Australian Journal of Basic and Applied Sciences

2025; 19(5): pages 1-14
https://doi.org/10.22587/ajbas.2025.19.5.1

ORIGINAL ARTICLE AENSI Publications

Enviromic prediction of sugarcane yield using nonlinear Gaussian kernels and climatic data

Tays Silva Batista[1], Wagner Barbosa[2], Germano Costa-Neto[3], Felipe Lopes da Silva[4], Bruno Portela Brasileiro[5] and Luiz Alexandre Peternelli[1*]

[1]Departamento de Estatística, Universidade Federal de Viçosa, Viçosa, MG, 36570–900, Brazil
[2]EPAMIG Sudeste – Vila Gianetti, 46/47 – Universidade Federal de Viçosa Viçosa, MG, 36570-075, Brazil
[3]Syngenta Seeds, Ithaca, NY, United States of America, 14851, United States of America
[4]Departamento de Agronomia, Universidade Federal de Viçosa, Viçosa, MG, 36570–900, Brazil
[5]Universidade Federal do Paraná, Curitiba, PR, 81530-000, Brazil

Correspondence Author: Luiz Alexandre Peternelli, Departamento de Estatística, Universidade Federal de Viçosa, Viçosa, MG, 36570–900, Brazil. Phone: +55(31) 3612-6161; Email: peternelli@ufv.br

Received date: 20 April 2025, Accepted date: 20 September 2025

ABSTRACT: The integration of environmental covariates into prediction models has demonstrated strong potential to increase the accuracy of genotype selection in plant breeding. When the number of phenotypic observations is limited, data simulation can be an effective strategy to expand training datasets while preserving model robustness. In this study, we tested the hypotheses that (i) incorporating environmental covariates improves the predictive performance of sugarcane yield models and (ii) increasing the number of simulated genotypes reduces prediction errors. Sugarcane yield, measured in tons of stalks per hectare (TSH), was predicted using Bayesian mixed models that incorporated Gaussian environmental kernels derived from NASA Power remote sensing data. Due to the limited availability of empirical data (11 genotypes evaluated in 8 environments), synthetic genotypes were generated using a segmented regression approach based on genotype-specific fitness and stability parameters. These genotypes were generated based on real data and validated against a subset of observed genotypes previously excluded from model training. Three validation scenarios (3G, 5G, and 7G) were implemented, with an increasing number of genotypes used in the training phase (183, 185, and 187, respectively). Three model structures were evaluated, differing in the inclusion of genotypic, environmental, and GxE interactions. The results confirmed that both environmental covariates and genotype simulation contributed significantly to the reduction of root mean square error (RMSE), with reductions of up to 82.5%. These findings highlight the relevance of enviromics and data augmentation strategies in early-stage plant breeding trials

Keywords: Genotype-by-environment interaction; Bayesian mixed models; Remote sensing; Sythetic phenotypic data.

INTRODUCTION

In recent years, sugarcane production in Brazil has faced considerable climatic challenges, including frosts, droughts, and reduced rainfall in the Center-South region, which negatively impacted productivity during the 2021 and 2022 harvests (CONAB, 2022/23). However, the third estimate for the 2023/24 harvest reported a significant recovery, driven by favorable climatic conditions and increased investment in the sector, consolidating Brazil's position as the world's largest sugarcane producer. Sugarcane is a perennial crop grown mainly in tropical and subtropical regions (Rodrigues, 1995), with a commercial production cycle lasting approximately five years. Throughout its development, sugarcane growth is influenced by edaphoclimatic conditions and agronomic management practices (Manhães et al., 2015). Understanding the environmental dynamics of agricultural lands is essential to optimize crop performance, variety recommendations, and support decision-making in breeding programs (Scarpari et al., 2004; Caetano et al., 2017).

In plant breeding programs, multi-environment trials (METs) are essential for evaluating genotypes across different locations and growing seasons. These trials often reveal strong genotype-environment (G×E) interactions, which can limit selection gains (Tena et al., 2019). To better capture these patterns, recent advances have highlighted the importance of quantitatively characterizing environments through environmental descriptors (Des Marais et al., 2013). Envirotyping is an approach that involves collecting, processing, and modeling environmental data to provide a broader understanding of how environmental variation influences plant phenotypic responses (Xu et al., 2016; Van Eeuwijk et al., 2018). Its large-scale application, known as enviromics, integrates remote sensing, bioinformatics, and predictive modeling to investigate genotype-environment (G×E) interactions in greater depth (Crossa et al., 2022).

