2019). This pedigree-based approach is widely used in conventional breeding programs to estimate breeding values and assess genetic variance components more accurately (CROSSA et al., 2006).

Beyond pedigree-based evaluations, mixed models have been extended to incorporate genomic information, enabling the implementation of genomic selection (GS) (VANRADEN, 2008). Genomic BLUP (GBLUP) replaces the pedigree-based relationship matrix with a genomic relationship matrix derived from molecular markers, further improving the accuracy of genetic predictions (SU et al., 2014). However, despite the increased prediction accuracy, the implementation of genomic selection remains challenging for some breeding programs due to the high costs associated with genotyping large populations. As a result, the feasibility of genomic selection must be carefully evaluated based on the economic and logistical constraints of each breeding program (WARTHA; LORENZ, 2021).

1.2.3 Maize Crop

Maize (Zea mays L.) is a crop of paramount importance due to its versatility and utilization across multiple areas, including human and animal nutrition, biofuel production, and various industrial sectors (PAES, 2006). Its grains are sources of carbohydrates, proteins, and fibers and can be consumed in diverse forms, such as boiled kernels, flour, and as secondary ingredients in various food products (MôRO; FRITSCHE-NETO, 2017).

The first maize plants were domesticated from teosinte (Zea mays ssp.) in parts of Mexico and Central America. Indigenous populations in these regions selected plants with desirable traits, eventually cultivating maize varieties extensively used in their societies (MATSUOKA et al., 2002). Additionally, Kistler et al. (2018)suggest that, apart from the previously mentioned regions, parts of the Amazon also represent secondary domestication centers, which played a significant role in establishing maize cultivation in Brazil.

With technological advancements, maize cultivation has benefited significantly from agricultural biotechnology, including the development of genetically modified varieties resistant to herbicides and pests, such as Bt maize and Herculex. These innovations have reduced dependence on chemical pesticides, thus promoting more sustainable agricultural practices (YASSITEPE et al., 2021). Maize has also played an essential role in integrated Crop-Livestock-Forestry systems (ILPF), where it is utilized in rotation with pastures, enhancing soil structure, and assisting in reducing greenhouse gas emissions, thereby promoting more sustainable agricultural practices (BALBINO et al., 2012).

In the 2022/23 season, Brazil was the third-largest global maize producer, yielding approximately 131.9 million tons, following the United States and China, which produced 348 million and 277.2 million tons, respectively. In 2023, Brazil became the leading maize exporter, exporting 55 million tons, surpassing the United States, which previously held this position (USDA, 2023). The total planted area in Brazil was 21,081.4 thousand hectares, divided into three crop cycles: first, second, and third crops, with approximately 4,033 thousand, 16,417 thousand, and 631 thousand hectares, respectively. The average grain yield was 5,622 kg/ha.

In 2023, production reached 25,309.3 thousand tons for the first crop, 91,235.1 thousand tons for the second crop, and 2,334.6 thousand tons for the third crop. Maize cultivation spans the entire Brazilian territory, with Mato Grosso, Paraná, and Goiás being responsible for 33%, 15%, and 12% of national production, respectively (CONAB, 2023). Differences in productivity among these states result from variations in climate, topography, production systems, and technological levels, among other factors (COELHO; CRUZ; FILHO, 2004; MALDANER et al., 2014). Furthermore, climate change significantly impacts maize production, necessitating adaptive strategies such as early-maturing varieties and irrigation technologies to mitigate adverse effects (ZHANG et al., 2022).

1.2.4 Maize Varieties

Choosing the appropriate maize variety is a critical step in establishing and achieving the success of a maize crop, as this decision significantly influences final productivity. Therefore, factors such as productive potential, the technological level to be adopted, investment in agricultural inputs by the producer, as well as variety stability and adaptability must be considered. There are two main types of maize varieties: open-pollinated varieties (OPVs) and hybrids (SILVA et al., 2015).

Open-pollinated varieties consist of selected individuals that mate randomly. Generally, they have greater stability and adaptability but lower yield potential compared to hybrids due to a broader genetic base and less effective exploitation of heterosis.

Furthermore, OPVs remain in Hardy-Weinberg equilibrium; hence, in the absence of mutation, selection, and migration, allele frequencies remain constant, allowing producers to use part of their grain production as seeds for subsequent planting without significant productivity losses. These varieties are predominantly grown by producers utilizing lower technological input levels (BORéM; MIRANDA; FRITSCHE-NETO, 2017).

The first maize hybrids were developed in the late 19th century through George H. Shull's research in the United States. Shull (1908) described experiments involving the crossing of inbred lines, demonstrating that resulting plants were more productive and vigorous than their parental lines. This phenomenon, known as heterosis or hybrid vigor, refers to hybrids exhibiting superior characteristics compared to the parental average, including increased growth, disease resistance, and yield.

Despite their higher productivity, the adoption of single-cross hybrids was initially hindered by difficulties in seed production due to low productivity of parental lines. As a solution, double-cross hybrids, involving four parental lines instead of two, became popular. Producers then transitioned from open-pollinated varieties to double-cross hybrids (BERNARDO, 2021). Commercial utilization of single-cross hybrids became feasible only after recurrent selection strategies were adopted to enhance seed production efficiency from the base populations. Since the 1960s, single-cross hybrids have become predominant maize varieties (BERNARDO, 2021).

In Brazil, initial maize hybrid breeding efforts began in the 1930s at the Federal University of Viçosa (UFV) and the Agronomic Institute of Campinas (IAC). At IAC, the widely popular Cateto variety was used to produce inbred lines, which subsequently led to the development of the first hybrids characterized by hard, orange kernels. Concurrently, researchers Antônio Secundino and Gladstone de Almeida at UFV initiated a hybrid maize breeding program, crossing Cateto and dent maize inbred lines to produce semi-dent hybrids. These semi-dent hybrids demonstrated significantly higher productivity compared to hybrids derived solely from Cateto lines developed by IAC, thus gaining popularity in Brazil. The two researchers from Viçosa founded Agroceres, the first private company dedicated to hybrid maize development in Brazil (FILHO et al., 2000).

In the second half of the 20th century, the widespread adoption of single-cross hybrids resistant to diseases, tolerant to environmental stresses, and with higher productivity significantly boosted maize production in Brazil (SILVA et al., 2015). Initially, the adoption of these hybrids was limited to regions where producers had resources to adopt advanced agricultural practices, mainly concentrated in the South and Southeast regions. Their recognized productivity and profitability continue to make them the most cultivated varieties today (SCAPIM, 2009). According to Embrapa researchers, as of 2022, due to agricultural expansion and improved breeding programs, hybrids represent more than 95% of maize seeds sold in Brazil, with 50.19% being single-cross hybrids, 1.93%

triple-cross hybrids, 1.16% double-cross hybrids, and the remaining varieties not specified by companies.

The primary objective of hybrid production is to exploit heterosis by crossing genetically distinct and complementary inbred lines (MôRO; FRITSCHE-NETO, 2017). Hybrids are classified as single, triple, and double hybrids, produced by crossing two, three, or four inbred lines, respectively. A greater number of parental lines involved in the cross increases genetic variability, providing adaptability to adverse conditions but decreasing yield potential (SOUZA, 2018).

Single-cross hybrids possess a narrower genetic base and present the highest yield potential among maize varieties, although with lower adaptability to unfavorable conditions compared to other hybrids. They exhibit uniform characteristics such as plant height, maturity, and grain quality. These hybrids are recommended for producers who invest heavily in agricultural inputs, as seed production involves less productive and disease-susceptible inbred lines, leading to higher seed costs. They also demand high fertilizer use and pest control (SOUZA, 2018).

Double-cross hybrids result from crossing two single-cross hybrids and typically exhibit higher resilience to adverse environmental conditions due to their broader genetic base, although with lower yield potential compared to other hybrids (SOUZA, 2018).

Triple-cross hybrids, obtained by crossing a single-cross hybrid with an additional inbred line, usually have enhanced stability compared to single-cross hybrids, enabling consistent productivity under environmental stresses such as temperature fluctuations, water availability, and disease incidence. They also exhibit high yield potential, making them suitable for producers facing environmental challenges. Economically, triple-cross hybrids are attractive because seed production is more efficient and cost-effective, as seeds originate from a productive single-cross hybrid rather than less productive inbred lines. This economic advantage, combined with resilience and adaptability, makes triple-cross hybrids a strategic and accessible choice for producers seeking to balance productivity and cost (SOUZA, 2018).

1.2.5 Environmental factors driving genotype-by-environment interaction

Maize cultivation is widely disseminated across Brazil, with diverse maize varieties adapted to various climatic conditions, reflecting significant genetic variability. This genetic diversity enables maize to respond effectively to different selective pressures, favoring genotypes carrying favorable alleles for specific environmental conditions. In other words, maize's extensive genetic variability facilitates the development of genotypes adapted to a wide range of abiotic and biotic stresses (JONG; BREWBAKER; LEE, 1982;

BADU-APRAKU et al., 2011; SSERUMAGA et al., 2018; REZENDE et al., 2020).

In low-altitude tropical regions of Brazil (up to 700 meters above sea level), maize growth and productivity are strongly influenced by abiotic factors. High solar radiation, nighttime temperatures, and humidity represent considerable challenges for maize production, negatively impacting grain yield (JONG; BREWBAKER; LEE, 1982; DURãES, 2006). Temperature, particularly, is critical for maize metabolic processes. The optimal daytime temperature of around 30°C promotes growth and leaf emergence, enhancing photosynthesis and grain filling (DURãES, 2006). However, elevated nighttime temperatures (>24°C) increase plant respiration rates, leading to higher energy consumption and reduced net photosynthetic rates, consequently decreasing biomass accumulation and grain yield (SUNOJ et al., 2016; WANG et al., 2020). Thus, temperature amplitude, more pronounced in low-altitude regions, is crucial for maize production (ALAM et al., 2017; WANG et al., 2020).

Higher-altitude regions experience lower night temperatures, potentially delaying plant growth and development. Reduced air density at higher altitudes can affect photosynthetic rates and water-use efficiency. Additionally, increased ultraviolet radiation can damage plant tissues and reduce productivity. Early autumn or late spring frosts can further compromise maize crops, while mountainous soils often present limitations in depth and nutrient availability (FAO, 2018).

Water stress particularly reduces the conversion of light energy into biomass and lowers stomatal conductance, negatively impacting grain productivity (BENEšOVá et al., 2012). Extended drought periods during critical phenological stages can cause grain yield losses exceeding 70% (BERGAMASCHI et al., 2004). Costa et al. (1988) reported that reductions in water availability significantly decreased maize productivity parameters, mainly due to reduced leaf area and associated physiological processes such as photosynthesis and solute translocation to grains.

According to Ribaut et al. (2009), critical maize growth stages include germination and emergence, during which soil compaction or low temperatures can reduce plant populations. During early vegetative growth, water deficits significantly impair crop development. Pollination is particularly sensitive, as water scarcity or high temperatures can result in poor pollination and reduced grain number. Finally, the grain-filling stage is vital for final grain weight and can be negatively impacted by nutritional deficiencies or diseases, thus affecting both grain quality and yield (RIBAUT et al., 2009).

Soil nutrient availability significantly influences maize growth and productivity (MAFOUASSON et al., 2018). Nitrogen (N), in particular, is essential for maximizing grain yield (ADRIAANSE; HUMAN, 1993). Maize responses to nitrogen fertilization depend on soil type and climatic conditions (FERNANDES et al., 2005). Additionally, maize genotypes differ in their nitrogen responsiveness (BADU-APRAKU et al., 2011). Efficient nitrogen use is crucial for reducing fertilizer costs, minimizing environmental losses through leaching and volatilization, enhancing grain productivity, and reducing environmental impacts (BALIGAR; FAGERIA; HE, 2001; HAEGELE et al., 2013; RODRIGUES et al., 2017). In research by rodrigues (2017), commercial maize hybrids were evaluated under low and high nitrogen availability conditions, demonstrating that hybrids responded differently in terms of yield, highlighting the importance of adjusting management practices based on genotype demands and soil conditions.

Given the variability in cultivation environments and environmental influences, it is crucial to develop maize varieties adapted to broad and specific conditions, exhibiting drought tolerance, resistance to pests and diseases, and high productivity under diverse soil and climatic conditions (DHAKAL et al., 2022). Studies on triple hybrids in drought-prone areas indicate some hybrids possess significant genetic potential for early maturity and drought resistance, underscoring the importance of selecting suitable genotypes for each environmental condition (IZGE; DUGJE, 2011).

2 An enhanced WAASB: relationship information can improve multi-trait multi-environment selection of untested maize lines

2.1 Abstract

In this chapter, we show how a model enriched with relationship between genotypes information can benefit WAASB i) by directly imputing missing values in the GEI matrix based on the genetic covariance; and ii) by estimating the stability of parental lines through their progenies' performance. We used grain yield (GY), plant height (PH) and ear height (EH) data from 160 three-way hybrids and seven commercial checks evaluated in eight trials (environments) laid out alpha-lattice. We built a numerator relationship matrix from the pedigree information, and used it in a compound symmetry model to obtain the GEI matrix. This matrix was used to compute the genotypes' WAASB. We also built a selection index that considered both performance and stability of all three traits. To assess how efficient the imputation via genetic covariance was in relation to a general imputation method (EM-SVD), we simulated 200 datasets in four scenarios of high and low heritability and GEI. In each dataset, we randomly removed values of genotype- environment combinations to mimic data imbalance. Our analytical pipeline provided means for selecting parental lines, single cross hybrids and three-way hybrids. Considering only line selection, the expected genetic gains for GY, PH and EH are 4%, 2% and -2%. The imputation using genetic covariance outperformed the EM-SVD method in all scenarios, and provided more reliable values. The enhancement proposed to the WAASB accredits it as an efficient yet practical tool that can be integrated into routine decision-making in breeding programs.

2.2 Introduction

Maize breeding relies on multi-environmental trials (METs) to properly estimate the stability of selection candidates before recommending for the target population of environments (TPE). METs are mandatory in breeding pipelines due to the genotype-by-environment interaction (GEI), a confounding factor that impedes the selection/recommendation for a given location to be valid in another one with different environmental conditions (CROSSA, 2012; SMITH; CULLIS; THOMPSON, 2005). This is mostly true when GEI is of rank-type, meaning that there is change of rank between locations, i.e., the genetic correlation between environments is low/negative (AASTVEIT; AASTVEIT, 1993; POUPON et al., 2023). A rank-type GEI is the rule for polygenic traits, which are the most commercially important in breeding programs (COOPER; DELACY, 1994).

Stability is a complex genetic trait, and as such, can be a breeding target. For a genotype to be considered for recommendation in the TPE, it needs to hold high-performance alleles, and further genes that allow this performance to be constant, irrespective of environmental conditions (stability types I and II); or, at least, predictable, i.e., responsive to good environmental inputs, and resilient to harmful ones (stability type III) (LIN; BINNS, 1994). Identifying such ideotypes is not an easy task, and several statistical approaches have been proposed so far (SMITH; CULLIS; THOMPSON, 2005; MALOSETTI; RIBAUT; EEUWIJK, 2013; CROSSA, 1990). WAASB (Weighted Average of Absolute Scores of BLUPs)(OLIVOTO et al., 2019a) is one of the latest. It is based on the principles of AMMI (Additive Main effects and Multiplicative Interaction) (JR., 1988), as it also decomposes the two-way genotype-by-environment table (environments in the row, genotypes in the column) using singular value decomposition (SVD). The difference is that WAASB uses a single value built with all principal components (PCs) to indicate a genotype's stability, taking the explicative power of each PC into consideration. This simplifies the decision-making and allows stability to be inserted into selection indices such as the WAASBY, for example (OLIVOTO et al., 2019b). For being such a simple yet effective method, and because it is implemented in an open-source package `metan` (OLIVOTO; LúCIO, 2020), WAASB has been broadly used for obtaining candidates' stability in various species, such as wheat (CASAGRANDE et al., 2024), soybean (NATARAJ et al., 2021) and sorghum (BEHERA et al., 2024).

The original proposal of WAASB has two gaps: first, it does not allow for the assessment of the parental lines' stability, and it is sensible to imbalance, as the GEI matrix needs to be complete for the SVD decomposition to be possible (JR.; ZOBEL, 1990). In the original approach, which did not account for relationship information, values in the GEI matrix were imputed using EM-SVD (TROYANSKAYA et al., 2001) to address any imbalance in the dataset. This imputation was necessary to fill gaps because compound symmetry models do not predict candidate performance in environments where they were not tested, provided no relationship information is incorporated. While imputation is a viable statistical solution that performs well despite introducing additional uncertainty into the analyses (ANGELINI; CERVIGNI; QUAGLINO, 2024; PADEREWSKI; RODRIGUES, 2014), we hypothesize that a statistical-genetic approach would be more adequate. By incorporating relationship information (whether derived from pedigree or genomic sources), the model can "impute" missing values based on genetic covariances among related genotypes (PIEPHO et al., 2008; MISZTAL; LEGARRA; AGUILAR, 2009). This approach eliminates the need for explicit imputation, as including a relationship matrix in the model naturally accounts for these gaps, offering a more biologically grounded and efficient solution.

Incorporating relationship-enriched models would also enable the evaluation of parental lines' performance and stability through their progenies. This approach addresses a key limitation in breeding programs that lack the financial resources to test lines in METs throughout the breeding pipeline. Typically, in such programs, lines are evaluated in only a few environments, and only hybrids derived from lines with the best general combining abilities (GCAs) undergo thorough MET evaluation before recommending for the whole TPE (DIAS et al., 2020; FARIA et al., 2022). In Brazil, this evaluation is mandatory in value of cultivation and use (VCU) trials before a hybrid can be released as a commercial cultivar. Notably, this process overlooks the stability of parental lines. In such context, and in the current implementation, WAASB could only assess the stability of hybrids, as they are the only entries with collected data. However, since stability is an inheritable trait, valuable information is being lost. By integrating relationship information into the model, METs with hybrids would serve a dual purpose: selecting high-performing and stable hybrids while simultaneously identifying parental lines that contributed favourable alleles for both performance and stability (RINCENT et al., 2019). Intuitively, this would lead to the selection of lines for parenting the next generation that not only enhance hybrid performance but also promote stability, ensuring long-term genetic gains.

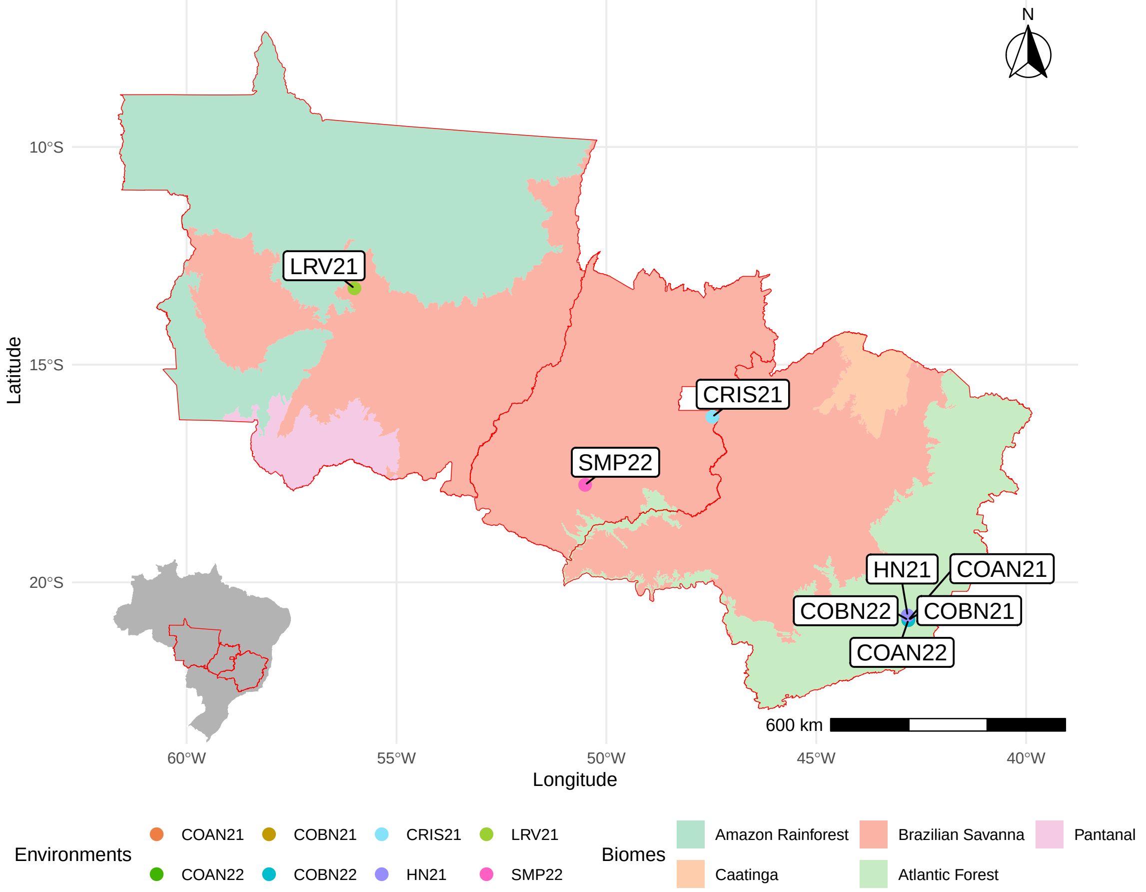
This study aims to test the two previously described hypotheses by presenting a comprehensive multi-environment, multi-trait analytical pipeline for maize breeding. First, we demonstrate how WAASB, derived from a relationship-enriched model, can effectively fill gaps in the GEI matrix while providing stability estimates for parental lines. Next, using a selection intensity of 20%, we identify the top parental lines, F_1 hybrids, and three-way hybrids based solely on the latters' performance data, considering three key traits—grain yield, plant height, and ear height—along with their respective stability measures. Finally, we use simulated datasets to compare imputation based on genetic covariance (statistical-genetic principles) with EM-SVD-based imputation (purely statistical principles), evaluating their effectiveness under different missing data scenarios.

2.3 Material and Methods

2.3.1 Plant material

A hundred and sixty hybrids derived from three-way crosses were evaluated across eight tropical environments in Brazil (Table 1, Figure 2). These hybrids were generated using 19 parental lines. Initially, the lines were intercrossed to produce 61 single cross hybrids, which were subsequently crossed with different lines (not used as parents in the first crossing) to create the 160 three-way hybrids. Each environment was defined as a combination of location and year, capturing both spatial and temporal variability. The field trials were conducted using an alpha-lattice design with two replications, comprising 11 blocks and 15 genotypes per block. Alongside the experimental hybrids, a set of commercial checks was included for benchmarking purposes. The checks—DKB230PRO3, BM709PRO2, B2688PW, B2433PWU, 22S18TOP3, VA42B, and BM270—varied across environments.

Figure 2 – Geographic distribution of the eight field trials conducted for evaluating 160 hybrids. Distinct colored points mark each location, while the biomes are the shaded areas. The inset map provides spatial context, indicating the position of these states within Brazil.



Two of the four trials conducted in Coimbra were established under low-nitrogen conditions. For these trials, only the initial fertilization was applied (380 kg ha^{-1} of N-P-K 08-28-16 at sowing). In all other environments, location-specific best management practices recommended for the species were followed. We measured the plant height (in cm), ear height (in cm) and the grain yield (GY) of each plot, adjusted to 14.5% of grain moisture, and converted the values to kg ha^{-1}.

Table 1 – Description of the eight environments where 160 hybrids and seven checks were evaluated: geographical location, coordinates (Longitude, Latitude and Altitude, in meters above sea level), and year and season in which it was conducted.