Although the use of enviromics in sugarcane is still limited, some studies have begun to explore its application in this crop (Resende et al., 2024; Montes et al., 2021). However, most of these efforts have relied on a small number of environmental variables, often restricted to localized, small-scale analyses (Ramburan et al., 2011; Ramburan et al., 2012; Cursi et al., 2021). In contrast, enviromics has already shown promising results in other crops, such as corn, wheat, rice, and common beans (Jarquín et al., 2014; Millet et al., 2019; Heinemann et al., 2022; Heinemann et al., 2024). Bayesian mixed models offer a powerful framework to incorporate multiple sources of variation, including G×E effects and environmental covariates, into genomic and phenotypic predictions (De Los Campos et al., 2009; Montesinos-López et al., 2018). However, one of the main challenges of these models is the limited number of genotypes available for training. To address this limitation, data augmentation through genotype simulation has been proposed as a strategy to increase training diversity and improve model generalization (Silva et al., 2015; Muetanene et al., 2023).

In this study, we tested the hypothesis that (i) the inclusion of environmental covariates improves the predictive performance of sugarcane yield models and (ii) increasing the number of genotypes through simulation reduces prediction errors. To this end, we applied Bayesian mixed models incorporating Gaussian environmental kernels derived from remote sensing data to predict sugarcane yield (TSH – tons of stalks per hectare). Our objective was to evaluate the contribution of enviromics and data augmentation strategies to improve predictive accuracy in early-stage sugarcane breeding trials.

2. MATERIAL AND METHODS

2.1 Description of the study area

The experimental data were obtained from the experimental phase of the sugarcane breeding program at the Federal University of Viçosa (PMGCA-UFV). Eleven clones were evaluated across eight locations (Fig. 1). In each location, the experiments were arranged in a randomized block design with three replications. The experimental plots consisted of four rows, each five meters long, with a spacing of 1.4 meters between rows.