Environment	Location	Longitude	Latitude	Altitude	Year	Season
COAN21	Coimbra (MG)	42º 48’ 30” W	20º 50’ 30” S	713	2021	First
COBN21	Coimbra (MG)	42º 48’ 30” W	20º 50’ 30” S	713	2021	First
COAN22	Coimbra (MG)	42º 48’ 30” W	20º 50’ 30” S	713	2022	First
COBN22	Coimbra (MG)	42º 48’ 30” W	20º 50’ 30” S	713	2022	First
HN21	Viçosa (MG)	42º 49’ 26” W	20º 45’ 40” S	661	2021	First
SMP22	Santa Helena de Goiás (GO)	50º 30’ 23.1” W	17º 45’ 35.5” S	562	2022	First
CRIS21	Cristalina (GO)	47º 28’ 24” W	16º 11’ 50” S	994	2021	Second
LRV21	Lucas do Rio Verde (MT)	56º 00’ 11” W	13º 14’ 44” S	417	2021	Second

2.3.2 Statistical procedures

In the mathematical notations below, V is the number of genotypes ($v \in 1, 2, \ldots, V$), M is the number of environments ($m \in 1, 2, \ldots, M$), and N is the number of plots (with $N = \sum_m^M N_m$). All analyses were done in the R software environment - version 4.2 (R Core Team, 2024) - and the plots built using the `ggplot2` package (WICKHAM, 2016). We solved the linear mixed models (HENDERSON, 1975) and obtained variance component estimates (via residual maximum likelihood) (PATTERSON; THOMPSON, 1971) and BLUPs (best linear unbiased predictions) using the `R` package `asreml` (The VSNi Team, 2023).

2.3.2.1 Individual analyses

We first performed a particularized assessment of each environment. The following model was used:

$$\boldsymbol{y} = \boldsymbol{1}\mu + \boldsymbol{X}\boldsymbol{r} + \boldsymbol{Z}_1\boldsymbol{g} + \boldsymbol{Z}_2\boldsymbol{b} + \boldsymbol{\varepsilon} \tag{2.1}$$

where $\boldsymbol{y}$ is the $N_m \times 1$ vector of phenotypic records, μ is the intercept, $\boldsymbol{r}$ is the 2×1 vector of fixed repetition effects, $\boldsymbol{g}$ is the $V \times 1$ vector of random additive-genetic effects, distributed as $\boldsymbol{g} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_g^2 \boldsymbol{A}\right)$, where $\boldsymbol{0}$ is a vector of zeros, σ_g^2 is the additive-genetic variance and $\boldsymbol{A}$ is the $V \times V$ numerator relationship matrix, built from the pedigree using the package `AGHmatrix` (AMADEU et al., 2023); $\boldsymbol{b}$ is the 30×1 vector of random block effects, distributed as $\boldsymbol{b} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_b^2 \boldsymbol{I}_{30}\right)$, with σ_b^2 representing the variance of blocks, and $\boldsymbol{I}_{30}$, an identity matrix whose order is given by its subscript; and $\boldsymbol{\varepsilon}$ is the $N_m \times 1$ vector of residual effects $[\boldsymbol{\varepsilon} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_\varepsilon^2 \boldsymbol{I}_{N_m}\right)$, with σ_ε^2 being the residual variance]. The upper case letters $\boldsymbol{X}$, $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are the incidence matrix, whose number of columns is the same as the size of its respective vector, and number of rows is N_m.

Using Eq. 2.1, we checked the significance of the additive-genetic effects via the likelihood ratio test (LRT). We also computed the narrow-sense heritability (CULLIS; SMITH; COOMBES, 2006) and the coefficient of experimental variation as follows, respectively:

$$h_m^2 = 1 - \frac{\bar{\Delta}}{2\sigma_g^2} \tag{2.2}$$

$$CV_m = \frac{\sqrt{\sigma_\varepsilon^2}}{\mu} \tag{2.3}$$

where $\bar{\Delta}$ is the pairwise prediction error variance.

2.3.2.2 Multi-environment analyses

We used the following compound symmetry model to fit the multi-environment data:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{X}_1\boldsymbol{e} + \boldsymbol{X}_2\boldsymbol{r} + \boldsymbol{Z}_1\boldsymbol{g} + \boldsymbol{Z}_2\boldsymbol{ge} + \boldsymbol{Z}_3\boldsymbol{b} + \boldsymbol{\varepsilon} \tag{2.4}$$

where $\boldsymbol{e}$ is the $M \times 1$ vector of fixed environment effects, and $\boldsymbol{ge}$ is the $VM \times 1$ vector of genotype-by-environment interaction (GEI) effects, distributed as $\boldsymbol{ge} \sim \mathcal{N}\left(\mathbf{0}, \sigma_{ge}^2 \boldsymbol{I}_M \otimes \mathbf{A}\right)$, where σ_{ge}^2 is the GEI variance. All the other terms were previously described in Eq. 2.1, and have new dimensions in Eq. 2.4. $\boldsymbol{y}$ is a $N \times 1$ vector, so all incidence matrices have N rows. $\boldsymbol{r}$ and $\boldsymbol{b}$ have nested effects, so their original dimension is multiplied by the number of environments. Finally, we assumed that $\boldsymbol{\varepsilon}$ followed a block-diagonal structure, i.e. $\boldsymbol{\varepsilon} \sim \mathcal{N}\left(\mathbf{0}, \oplus_{m=1}^{M} \sigma_{\varepsilon_m}^2 \otimes \boldsymbol{I}_{N_m}\right)$, where $\oplus$ is the direct sum symbol.

Using the variance components provided by Eqs. 2.1 and 2.4, we quantified the proportion of rank- and scale-type GEI in each trait as follows (COOPER; DELACY, 1994):

$$\sigma_{ge_{scale}}^2 = 1 - \frac{Var\left(\sqrt{\sigma_{g_m}^2}\right)}{\sigma_{ge}^2} \tag{2.5}$$

2.3.2.3 AMMI and WAASB

We transformed $\boldsymbol{ge}$ from a vector of dimensions $VM \times 1$ (long format) into a centred matrix with dimensions $V \times M$ (wide format). This matrix, here called the GEI matrix, is denoted by $\boldsymbol{\Phi}$. Following an AMMI (Additive main effects and multiplicative interaction) framework, the GEI matrix was decomposed using singular value decomposition, as follows:

$$\boldsymbol{\Phi} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}' \tag{2.6}$$

where $\boldsymbol{\Lambda}$ is a $P \times P$ diagonal matrix of singular values, with $P \leq \min(V-1, M-1)$ and $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthonormal matrices with the singular vectors of $\boldsymbol{\Phi}\boldsymbol{\Phi}'$ and $\boldsymbol{\Phi}'\boldsymbol{\Phi}$, respectively. The matrices of the principal component scores for genotypes and the loadings for environments were calculated as $\boldsymbol{W}_g = \boldsymbol{U}\boldsymbol{\Lambda}$ (dimension $V \times P$) and $\boldsymbol{W}_e = \boldsymbol{V}\frac{\boldsymbol{\Lambda}}{\sqrt{M-1}}$ (dimension $M \times P$), respectively. We used the first two rows of $\boldsymbol{W}_g$ and $\boldsymbol{W}_e$ to perform AMMI-like analyzes, ie, AMMI1 (first principal component vs. trait means) and AMMI2 (first vs second principal components). Next, we used the same matrices to compute the weighted average of absolute scores of GEI's BLUPs (WAASB) (OLIVOTO et al., 2019a), given by:

$$WAASB_v = \frac{\sum_{p=1}^{P} |w_{pv} \times \tau_p|}{\sum_{p=1}^{P} \tau_p} \tag{2.7}$$

where w_{pv} is the score of candidate v in the p^{th} component, and τ_p is the variance explained by the p^{th} component. The lower the WAASB, the more stable the genotype. Conclusions drawn from WAASB have one advantage over AMMI: they are based on all the principal components, rather than only the first two. We also computed WAASBY (OLIVOTO et al., 2019a), an index that provides unified information on genotype mean performance and stability (computed using WAASB). The WAASBY was computed as follows:

$$WAASBY_v = \frac{\breve{g}_v \times \theta + \breve{w}_v \times (1 - \theta)}{\theta + (1 - \theta)} \tag{2.8}$$

where $\breve{g}_v$ and $\breve{w}_v$ represent the performances and WAASBs rescaled to range from 0 to 100, and θ is a weight that determines the importance of both performance (θ per se) and stability $(1 - \theta)$. Here, we assumed that $\theta = 0.75$. The WAASBY is necessary to compute an index that will aid in a selection that considers per-trait performance and stability, described next.

2.3.2.4 MTSI

Essentially, the multi-trait stability index (MTSI) (OLIVOTO et al., 2019b) is a genotype-ideotype Euclidean distance based on factor analysis. To compute the MTSI, we first computed the correlation between the WAASBY values of each trait. Let $\boldsymbol{Q}$ be the $V \times M$ WAASBY matrix. The correlation was computed by:

$$\boldsymbol{R} = \boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{D}^{-1} \tag{2.9}$$

where $\boldsymbol{D}$ is a diagonal matrix whose elements are the standard deviation of the columns of $\boldsymbol{Q}$, and $\boldsymbol{C}$ is the matrix of covariances between the WAASBYs of different traits. Next, we performed the eigendecomposition of $\boldsymbol{R}$, as follows:

$$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U} \tag{2.10}$$

where $\boldsymbol{U}$ is the matrix of eigenvectors, and $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues. We retained two eigenvalues, for they explain 88% of the total variance when combined (64% + 24%); and their corresponding eigenvectors. The K remaining eigenvectors composed the initial loadings, as follows:

$$\boldsymbol{L} = \boldsymbol{U}_K\boldsymbol{\Lambda}^{\frac{1}{2}} \tag{2.11}$$

where $\boldsymbol{U}_K$ contains the first K eigenvectors, and $\boldsymbol{\Lambda}^{\frac{1}{2}}$ represent the square root of the K retained eigenvalues. $\boldsymbol{L}$ was rotated using the varimax method to ensure interpretability. After rotation, the canonical loadings (say, $\ddot{\boldsymbol{L}}$) were obtained, from which the matrix of scores ($\boldsymbol{F}$) was computed, as follows:

$$\boldsymbol{F} = \breve{\boldsymbol{Q}}\left(\ddot{\boldsymbol{L}}'\boldsymbol{R}\right)' \tag{2.12}$$

where $\breve{\boldsymbol{Q}}$ is a matrix of standardized WAASBY means.

The next step was to obtain the ideotypes' scores. Recall that when we computed the WAASBYs (Eq. 2.8), the means and WAASBs were rescaled to 0 to 100. In other words, irrespective of the selection objective (i.e., whether to increase or decrease the trait mean), the desirable WAASBY will always be 100. Therefore, the matrix of ideotypes is a vector of 100s. This vector substitutes $\breve{\boldsymbol{Q}}$ in Eq. 2.12 to obtain the ideotypes' scores. Finally, the MTSI is computed as:

$$MTSI_v = \left[\sum_{k=1}^{K}\left(F_{vk} - \acute{F}_k\right)^2\right]^{\frac{1}{2}} \tag{2.13}$$

where F_{vk} is the score of the v^{th} candidate in the k^{th} factor, and $\acute{F}_k$ is the score of the ideotype in the k^{th} factor. A desirable candidate has a low MTSI, meaning a closer distance to the ideotype.

We built two functions to reproduce the analysis performed herein. These functions provide the same outputs but use different resources to solve the compound symmetry mixed model: one uses `asreml` , and the other uses `sommer` (COVARRUBIAS-PAZARAN, 2016). The original proposal (without using relationship information) is implemented in the package `metan` (OLIVOTO; LúCIO, 2020), and leverages resources of the package `lme4` (BATES et al., 2015). The difference in algorithms may result in discrete differences. `metan` is freely available from CRAN.