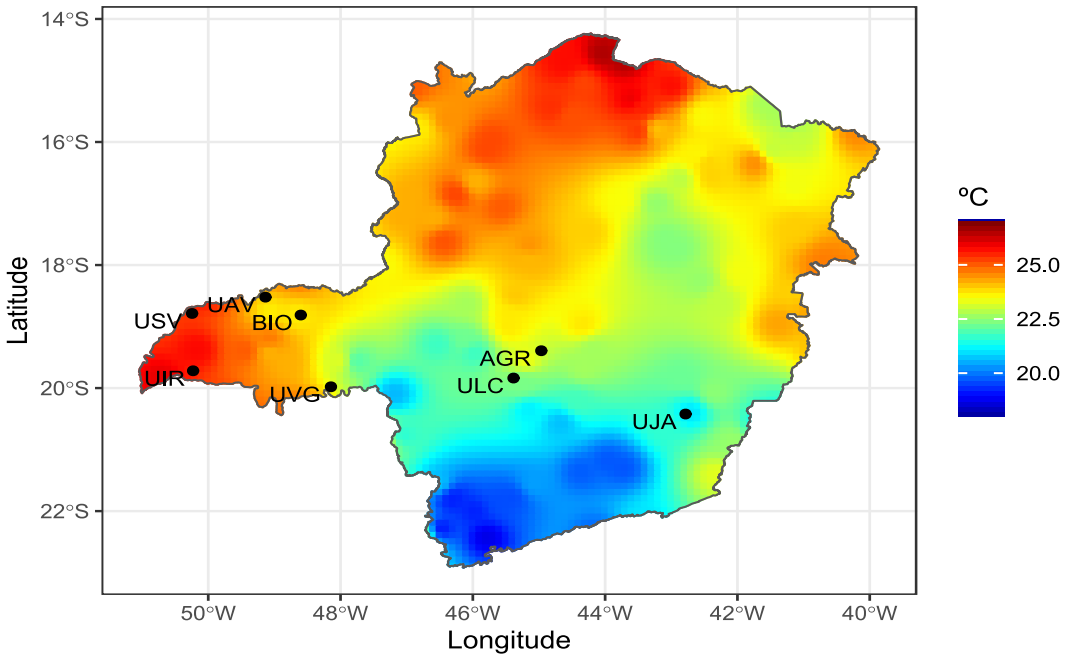

Figure 1: Average temperatures in Minas Gerais state. The points on the map indicate the geographic location of the experiments conducted at the following mills (name; municipality): UAV (Usina Alvorada; municipality), UIR (Usina Coruripe; municipality), AGR (Usina Agropéu; municipality), BIO (Usina Bioaroeira; municipality), USV (Usina Santa Vitória; municipality), ULC (Usina Santa Lúcia; municipality), UJA (Usina Jatiboca; municipality), and UVG (Usina Delta Sucroenergia; municipality).

2.2 Collection and Expansion of Sugarcane TSH Data

The analysis included records of tons of stalks per hectare (TSH) from 2021, originating from 11 clones cultivated in eight distinct locations, corresponding to the ratoon cane harvest (second harvest). Due to the insufficient number of observations for training the proposed models, it was necessary to use data augmentation to expand the training set, thereby generating synthetic data (Muetanene et al., 2023). The simulation of new individuals was carried out based on the information already available.

The expansion of TSH values by genotype and environments was based on the method developed by Barbosa et al. (2023), who employed the adaptability and stability methodology proposed by Verma et al. (1978). In the present study, however, this approach was modified by replacing the adaptability and stability method with that of Cruz, Torres, and Vencovsky (1989). The segmented regression model was applied to each observed genotype using its performance across eight environments. The environmental index was computed as the deviation of each environment's mean from the overall mean. Based on this, three parameters were estimated per genotype: the intercept ($\hat{\beta}_{0i}$), the slope for favorable environments ($\hat{\beta}_{1i}$), and the deviation from linearity ($\hat{\beta}_{2i}$). The sum ($\hat{\beta}_{1i} + \hat{\beta}_{2i}$) represents the response in unfavorable environments. These values were used to classify genotypes according to the decision rules in Table 1. The stability parameter $\hat{\sigma}^2$ (δ_{ij}) was obtained from the residual variance of the regression and was used to define whether each genotype was stable or unstable. These parameter combinations guided the simulation of new genotypes into 18 adaptability × stability classes.

This methodology, based on segmented regression, comprises the adaptability and stability parameters, which are: the mean ($\hat{\beta}_{0i}$) and the linear response to favorable ($\hat{\beta}_{1i}$) and unfavorable environments ($\hat{\beta}_{1i} + \hat{\beta}_{2i}$). Therefore, the model $Y_{ij} = \beta_{0i} + \beta_{1i} I_i + \beta_{2i}\, T(I_j) + \delta_{ij} + \bar{\varepsilon}_{ij}$, where, $\hat{\sigma}^2_{\delta_i}$ represents the regression deviation of each genotype (Cruz et al., 1999), is used to generate new genotypes in the following classes (Table 1):

Table 1. Genotype adaptability and stability classes, determined by the values of the parameters β_{1i}, β_{2i} and $\hat{\sigma}^2_{\delta_i}$ as described in the method by Cruz, Torres and Vencovsky (1989).

Classes	**Beta parameter values**		**Adaptability**
	Environment		
	unfavorable	favorable	
1	$\beta_{1i} < 1$	$\beta_{1i} + \beta_{2i} < 1$	Specific for unfavorable environments
2	$\beta_{1i} < 1$	$\beta_{1i} + \beta_{2i} = 1$	Specific for unfavorable environments
3	$\beta_{1i} < 1$	$\beta_{1i} + \beta_{2i} > 1$	Ideal
4	$\beta_{1i} = 1$	$\beta_{1i} + \beta_{2i} < 1$	Not recommended
5	$\beta_{1i} = 1$	$\beta_{1i} + \beta_{2i} = 1$	Overall
6	$\beta_{1i} = 1$	$\beta_{1i} + \beta_{2i} > 1$	Specific for favorable environments
7	$\beta_{1i} > 1$	$\beta_{1i} + \beta_{2i} < 1$	Not recommended
8	$\beta_{1i} > 1$	$\beta_{1i} + \beta_{2i} = 1$	Not recommended
9	$\beta_{1i} > 1$	$\beta_{1i} + \beta_{2i} > 1$	Specific for favorable environments