2.3.2.5 Simulation

The simulation was conducted using the `AlphaSimR` (GAYNOR; GORJANC; HICKEY, 2021) and `FieldSimR` (WERNER; GEMENET; TOLHURST, 2024) packages. We first ran a 20-year burn-in phase of a standard maize breeding pipeline, see Figure 4 and Table 2 for our specific setup (BANčIč et al., 2025), then used the resulting genetic structure to generate three-way hybrids. With this dataset, we assessed the accuracy of data imputation relative to relationship-based approaches for filling gaps in $\boldsymbol{\Phi}$.

The founder population was simulated using the Markovian Coalescence Simulator (CHEN; MARJORAM; WALL, 2009). The genome consisted of ten chromosomes, following the default maize parameters in the `runMacs` function of `AlphaSimR`, which simulates a 200 cM genome ($\sim 2 \times 10^8$ bp) with mutation and recombination rates of 5×10^{-6} and

Table 2 – Details of the parameters used for in simulating a maize breeding pipeline, following Bančič, Gorjanc e Tolhurst (2024) and Bernardo (2009). The term "YT" refers to Yield Trial, "No. Envs." is the number of environments, "No. Reps." is the number of replicates per trial, and h^2 is the narrow-sense heritability. The pipeline begins with 40 parents—20 males (pool A) and 20 females (pool B)—intercrossed to produce 100 F_1 hybrids per year. From each hybrid, 50 doubled haploid (DH) lines are derived. These DH lines undergo sequential testcross evaluations: first with a single tester, where 400 lines are selected, and then with three testers, reducing the selection to 40 lines. Next, the selected lines from each pool are crossed with the top ten elite parents from the opposite pool, generating 20 hybrids. These hybrids undergo two sequential field trials, where seven hybrids are selected in the first stage, and one final hybrid is chosen in the second stage. Additionally, the trials identify the top ten DH lines with the highest general combining ability, which will serve as parents for the next breeding cycle.

Year	Stage	No. Envs	No. Reps.	h^2	Detail
1	Parent	-	-	-	20 parents. Each year, the top 10 parents substitute the bottom 10.
2	F1/DH/YT1	1	1	0.06	Produce 50 DHs per F1 cross (100 crosses are done per year). Testcross with one tester and select 400 DH lines.
3	YT2	3	2	0.11	Testcross with three testers. Select 40 DH lines.
4	YT3	10	3	0.2	Cross the DH lines with 10 elite parents. Evaluate and select 40 hybrids.
5	YT4	20	5	0.34	Hybrid field trial, selection of top 10 elite parents for the next cycle and seven hybrids for pre-release trials.
6	YT5	50	15	0.8	Pre-release trial, selection of one hybrid for release.

4×10^{-6}, respectively, and an initial effective population size of 100 (HICKEY et al., 2014). Each chromosome had 1700 segregating sites and 300 QTLs, representing a highly polymorphic trait. The founder population included 40 maize inbred lines (20 males, 20 females), with mean and variance parameters reflecting grain yield values observed in our empirical data set. Dominance deviation was set at 0.92 (mean) and 0.3 (variance) (BANčIč et al., 2025).

We then simulated a target population of environments (TPE) comprising 1000 environments. An unstructured covariance matrix was used, with mean environmental correlations of 0.8 (low GEI) and 0.2 (high GEI) (POUPON et al., 2023; AASTVEIT; AASTVEIT, 1993). Throughout the breeding pipeline, environments were sampled from the TPE for multi-environment trials (MET), with the number of environments per MET varying by breeding stage (Table 2). Each stage included randomized complete block trials (number of blocks varying by stage) with plot-level errors, which were modelled as heterogeneous across the MET. For further details on the simulation framework for plot-level errors and TPEs, see Bančič, Gorjanc e Tolhurst (2024) and Werner, Gemenet e Tolhurst (2024). While the example made available by Bančič, Gorjanc e Tolhurst (2024) focused on a wheat line-breeding pipeline, we adapted their methodology to a hybrid breeding pipeline following Bančič et al. (2025).

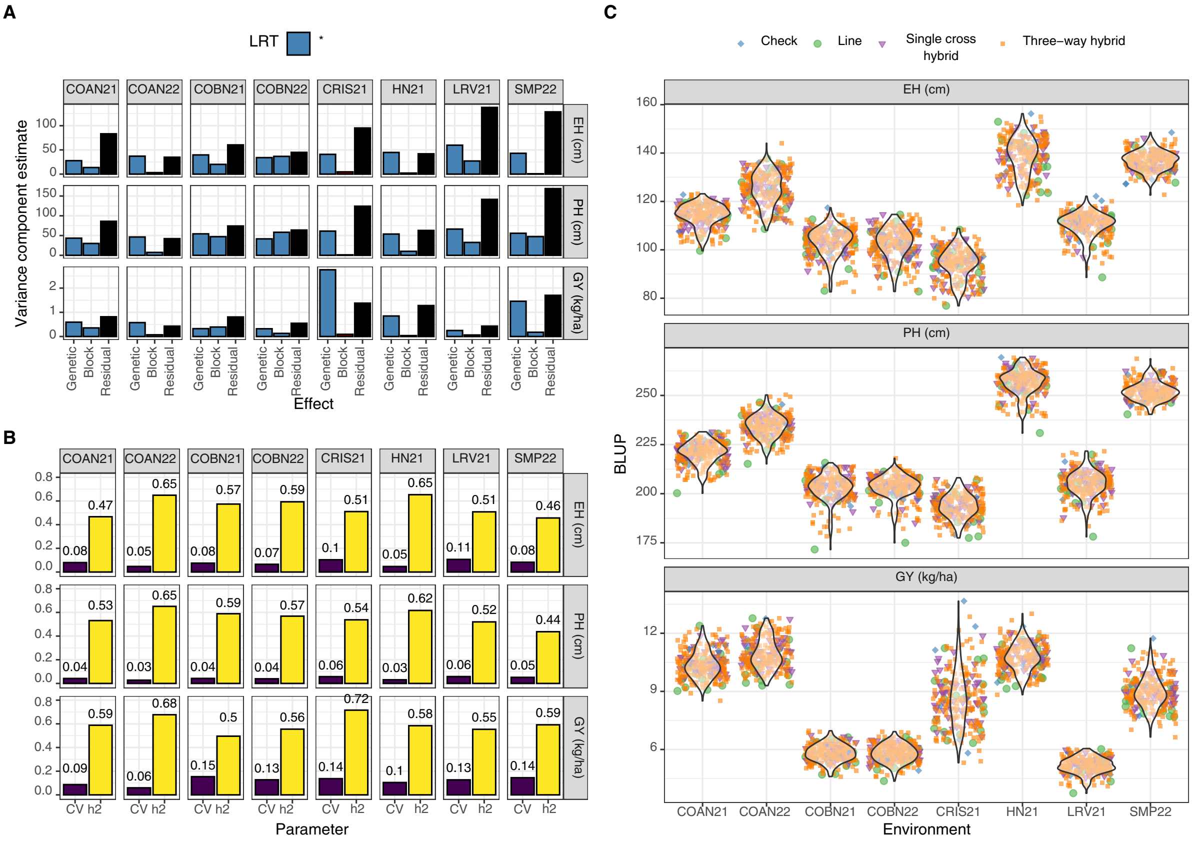
After the burn-in period, we used the selected parents for a new breeding cycle to generate F_1 hybrids, followed by three-way hybrids. At this stage, we simulated data sets under two heritability scenarios: low ($h^2 = 0.2$) and high ($h^2 = 0.8$). For each scenario, we sampled 50 environments from the TPE to form the MET. This resulted in four test scenarios: high heritability with high GEI (HHHG), low heritability with high GEI (LHHG), high heritability with low GEI (HHLG), and low heritability with low GEI (LHLG). To simulate an imbalanced MET, we randomly removed data from 20% to 80% (in 5% increments) of genotype-environment combinations (mimicking a CV2 scheme) (BURGUEñO et al., 2012), introducing gaps in $\mathbf{\Phi}$. In total, there were 5,000 genotype-environment combinations (×3 replications per environment, totaling 15,000 data points). Thus, for example, after removing 20% in the first scenario, data from 4,000 combinations (12,000 data points) are retained. In the last scenario, there were only 1,000 combinations and 3,000 data points. We then fitted two compound symmetry models (similar to Eq. 2.4 but without the random block effect): one using the numerator relationship matrix (*model A*) and one without it (*model I*). In *model I*, missing values in $\mathbf{\Phi}$ were imputed using the EM-SVD algorithm (TROYANSKAYA et al., 2001). Finally, we assessed model performance by correlating predicted ge values with their true simulated counterparts. The entire process, from burn-in to three-way hybrid generation and model fitting, was repeated 50 times for robustness.

2.4 Results

2.4.1 Overview of per-environment analyses

The genetic effects were significant in all traits and environments (Figure 3A). The narrow-sense heritabilities and coefficients of experimental variation (CVs) varied between environments and traits. Overall, the CVs of plant (PH) and ear height (EH) were lower than those of grain yield (GY) (Figure 3B). The performances also varied according to the environment (Figure 3C). Candidates were more vigorous in high-nitrogen environments (coded as “COAN”) than in their low-nitrogen counterparts (coded as “COBN”). Remember that “COAN” and “COBN” trials were conducted at the same location. The highest variation in grain yield was observed in the second-season environment CRIS21, which coincidently was the environment that had the shortest plants, on average. The other second-season environment, LRV21, also had plants with low vigour (short plants with low grain yield), comparable to the low-nitrogen environments ("COBN"). The first-season environments HN21 and SMP22 had the tallest plants, but the grain yield did not surpass the high-nitrogen environments ("COAN").

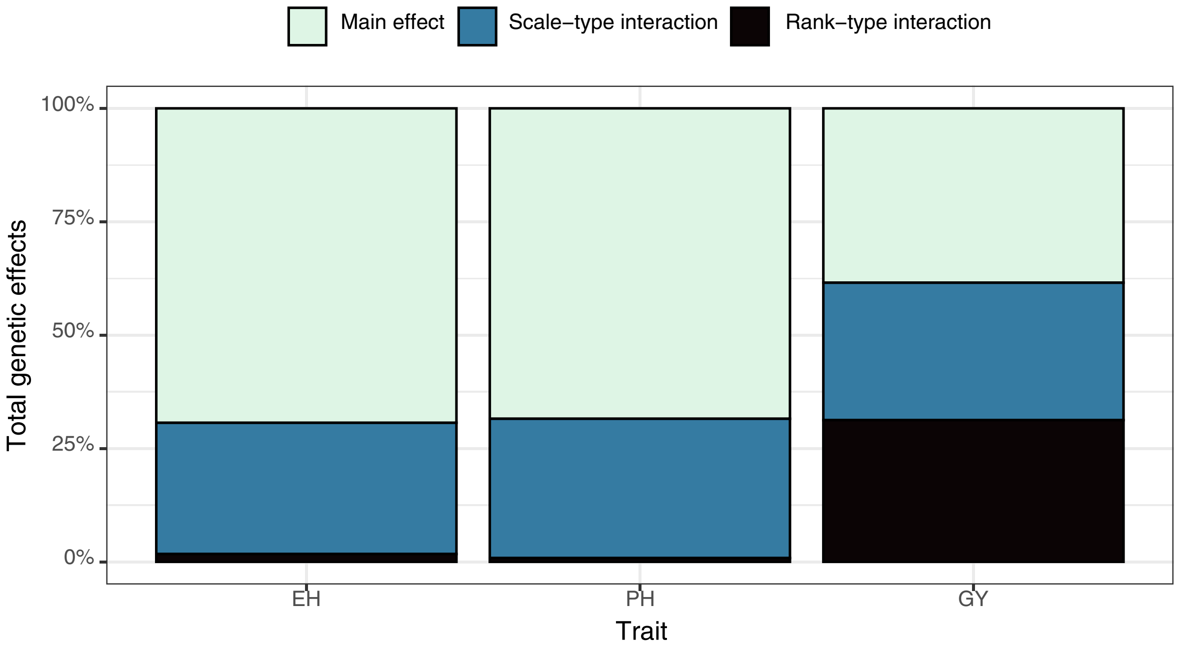
Figure 3 – Single environment analyses' results per trait: variance component estimates (A), coefficient of experimental variation and narrow-sense heritability (B), and distribution of genetic values (BLUPs, in C). The legend in A illustrates the results of the likelihood ratio tests (LRT), with "*" and "ns" representing significant and non-significant effects, respectively, considering an $\alpha = 0.05$ in a one-degree of freedom chi-square distribution. In A and B, the labels at the top distinguish the environment, and on the right, the traits (ear height, EH, in cm; plant height, PH, in cm; and grain yield, GY, in kg/ha). In C, the colour and shape of the points distinguish checks, hybrids (three-way) and parents (lines and single cross hybrids).