Additionally, the simulated genotypes were classified according to their stability. For each combination of the parameter values β_{1i} and $\beta_{1i} + \beta_{2i}$ value greater than or equal to zero is assigned to $\hat{\sigma}^2_{\delta_i}$ in order to determine whether the genotype is unstable or stable, respectively, resulting in 18 classes of adaptability and stability.

For example, for genotype G1, the estimated parameters were β_{1i} = 0.9 and $\beta_{1i} + \beta_{2i}$ = 1.2, indicating greater response in favorable environments. According to Table 1, G1 is assigned to Class 3 (Ideal adaptability). If the regression deviation $\hat{\sigma}^2(\delta_{ij})$ < threshold, G1 is considered stable. Otherwise, unstable. This classification places G1 into one of 18 possible adaptability × stability combinations used for simulation. We acknowledge the potential for class imbalance, given the limited number of observed genotypes (3G, 5G, 7G scenarios). To address this, genotypes were not simulated evenly across all 18 adaptability × stability classes. Instead, we prioritized biologically realistic parameter combinations, based on observed trends. Moreover, the final models were validated using independent genotypes (held out from the simulation process) to reduce the risk of overfitting.

2.3 Collection of Climatic and Ecophysiological Data

The climate and ecophysiological data will henceforth be referred to as environmental data, which were obtained in 2019, 2020, and 2021 from the NASA Power remote sensing database (https://power.larc.nasa.gov). Using the get_weather function from the EnvRtype package (Costa-Neto et al., 2021) in the R programming environment (R Core Team, 2019), it is possible to download climate data from various locations (Costa-Neto et al., 2021). All environmental covariates extracted were considered for all locations,

The data used in this study include: all-sky insolation incident on a horizontal surface, downward thermal infrared (longwave) radiative flux, measured in megajoules per square meter (MJ m^{-2} d^{-1}); wind speed at 10 meters above the Earth's surface (m s^{-1}); minimum air temperature, maximum air temperature, and dew point temperature (°C); relative air humidity (%); and precipitation (mm d^{-1}). Some variables characterizing ecophysiological processes, such as evapotranspiration, vapor pressure deficit, the slope of the vapor pressure deficit curve, and global solar radiation, were computed using the methods described by Soltani & Sinclair (2012).

The environmental data were used to create a covariate matrix (W), as proposed by Costa-Neto et al. (2021b). Subsequently, the Gaussian Kernel method was applied to matrix W to calculate environmental similarity, a nonlinear method commonly used in genomic prediction. Both techniques were performed using the EnvRtype package through the functions W_matrix and env_kernel, respectively. The proposed method is similar to those used in modeling genomic effects (Eq. 1) (Costa-Neto et al., 2021a).

$$K_e = \exp\left(\frac{hD^2_{ii'}}{Q}\right) \quad (1)$$

Where: The bandwidth factor (h) (assumed to be h = 1 by default) is multiplied by the Euclidean distance $D^2_{ii'} = \sum_k (w_{ik} - w_{i'k})^2$ for each pair of elements $W = \{w_{i,}, w_{i'}\}$. This implies that environmental similarity is determined by the distance between environments, as calculated from the environmental covariates. The scalar variable Q represents the quantile used to weight the environmental distance, assuming $Q = 0{,}5$, which corresponds to the median value of $D^2_{ii'}$.

In this study, the Gaussian kernel bandwidth was set to h = 1, which corresponds to the default setting in the env_kernel() function of the EnvRtype package (Costa-Neto et al., 2021). This value is commonly adopted in studies involving enviromic kernels when prior information on scale sensitivity is unavailable (e.g., Jarquín et al., 2014; Costa-Neto et al., 2021b). Although no formal sensitivity analysis was performed in this study, future work may explore the influence of different bandwidth values (e.g., h < 1 or h > 1) to assess their impact on environmental similarity and model performance.