2.4.2 Multi-environment analyses and WAASB

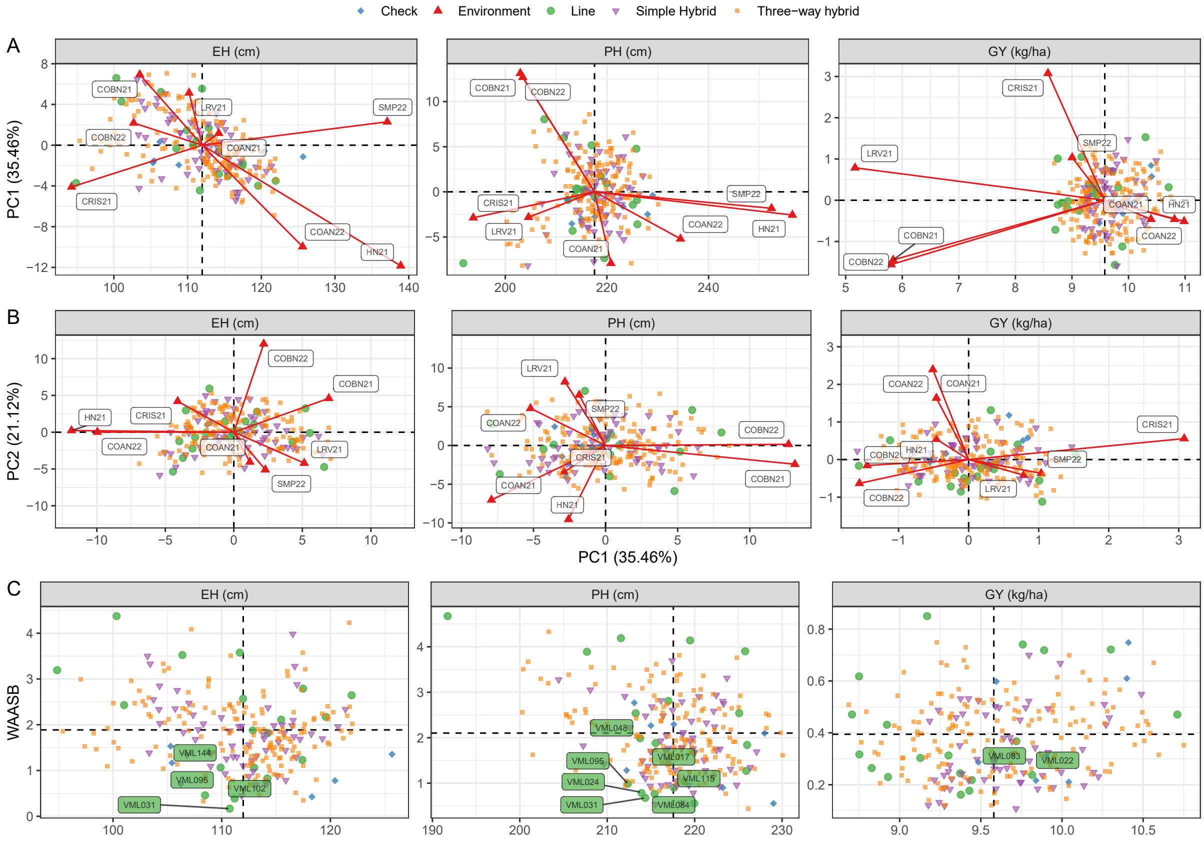
The different performances across environments observed in Figure 3C is a hint of the existence of genotype-by-environment interaction (GEI). The joint analysis confirms it (Figure 4). The GEI in PH and EH is mostly scale-type, meaning that genotypes tend to respond similarly to environmental inputs in these traits. On the other hand, rank-type GEI is slightly prevalent for GY, i.e., there is some ranking alteration between environments, meaning selection changes. It is important to stress that the main and GEI effects were significant for all traits, according to the LRT.

Figure 4 – Decomposition of the total genetic effects (in percentage, y-axis) into the main effect, rank-type genotype-by-environment interaction (GEI) effect, and scale-type GEI effect per trait (x-axis).



The presence of GEI in the dataset indicates that selection based solely on the mean may be insufficient. The AMMI results (Figures 5A and 5B) highlighted notable differences and similarities among environments, aligning with patterns previously observed in Figure 3C. For example, the "COAN" and "COBN" environments consistently formed distinct clusters across traits, while CRIS21 stood out as the most divergent environment for GY. Regarding parental selection, the plots revealed a limited number of high-performing, stable lines, with most genotypes exhibiting adaptation to specific environments. This is particularly evident in the WAASB plot (Figure 5C). For stability, lower WAASB values (i.e., farther down the *y-axis*) are desirable. For PH and EH, where reduced trait means are preferred, genotypes in the first quadrant are ideal. Conversely, for GY, higher means are preferred, making candidates in the fourth quadrant the focus of selection. Using the mean WAASB and mean trait values as truncation thresholds, only four, seven, and two lines were suitable for single-trait selection across PH, EH, and GY, respectively. Only two lines, VML095 and VML031, were selected for more than one trait, being identified as desirable for both PH and EH.

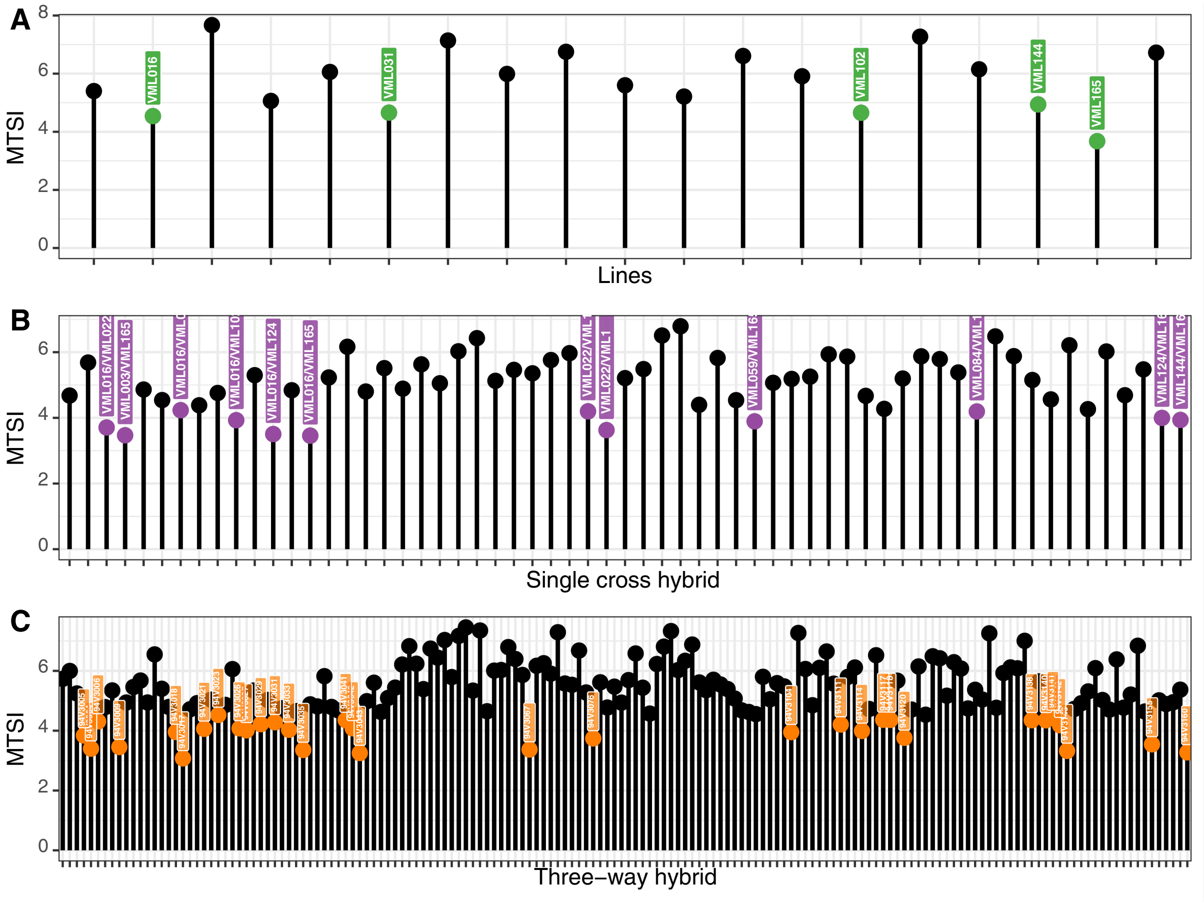
Figure 5 – Stability analyses per trait: Additive main effects and multiplicative interaction (AMMI) 1 (in A: first component-scores in the y-axis vs trait mean in the x-axis), AMMI2 (in B: first-component scores in the y-axis vs second-component scores in the x-axis) and weighted average of absolute scores of GEI's BLUPs (WAASB, in C: WAASB in the y-axis vs trait mean in the x-axis). Points have colours and shapes set according to the factor they refer to (check, line, single cross hybrid or three-way hybrid). The highlighted lines are the ones that would be selected considering the mean WAASB and mean trait values as truncation points.



2.4.3 Multi-trait multi-environment selection

The WAASB biplot usefulness is limited to the single-trait framework. As an alternative for multi-trait selection - considering both the performance and stability of all analysed traits - we used the multi-trait stability index (MTSI). We defined a selection intensity of 20%, translating into the selection of five parental lines (VML016, VML031, VML102, VML144, VML165), twelve single cross and 32 three-way hybrids. The selection of these lines represents an expected gain of $\approx 4\%$ for GY, $\approx 2\%$ for PH and $\approx -2\%$ for EH. We acknowledge that the selection of maize hybrids based only on additive effects may be suboptimal, but opted to show the results either way to highlight the potential of selecting both progenies and parents using a single trial network, considering the stability of both. Also, it is worth mentioning the correlation between genetic main effects of these traits: 0.31 between GY and EH, 0.67 between GY and PH, and 0.52 between PH and EH.

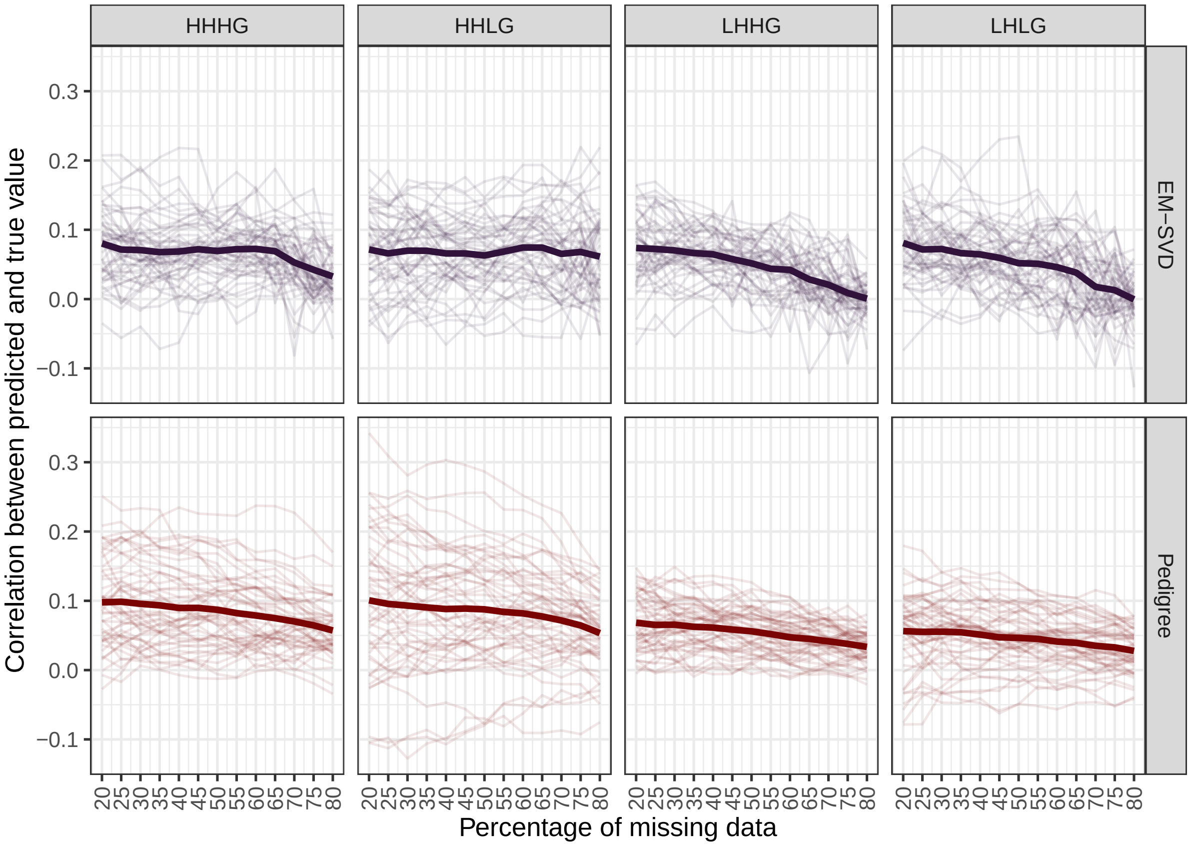
Figure 6 – Multi-trait stability index (MTSI) values (y-axis) of lines, single cross hybrids and three-way hybrids (x-axis). The lower the MTSI value, the closer the candidate is to the ideotype (desirable). The top 20% candidates are highlighted.



2.4.4 Simulation results

Leveraging relationship information to fill gaps in $\mathbf{\Phi}$ generally outperforms data imputation across all scenarios, though the difference in performance is not substantial (Figure 7). Overall, as the percentage of missing values increases, the correlation between predicted and true values declines, with the most pronounced effect observed in the low heritability, high GEI (LHHG) scenario. The variability in results across replications, represented by the translucent lines, highlights the impact of data set composition background (i.e., the simulated burn-in phase) and genotype-environment combinations that were retained after removing data to simulate imbalanced scenarios. For instance, in the high heritability, low GEI (HHLG) scenario, predictions using pedigree-based information can yield correlations as high as 0.35 or as low as -0.1, emphasizing the influence of missing data structure on model performance.

Figure 7 – Correlation between predicted and true (simulated) $g + ge$ values (y-axis) considering four scenarios varying heritability (h^2) and magnitude of the genotype-by-environment interaction (GEI, represented by $\bar{\rho}_{mm'}$, the base correlation between environments), different rates of missing data (x-axis, from 20% to 80% by 5%) and strategy to fill gaps in the GEI matrix (imputation or using relationship information via pedigree). The four scenarios were: HHHG - high heritability ($h^2 = 0.8$) and high genotype-by-environment interaction (GEI, $\bar{\rho}_{mm'} = 0.2$), HHLG - high heritability and low GEI ($\bar{\rho}_{mm'} = 0.8$), LHHG - low heritability ($h^2 = 0.2$) and high GEI, and LHLG - low heritability and low GEI. The translucent lines represent the progress in each one of the 50 replications, and the solid (opaque) line, the mean.



2.5 Discussion

This study tackles two problems using a single solution. By fitting a pedigree-based compound symmetry model, we extend the WAASB method to evaluate parental line stability based on progeny performance, eliminating the need for dedicated multi-environment trials (MET) to assess line performance. Additionally, through simulations, we demonstrate that models incorporating relationship information more effectively handle data imbalance— a common issue in plant breeding. Notably, while our approach relies on pedigree-based relationships, adapting it to genomic-based relationships is straightforward.

2.5.1 Stability of parental lines

The essence of plant breeding lies in the accumulation of favourable alleles, passed from one cycle to the next through selected lines that produce the next generation. In hybrid breeding programs, multi-environment trials (MET) primarily aim to assess hybrid performance stability across a sample of the target population of environments (TPE) (BERNARDO, 2009; PRASANNA et al., 2022; MASUKA et al., 2017). Line selection is largely driven by general combining ability, but since the correlation between performance and stability is far from perfect (see Figure 5C), stability-related alleles are often overlooked. As a result, lines that could contribute to hybrid stability may be discarded if stability is not explicitly considered in selection.

Classical stability estimation methods rely on observed data, from Finlay e Wilkinson (1963) regression to AMMI (JR., 1988) and GGE biplot (YAN et al., 2000b), methods that are still widely used. However, these approaches overlook parental stability unless specific field trials are conducted. New mixed model approaches allow an indirect assessment of parental stability through progeny performance leveraging relationship information (HENDERSON, 1975; SMITH; CULLIS; THOMPSON, 2005). One such method is WAASB, which builds on AMMI principles but differs fundamentally by using all principal components from the GEI matrix decomposition within a compound symmetry mixed model to build a single stability value, rather than just the first two or three components for biplot visualization (OLIVOTO et al., 2019a). This distinction is crucial, as graphical analysis becomes increasingly complex with more environments and genotypes (see Figures 5A and 5B). In our study, both AMMI and WAASB were derived from a compound symmetry model enriched with relationship information, enabling the estimation of parental genetic values. The conclusions drawn from both methods aligned: most lines exhibited specific adaptation, while crossing complementary lines produced broadly adapted single cross and three-way hybrids. This is particularly relevant for grain yield, where total genetic effects are largely controlled by rank-type GEI, as expected for a highly polymorphic trait.

For the proposed method to be applicable in a breeding pipeline, it must account for the multi-trait nature of routine selection. This represents a key challenge for plant breeders: selecting lines that carry favourable alleles for multiple traits (in our case, PH, EH, and GY) while ensuring these alleles contribute to stable performance across environments in the TPE. A practical approach to simplifying this selection is constructing a selection index based on the genotype-ideotype distance using the WAASBY of each trait. The WAASBY of a candidate is a weighted sum of its performance (represented by the BLUP from the compound symmetry model) and its stability (represented by its WAASB value) (OLIVOTO et al., 2019a). The complexity of multi-trait selection is evident in the expected genetic gains: while plant height would increase, ear height would decrease. This outcome is possible because the additive genetic correlation between these traits is moderate (0.52). Studies on the genetic architecture of PH and EH indicate that, despite their close relationship and shared QTLs, several specific QTLs influence each trait independently (CAI et al., 2012; SHU et al., 2023; PEIFFER et al., 2014).

2.5.2 Strategies for filling the GEI matrix

The imbalanced structure of plant breeding MET poses a significant challenge for methods that rely on the GEI matrix to compute stability metrics. For instance, to address missing data in AMMI models, Jr. e Zobel (1990) proposed an imputation method based on the expectation-maximization (EM) algorithm, which was later evaluated and validated under various missing data scenarios (PADEREWSKI; RODRIGUES, 2014). Several other imputation methods have since been developed and compared to existing algorithms (ANGELINI; CERVIGNI; QUAGLINO, 2024; ARCINIEGAS-ALARCóN et al., 2014; ARCINIEGAS-ALARCóN; GARCíA-PEñA; RODRIGUES, 2020), but none have explicitly considered the genetic relationships between selection candidates. It is worth noting that leveraging genetic covariances to address missing data is not a novel concept. For instance, Rincent et al. (2019) integrated genomic data with environmental features to impute missing values before fitting an AMMI model. However, the relative advantage of this approach over conventional imputation methods had not been thoroughly assessed. Our study fills this gap by directly comparing a flexible and widely used imputation method, EM-SVD, against a relationship-based approach using a combination of statistical and genetic simulations (GAYNOR; GORJANC; HICKEY, 2021; WERNER; GEMENET; TOLHURST, 2024; BANčIč; GORJANC; TOLHURST, 2024).

The superiority of relationship-based imputation is both intuitive and expected, as it allows the compound symmetry model to fully exploit the genetic covariances between genotypes—an inherent statistical and genetic property of the model (BERNARDO, 2020; MISZTAL; LEGARRA; AGUILAR, 2009; VIANA et al., 2010). However, a deeper analysis of the simulation results reveals two key points. First, the generally low correlation between predicted and true (simulated) values (mostly < 0.1) may stem from the data structure and model choice. As outlined in Materials and Methods, we based the simulation of GEI on an unstructured covariance matrix and accounted for heteroscedasticity in the residuals—both of which are common in plant breeding (BANčIč; GORJANC; TOLHURST, 2024; SMITH; CULLIS; THOMPSON, 2005). Given these complexities, the compound symmetry model may be overly simplistic, making it difficult to accurately capture the true GEI effect (CHAVES et al., 2023; ARAúJO et al., 2023). This limitation is particularly evident in the low heritability and high GEI scenario. Nevertheless, this does not undermine the validity of our findings. Compound symmetry models remain effective for estimating main effects and overall stability measures, making them a reliable and parsimonious tool for decision-making across the TPE, mainly when GEI is mostly of scale-type, like for EH and PH.

The second key point is the reliance of both statistical (EM-SVD) and statistical-genetic (genetic covariance) imputation methods on the available data: as the percentage of missing values increases, the overall quality of imputation declines. However, EM-SVD-based imputation exhibits more frequent irregular patterns—fluctuations in accuracy as missing data increases—compared to its statistical-genetic counterpart. This is likely due to the influence of genetic covariance in the latter, which prevents imputed values from deviating excessively from biologically plausible expectations. Finally, it is important to note that our approach is based on an expected identity-by-descent framework. We anticipate that imputation accuracy would further improve if identity-by-state information were incorporated using a genomic relationship matrix, providing even more precise estimates for missing values VanRaden (2008), Vitezica, Varona e Legarra (2013).

2.5.3 Concluding remarks

The analytical pipeline outlined in this study enabled the selection of parental lines, F_1 hybrids, and three-way hybrids based on both performance and stability across multiple traits. The effectiveness of our proposed approach was further validated through simulations, reinforcing the reliability of our findings. Future studies should investigate how relationship-based imputation performs under different missing data patterns (e.g., missing-at-random, not-missing-at-random) and compare it with alternative imputation strategies. The ultimate goal is to equip plant breeders with efficient yet practical tools that can be seamlessly integrated into routine decision-making in breeding programs.