2.4 Description of Bayesian Mixed Models

After generating the environmental similarity kernels, Bayesian mixed models were fitted using the kernel_model function from the EnvRtype package. This function implements an optimization algorithm for hierarchical Bayesian modeling, as described in the BGGE (Bayesian Genomic Genotype × Environment) package (Granato et al., 2018), which uses Gibbs sampling. All models and structures used in this study were estimated using a Markov Chain Monte Carlo (MCMC) procedure with 15,000 iterations, of which the first 3,000 iterations were discarded as burn-in, and thinning was applied every 20 iterations, yielding 600 posterior samples for inference. This setup follows previous studies that adopted similar configurations (Cuevas et al., 2017; Costa-Neto et al., 2021; Fradgley et al., 2023; Gevartosky et al., 2023).

The described model (Eq. 2) provides a generic framework that can be applied to various kernel combinations to model phenotypic variation across different locations.

$$y = 1\mu + X_f\beta + \sum_{s=1}^{k} g_s + \sum_{r=1}^{l} w_r + \varepsilon \quad (2)$$

where: the vector y combines the means of each genotype in each environment, with the scalar μ representing the overall mean. X_f denotes the incidence matrix associated with the vector of fixed effects β. The random effect vector for genetics (g_s) varies from 1 a k, and the environment-based effect (w_r) varies from 1 to l, and ε represents the residual.

In this study, three distinct model structures were fitted to the data. The MMD model (Main Genomic Effects + Single G×E Deviation) considers exclusively genotypic effects, where $\sum_{s=1}^{p} g_s \neq 0$ e $\sum_{r=1}^{q} w_r = 0$. In this model, g_s is associated with the main genotype effect plus the G×E (G + G×E), as described by Jarquín et al. (2014). With the inclusion of environmental variables, the models begin to consider $\sum_{s=1}^{p} g_s \neq 0$ e $\sum_{r=1}^{q} w_r \neq 0$, where g_s is related to the main genotypic effect (G) or the sum of genotypic effects and interaction (G+G×E). This results in the EMM models (Main Genomic Effects + Main Environmental Effects), which include the main environmental effects (w_r), and the EMDs models (Main Genomic Effects + Main Environmic Effects + Single G×E Deviation), which add the unique variation of the G×E interaction (single G×E deviation). The EMM and EMDs models follow the approach described by Souza et al. (2017).

The Bayesian framework adopted in this study follows the hierarchical structure described by Costa-Neto et al. (2021). The prior distribution for the random effects was assumed to be multivariate normal $b \sim N(0, \sigma^2_u \mathrm{K})$, where K represents the genomic or environmental kernel. Residuals were modeled as $\varepsilon \sim N(0, \sigma^2_\varepsilon \mathrm{I})$, and the prior distributions for the variance components were specified as inverse chi-squared: $\sigma^2_u \sim \chi^{-2}(v_u,\ SC_u), \sigma^2_\varepsilon \sim \chi^{-2}(v_\varepsilon,\ SC_\varepsilon)$.

Posterior estimates were summarized using the posterior mean and the 95% credible interval (CI) from the 600 retained MCMC samples, which supported inference on model components and comparison of predictive performance across structures.

The training and validation sets were organized into three distinct scenarios, each with a defined number of genotypes used in the simulation. Initially, eight of the 11 genotypes were removed from the dataset for validation. From the remaining three, ten genotypes

were generated in each class, totaling 18 classes, which were used for training (3G). Next, in the second scenario, six genotypes were removed for validation, while the remaining five were used for expansion and included in the training (5G). In the third scenario, four genotypes were removed for validation, and the expansion was performed using the remaining genotypes, which were then used for training (7G). Thus, the total number of genotypes used in training was 183, 185, and 187 for the 3G, 5G, and 7G scenarios, respectively.

2.5 Análise estatística

The relationship between the observed and estimated values of sugarcane productivity was assessed using the root mean square error (RMSE), expressed in tons per hectare:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2}{n}} \quad (3)$$

Where $\hat{Y}_i$ and Y_i are the $i-th$, estimated and observed values of the variable, respectively, and n is the sample size.