References

AASTVEIT, A. H.; AASTVEIT, K. Effects of genotype-environment interactions on genetic correlations. *Theoretical and Applied Genetics*, v. 86, n. 8, p. 1007–1013, set. 1993. ISSN 1432-2242. 23, 30

ADRIAANSE, F. G.; HUMAN, J. J. Effect of time of application and nitrate: ammonium ratio on maize grain yield, grain n concentration and soil mineral n concentration in a semi-arid region. *Field Crops Research*, v. 34, n. 1, p. 57–70, 1993. ISSN 0378-4290. 20

ALAM, M. M. et al. Effect of drought stress on grain yield, water use efficiency and physiological characteristics in maize genotypes. *Scientific Reports*, v. 7, p. 13284, 2017. 20

AMADEU, R. R. et al. AGHmatrix: genetic relationship matrices in R. *Bioinformatics*, v. 39, n. 7, p. btad445, jul. 2023. 26

ANGELINI, J.; CERVIGNI, G. D. L.; QUAGLINO, M. B. New imputation methodologies for genotype-by-environment data: an extensive study of properties of estimators. *Euphytica*, v. 220, n. 6, p. 92, maio 2024. 23, 38

ANNICCHIARICO, P. Cultivar adaptation and recommendation from alfalfa trials in northern italy. *Journal of Genetics and Plant Breeding*, v. 46, p. 269–278, 1992. 14

ARAúJO, M. d. S. et al. Modeling covariance structures for genetic and non-genetic effects in cowpea multi-environment trials. *Agronomy Journal*, v. 115, n. 3, p. 1248–1256, 2023. ISSN 1435-0645. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/agj2.21321. 38

ARAúJO, M. S. et al. GIS-FA: an approach to integrating thematic maps, factor-analytic, and envirotyping for cultivar targeting. *Theoretical and Applied Genetics*, v. 137, n. 4, p. 80, March 2024. 15

ARCINIEGAS-ALARCóN, S. et al. An alternative methodology for imputing missing data in trials with genotype-by-environment interaction: some new aspects. *Biometrical Letters*, v. 51, n. 2, p. 75–88, dez. 2014. ISSN 1896-3811. 38

ARCINIEGAS-ALARCóN, S.; GARCíA-PEñA, M.; RODRIGUES, P. C. New multiple imputation methods for genotype-by-environment data that combine singular value decomposition and Jackknife resampling or weighting schemes. *Computers and Electronics in Agriculture*, v. 176, p. 105617, set. 2020. ISSN 0168-1699. 38

BADU-APRAKU, B. et al. Relationships among traits of tropical early maize cultivars in contrasting environments. *Agronomy Journal*, v. 103, n. 3, p. 717–729, 2011. 19, 20

BALBINO, L. C. et al. *Integração lavoura-pecuária-floresta (ILPF): uma inovação tecnológica no Brasil*. Brasília: Embrapa, 2012. 133 p. 17

BALIGAR, V. C.; FAGERIA, N. K.; HE, Z. L. Nutrient use efficiency in plants. *Communications in Soil Science and Plant Analysis*, Taylor & Francis, v. 32, n. 7-8, p. 921–950, 2001. 21

BANčIč, J.; GORJANC, G.; TOLHURST, D. J. A framework for simulating genotype-by-environment interaction using multiplicative models. *Theoretical and Applied Genetics*, v. 137, n. 8, p. 197, ago. 2024. ISSN 1432-2242. 9, 30, 31, 38

BANčIč, J. et al. Plant breeding simulations with AlphaSimR. *Crop Science*, v. 65, n. 1, p. e21312, 2025. ISSN 1435-0653. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/csc2.21312. 29, 30, 31

BATES, D. et al. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, v. 67, n. 1, p. 1–48, 2015. 29

BEHERA, P. P. et al. WAASB-based stability analysis and simultaneous selection of forage sorghum for forage yield traits. *Euphytica*, v. 220, n. 9, p. 145, ago. 2024. ISSN 1573-5060. 23

BENEšOVá, M. et al. The physiology and proteomics of drought tolerance in maize: early stomatal closure as a cause of lower tolerance to short-term dehydration? *PLoS One*, v. 7, n. 6, p. e38017, 2012. 20

BERGAMASCHI, H. et al. Distribuição hídrica no período crítico do milho e produção de grãos. *Pesquisa Agropecuária Brasileira*, v. 39, n. 9, p. 831–839, September 2004. 20

BERNARDO, R. Genomewide selection for rapid introgression of exotic germplasm in maize. *Crop Science*, v. 49, n. 2, p. 419–425, 2009. ISSN 1435-0653. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2135/cropsci2008.08.0452. 9, 30, 37

BERNARDO, R. Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. *Heredity*, v. 125, n. 6, p. 375–385, dez. 2020. ISSN 0018-067X, 1365-2540. 38

BERNARDO, R. Predictive breeding in maize during the last 90 years. *Crop Science*, v. 61, n. 5, p. 2872–2881, 2021. 18

BORéM, A.; MIRANDA, G. V.; FRITSCHE-NETO, R. *Melhoramento de Plantas*. 7. ed. Viçosa, MG: Editora UFV, 2017. 543 p. ISBN 9788572695671. 18

BURGUEñO, J. et al. Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. *Crop Science*, v. 52, n. 2, p. 707–719, 2012. ISSN 1435-0653. 31

CAI, H. et al. Identification of QTLs for plant height, ear height and grain yield in maize (*Zea mays* L.) in response to nitrogen and phosphorus supply. *Plant Breeding*, v. 131, n. 4, p. 502–510, ago. 2012. ISSN 0179-9541, 1439-0523. 38

CASAGRANDE, C. R. et al. Envirotypes applied to evaluate the adaptability and stability of wheat genotypes in the tropical region in Brazil. *Euphytica*, v. 220, n. 2, p. 27, fev. 2024. ISSN 0014-2336, 1573-5060. 23

CHAVES, S. F. S. et al. Analysis of repeated measures data through mixed models: An application in *Theobroma grandiflorum* breeding. *Crop Science*, v. 63, n. 4, p. 2131–2144, 2023. ISSN 0011-183X, 1435-0653. 14, 38

CHAVES, S. F. S. et al. Probbreed: a novel tool for calculating the risk of cultivar recommendation in multienvironment trials. *G3 Genes|Genomes|Genetics*, v. 14, n. 3, p. jkae013, 01 2024. ISSN 2160-1836. 14

CHEN, G. K.; MARJORAM, P.; WALL, J. D. Fast and flexible simulation of DNA sequence data. *Genome Research*, v. 19, n. 1, p. 136–142, jan. 2009. ISSN 1088-9051, 1549-5469. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab. 29

COELHO, A. M.; CRUZ, J. C.; FILHO, I. A. P. *Desafios para a obtenção de altas produtividades de milho*. Sete Lagoas, MG, Brasil, 2004. 20 p. (Comunicado Técnico, 99). 17

CONAB. *Acompanhamento de Safra Brasileira: Grãos, Safra 2022/23*. 2023. Acesso em: 21 mar. 2025. Disponível em: <https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos/item/21369-12-levantamento-safra-2022-23>. 17

COOPER, M.; DELACY, I. H. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *Theoretical and Applied Genetics*, v. 88, n. 5, p. 561–572, jul. 1994. ISSN 1432-2242. 23, 27

COVARRUBIAS-PAZARAN, G. Genome-assisted prediction of quantitative traits using the R Package sommer. *PLOS ONE*, v. 11, n. 6, p. e0156744, jun. 2016. ISSN 1932-6203. 29

CROSSA, J. Statistical Analyses of Multilocation Trials. In: *Advances in Agronomy*. [S.l.]: Elsevier, 1990. v. 44, p. 55–85. ISBN 978-0-12-000744-8. 23

CROSSA, J. From Genotype × Environment Interaction to Gene × Environment Interaction. *Current Genomics*, v. 13, n. 3, p. 225–244, maio 2012. ISSN 1389-2029. 22

CROSSA, J. et al. Modeling genotype × environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. *Crop Science*, v. 46, n. 4, p. 1722–1733, 2006. 16

CROSSA, J.; CORNELIUS, P. L. Sites regression and shifted multiplicative model clustering of cultivar trial sites under heterogeneity of error variances. *Crop Science*, v. 37, n. 2, p. 406–415, 1997. 14

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. *Modelos biométricos aplicados ao melhoramento genético*. 4. ed. Viçosa, MG: Editora UFV, 2012. v. 1. 544 p. 12, 14

CULLIS, B. R.; SMITH, A. B.; COOMBES, N. E. On the design of early generation variety trials with correlated data. *Journal of Agricultural, Biological, and Environmental Statistics*, v. 11, n. 4, p. 381–393, dez. 2006. ISSN 1085-7117, 1537-2693. 26

DHAKAL, K. et al. Three-way cross white kernel hybrid maize out-yielded commercial variety tested under two contrasting environments. *Journal of Agriculture and Food Research*, v. 7, p. 100279, 2022. ISSN 2666-1543. 21

DIAS, K. O. G. et al. Novel strategies for genomic prediction of untested single-cross maize hybrids using unbalanced historical data. *Theoretical and Applied Genetics*, v. 133, n. 2, p. 443–455, fev. 2020. 24

DIAS, K. O. G. et al. Leveraging probability concepts for cultivar recommendation in multi-environment trials. *Theoretical and Applied Genetics*, v. 135, n. 4, p. 1385–1399, 2022. ISSN 1432-2242. 14

DURãES, F. O. M. Limitações fisiológicas do milho nas condições de plantio nas regiões tropicais baixas. In: EMBRAPA MILHO E SORGO. *Anais do evento da Embrapa Milho e Sorgo*. [S.l.], 2006. 20

EBERHART, S. A.; RUSSELL, W. A. Stability parameters for comparing varieties. *Crop Science*, v. 6, n. 1, p. cropsci1966.0011183X000600010011x, 1966. 14

ELIAS, A. A. et al. Half a century of studying genotype × environment interactions in plant breeding experiments. *Crop Science*, v. 56, n. 5, p. 2090–2105, 2016. 14

FAO. *Climate Smart Agriculture: Sourcebook*. Rome: FAO, 2018. Disponível em: <https://www.fao.org/climate-smart-agriculture-sourcebook/en/>. 20

FARIA, S. V. et al. Phenotypic and molecular characterization of a set of tropical maize inbred lines from a public breeding program in Brazil. *BMC Genomics*, v. 23, n. 1, p. 54, jan. 2022. ISSN 1471-2164. Number: 1. 24

FERNANDES, F. C. S. et al. Doses, eficiência e uso de nitrogênio por seis cultivares de milho. *Revista Brasileira de Milho e Sorgo*, Sete Lagoas, v. 4, n. 2, p. 195–204, 2005. 20

FILHO, A. B. et al. *Uma história brasileira do milho: o valor dos recursos genéticos*. Brasília: Paralelo 15, 2000. 523 p. 18

FINLAY, K. W.; WILKINSON, G. N. The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research*, v. 14, n. 6, p. 742, 1963. ISSN 0004-9409. 14, 37

GAUCH, H. G.; ZOBEL, R. W. Predictive and postdictive success of statistical analyses of yield trials. *Theoretical and Applied Genetics*, v. 76, p. 1–10, 1988. 12, 14

GAYNOR, R. C.; GORJANC, G.; HICKEY, J. M. AlphaSimR: an R package for breeding program simulations. *G3 Genes|Genomes|Genetics*, v. 11, n. 2, p. jkaa017, fev. 2021. ISSN 2160-1836. 29, 38

HAEGELE, J. W. et al. Changes in nitrogen use traits associated with genetic improvement for grain yield of maize hybrids released in different decades. *Crop Science*, v. 53, n. 4, p. 1256–1268, 2013. 21

HENDERSON, C. R. Best Linear Unbiased Estimation and Prediction under a selection model. *Biometrics*, v. 31, n. 2, p. 423, jun. 1975. ISSN 0006341X. 12, 16, 26, 37

HICKEY, J. M. et al. Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Science*, v. 54, n. 4, p. 1476–1488, 2014. ISSN 1435-0653. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2135/cropsci2013.03.0195. 30

IZGE, A. U.; DUGJE, I. Y. Performance of drought tolerant three-way and top cross maize hybrids in sudan savanna of north eastern nigeria. *Journal of Plant Breeding and Crop Science*, v. 3, n. 11, p. 269–675, 2011. 21

JONG, S. K.; BREWBAKER, J. L.; LEE, C. H. Effects of solar radiation on the performance of maize in 41 successive monthly plantings in hawaii. *Crop Science*, v. 22, n. 1, p. 13–18, 1982. 19, 20

JR., H. G. G. Model selection and validation for yield trials with interaction. *Biometrics*, v. 44, n. 3, p. 705–715, 1988. ISSN 0006-341X. Publisher: [Wiley, International Biometric Society]. 23, 37

JR., H. G. G.; ZOBEL, R. W. Imputing missing yield trial data. *Theoretical and Applied Genetics*, v. 79, n. 6, p. 753–761, jun. 1990. ISSN 0040-5752, 1432-2242. 23, 38

KISTLER, L. et al. Multiproxy evidence highlights a complex evolutionary legacy of maize in south america. *Science*, v. 362, n. 6420, p. 1309–1313, 2018. 16

LIN, C. S.; BINNS, M. R. Concepts and methods for analyzing regional trial data for cultivar and location selection. In: JANICK, J. (Ed.). *Plant Breeding Reviews*. 1. ed. [S.l.]: Wiley, 1994. p. 271–297. ISBN 978-0-471-57344-9 978-0-470-65049-3. 23

LIN, C. S.; BINNS, M. R.; LEFKOVITCH, L. P. Stability analysis: Where do we stand? *Crop Science*, v. 26, n. 5, p. cropsci1986.0011183X002600050012x, 1986. 14

MAFOUASSON, H. N. A. et al. Genotype-by-environment interaction and yield stability of maize single cross hybrids developed from tropical inbred lines. *Agronomy*, v. 8, n. 5, p. 62, 2018. ISSN 2073-4395. 20

MALDANER, L. J. et al. Exigências agroclimáticas da cultura do milho (zea mays). *Revista Brasileira de Energias Renováveis*, v. 3, p. 13–23, 2014. 17

MALOSETTI, M.; RIBAUT, J.-M.; EEUWIJK, F. A. van. The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. *Frontiers in Physiology*, v. 4, p. 44, 2013. ISSN 1664-042X. 14, 23

MASUKA, B. et al. Gains in maize genetic improvement in Eastern and Southern Africa: I. CIMMYT hybrid breeding pipeline. *Crop Science*, v. 57, n. 1, p. 168–179, 2017. ISSN 1435-0653. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2135/cropsci2016.05.0343. 37

MATSUOKA, Y. et al. A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences of the United States of America*, v. 99, n. 9, p. 6080–6084, 2002. 16

MISZTAL, I.; LEGARRA, A.; AGUILAR, I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science*, v. 92, n. 9, p. 4648–4655, set. 2009. ISSN 0022-0302. 23, 38

MONTEVERDE, E. et al. Multienvironment models increase prediction accuracy of complex traits in advanced breeding lines of rice. *Crop Science*, v. 58, n. 4, p. 1519–1530, 2018. 12, 14

MôRO, G. V.; FRITSCHE-NETO, R. Importância e usos do milho no brasil. In: *Milho: do plantio à colheita*. Viçosa, MG, Brasil: Editora da UFV, 2017. p. 382. 13, 16, 19

NATARAJ, V. et al. WAASB-based stability analysis and simultaneous selection for grain yield and early maturity in soybean. *Agronomy Journal*, v. 113, n. 4, p. 3089–3099, 2021. ISSN 1435-0645. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/agj2.20750. 23

OLIVOTO, T.; LúCIO, A. D. metan: An R package for multi-environment trial analysis. *Methods in Ecology and Evolution*, v. 11, n. 6, p. 783–789, 2020. ISSN 2041-210X. 23, 29

OLIVOTO, T. et al. Mean performance and stability in multi-environment trials I: combining features of AMMI and BLUP techniques. *Agronomy Journal*, v. 111, n. 6, p. 2949–2960, 2019. 13, 14, 15, 23, 28, 37

OLIVOTO, T. et al. Mean performance and stability in multi-environment trials II: selection based on multiple traits. *Agronomy Journal*, v. 111, n. 6, p. 2961–2969, 2019. ISSN 1435-0645. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2134/agronj2019.03.0221. 23, 28

PADEREWSKI, J.; RODRIGUES, P. C. The usefulness of EM-AMMI to study the influence of missing data pattern and application to Polish post-registration winter wheat data. *Australian Journal of Crop Science*, v. 8, n. 4, p. 640–645, 2014. 23, 38

PAES, M. C. D. *Aspectos físicos, químicos e tecnológicos do grão de milho*. Sete Lagoas, MG, Brasil, 2006. 6 p. 13, 16

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. *Biometrika*, v. 58, n. 3, p. 545–554, 1971. ISSN 0006-3444, 1464-3510. 12, 16, 26

PEIFFER, J. A. et al. The Genetic Architecture Of Maize Height. *Genetics*, v. 196, n. 4, p. 1337–1356, abr. 2014. ISSN 0016-6731. 38

PIEPHO, H.-P. et al. BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica*, v. 161, n. 1-2, p. 209–228, maio 2008. ISSN 0014-2336, 1573-5060. 12, 15, 23

PLAISTED, R. L.; PETERSON, L. C. A technique for evaluating the ability of selections to yield consistently in different locations or seasons. *American Potato Journal*, v. 36, n. 11, p. 381–385, 1959. ISSN 1874-9380. 14

POUPON, V. et al. Genotype x environment interaction and climate sensitivity in growth and wood density of European larch. *Forest Ecology and Management*, v. 545, p. 121259, out. 2023. ISSN 03781127. 23, 30

PRASANNA, B. M. et al. Genetic trends in CIMMYT's tropical maize breeding pipelines. *Scientific Reports*, v. 12, n. 1, p. 20110, nov. 2022. ISSN 2045-2322. 37

R Core Team. *R: A Language and environment for statistical computing*. Viena, Áustria: R Foundation for Statistical Computing, 2024. Programmers: _:n2360. Disponível em: <https://www.R-project.org/>. 26

REZENDE, W. S. et al. Performance and yield stability of maize hybrids in stress-prone environments in eastern africa. *The Crop Journal*, v. 8, n. 1, p. 107–118, 2020. ISSN 2214-5141. 19, 20

RIBAUT, J. M. et al. Drought tolerance in maize. In: BENNETZEN, J.; HAKE, S. (Ed.). *Handbook of Maize: Its Biology*. New York: Springer, 2009. p. 311–344. 20

RINCENT, R. et al. Using crop growth model stress covariates and AMMI decomposition to better predict genotype-by-environment interactions. *Theoretical and Applied Genetics*, v. 132, n. 12, p. 3399–3411, dez. 2019. ISSN 1432-2242. 24, 38

RODRIGUES, M. C. et al. Genotypic variation and relationships among nitrogen-use efficiency and agronomic traits in tropical maize inbred lines. *Genetics and Molecular Research*, v. 16, n. 3, 2017. ISSN 1676-5680. 21

SCAPIM, C. A. *Melhoramento de Plantas: Conceitos e Técnicas*. 2. ed. Lavras, MG: Editora UFLA, 2009. 523 p. 18

SHU, G. et al. Identification of QTNs, QTN-by-environment interactions for plant height and ear height in maize multi-environment GWAS. *Frontiers in Plant Science*, v. 14, nov. 2023. ISSN 1664-462X. Publisher: Frontiers. 38

SHUKLA, G. K. Some statistical aspects of partitioning genotype-environmental components of variability. *Heredity*, v. 29, n. 2, p. 237–245, 1972. ISSN 1365-2540. 14

SHULL, G. H. The composition of a field of maize. *American Breeders Association Report*, v. 4, p. 296–301, 1908. 18

SILVA, G. O. et al. *Histórico da cultura do milho no Brasil*. [S.l.]: Embrapa Milho e Sorgo, 2015. 17, 18

SMITH, A. B.; CULLIS, B. R. Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. *Euphytica*, v. 214, n. 8, p. 143, July 2018. ISSN 1573-5060. 14

SMITH, A. B.; CULLIS, B. R.; THOMPSON, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *The Journal of Agricultural Science*, v. 143, n. 6, p. 449–462, dez. 2005. ISSN 1469-5146, 0021-8596. 16, 22, 23, 37, 38

SOUZA, J. C. Cultivares. In: LIMA, R. O.; BORéM, A. (Ed.). *Melhoramento de Milho*. Viçosa, MG: Editora UFV, 2018. p. 295–306. 13, 19

SSERUMAGA, J. P. et al. Grain-yield stability among tropical maize hybrids derived from doubled-haploid inbred lines under random drought stress and optimum moisture conditions. *Crop and Pasture Science*, v. 69, n. 11, p. 691–712, 2018. 19, 20

SU, G. et al. Comparison of genomic predictions using genomic relationship matrices built with different weighting factors to account for locus-specific variances. *Journal of Dairy Science*, v. 97, n. 10, p. 6547–6559, 2014. ISSN 0022-0302. 16

SUNOJ, V. S. J. et al. Diurnal temperature amplitude alters physiological and growth response of maize (zea mays l.) during the vegetative stage. *Environmental and Experimental Botany*, v. 130, p. 113–121, 2016. ISSN 0098-8472. 20

The VSNi Team. *asreml: Fits Linear Mixed Models using REML*. 2023. Disponível em: <www.vsni.co.uk>. 26

TROYANSKAYA, O. et al. Missing value estimation methods for DNA microarrays. *Bioinformatics*, v. 17, n. 6, p. 520–525, jun. 2001. ISSN 1367-4803. 23, 31

USDA. *World Agricultural Production: Global Market Analysis.* 2023. Accessed: 2024-09-12. Disponível em: <https://www.usda.gov>. 17

VANRADEN, P. Efficient methods to compute genomic predictions. *Journal of Dairy Science*, v. 91, n. 11, p. 4414–4423, nov. 2008. ISSN 00220302. 16, 39

VELAZCO, J. G. et al. Combining pedigree and genomic information to improve prediction quality: an example in sorghum. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, v. 132, p. 2055 – 2067, 2019. 12, 16

VIANA, J. M. S. et al. BLUP for genetic evaluation of plants in non-inbred families of annual crops. *Euphytica*, v. 174, n. 1, p. 31–39, jul. 2010. ISSN 1573-5060. 38

VITEZICA, Z. G.; VARONA, L.; LEGARRA, A. On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics*, v. 195, n. 4, p. 1223–1230, dez. 2013. 39

WANG, G. et al. Response of gene expression and alternative splicing to distinct growth environments in tomato. *International Journal of Molecular Sciences*, v. 18, 2017. 14

WANG, Y. et al. Reduction in seed set upon exposure to high night temperature during flowering in maize. *Physiologia Plantarum*, v. 169, n. 1, p. 73–82, 2020. 20

WARTHA, C. A.; LORENZ, A. Implementation of genomic selection in public-sector plant breeding programs: Current status and opportunities. *Crop Breeding and Applied Biotechnology*, 2021. 16

WERNER, C. R.; GEMENET, D. C.; TOLHURST, D. J. FieldSimR: an R package for simulating plot data in multi-environment field trials. *Frontiers in Plant Science*, v. 15, p. 1330574, abr. 2024. Publisher: Frontiers. 29, 31, 38

WICKHAM, H. *ggplot2: Elegant graphics for data analysis.* 2. ed. Cham: Springer, 2016. 26

WRICKE, G. Zur berechnung der ökovalenzb̈ei sommerweizen und hafer. *Pflanzenzüchtung*, Berlin, v. 52, p. 127–138, 1965. 14

YAN, W. et al. Cultivar evaluation and mega-environment investigation based on the gge biplot. *Crop Science*, v. 40, n. 3, p. 597–605, 2000. 14

YAN, W. et al. Cultivar evaluation and mega-environment investigation based on the GGE Biplot. *Crop Science*, v. 40, n. 3, p. 597–605, 2000. ISSN 1435-0653. 37

YASSITEPE, J. E. d. C. T. et al. Maize transformation: From plant material to the release of genetically modified and edited varieties. *Frontiers in Plant Science*, v. 12, p. 766702, 2021. Disponível em: <https://www.frontiersin.org/articles/10.3389/fpls.2021.766702/full>. 17

YATES, F.; COCHRAN, W. G. The analysis of groups of experiments. *The Journal of Agricultural Science*, v. 28, n. 4, p. 556–580, 1938. 14

ZHANG, Z. et al. The impact of climate change on maize production: Empirical findings and implications for sustainable agricultural development. *Frontiers in Environmental Science*, v. 10, p. 954940, 2022. 17