3. RESULTS

1 Descriptive Statistics of Climatic Variables

Analysis of climatic data across the different environments (AGR, BIO, UAV, UIR, UJA, ULC, USV, UVG) revealed variations in the statistics of the studied variables (Table 2). The Insolation Incident on a Horizontal Surface showed the highest UJA variance (31.25), indicating instability in solar radiation, while UAV showed the lowest (19.38), suggesting greater consistency. For maximum temperatures, UIR recorded the highest variance (12.57) and the highest absolute temperature (43.77 °C), while UJA had the lowest absolute temperature (17.05 °C). Regarding relative humidity, UAV exhibited the highest variance (304.36) and the lowest minimum humidity (19.46%), while UJA had the lowest variance (93.41) and the highest maximum humidity (94.04%). As for precipitation, UJA stood out with the highest variance (44.33), and UAV recorded the highest absolute precipitation (90.81 mm). These results highlight marked differences in climatic conditions across the environments, with UJA and UAV exhibiting the greatest fluctuations in solar radiation, humidity, and precipitation. The statistical results for all climatic variables are available in Appendix A.

Table 2. Descriptive Statistics of Climatic Variables for Different Environments

Variable	Stat	AGR	BIO	UAV	UIR	UJA	ULC	USV	UVG
All Sky Insolation Incident on a Horizontal Surface (MJ m-2 d-1)	var	24.44	20.58	19.38	24.21	31.25	24.23	20.77	22.06
	max	31.44	30.83	30.80	31.94	31.76	30.85	31.17	31.47
	min	3.76	4.67	3.77	4.33	3.78	2.66	4.24	3.01
Minimum Temperature (°C)	var	9.23	9.79	9.76	11.34	8.22	10.75	10.70	10.51
	max	26.25	25.57	28.00	28.93	22.11	26.89	29.22	27.77
	min	6.13	4.53	5.13	3.65	6.22	5.70	3.80	4.84
Maximum Temperature (°C)	var	9.68	10.25	11.77	12.57	9.37	10.03	11.46	10.89
	max	41.58	41.10	43.10	43.77	38.43	42.08	42.96	42.09
	min	19.95	19.02	21.02	17.98	17.05	19.37	18.70	17.88
Relative Humidity (%)	var	217.78	248.94	304.36	277.47	93.41	226.29	277.94	232.90
	max	91.88	91.97	92.73	93.43	94.04	92.17	93.80	91.52
	min	25.62	24.11	19.46	20.97	45.33	22.60	20.39	22.49
Precipitation (mm)	var	43.44	34.47	39.93	27.96	44.33	37.85	27.81	28.91
	max	62.58	67.87	90.81	66.69	63.82	49.98	54.69	51.35
	min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

2 Evaluation of Models Used in Predicting the TSH Variable

The root mean square error (RMSE) was used to assess which models best fit the TSH data. When examining the RMSE values, it is evident that, overall, the MMD model (without environmental information) yielded higher values than the EMM and EMD models, both enriched with environmental information. When analyzing the RMSE values in the MMD model, a notable feature is

the presence of extremely high values at the BIO and UJA locations. In these locations, the RMSE values exceeded 50 t ha^{-1} across all validation scenarios, especially at the BIO location, where they exceeded 100 t ha^{-1}. Additionally, other locations, such as UVG and UAV, showed RMSE values above 34 t ha^{-1}. In contrast, the AGR location, which recorded the fourth-lowest RMSE value, had a maximum of 27 t ha^{-1}, representing a difference of approximately 14 t ha^{-1} between the two. This difference is significant and indicates considerable variation in RMSE values across locations. Furthermore, the USV, UIR, and ULC locations subsequently showed lower RMSE values (Figure 2A). The boxplot in Figure 2B provides a more detailed comparison of the RMSE distribution across genotypes within each environment. It is evident that models using environmental covariates (EMM and EMD) yielded lower variability in prediction errors than the MMD model. The reduced dispersion of RMSE values in models EMM and EMD highlights their consistency across genotypes. This consistency is especially notable in environments like AGR, UIR, and UAV, where both the average error and its variability were reduced.

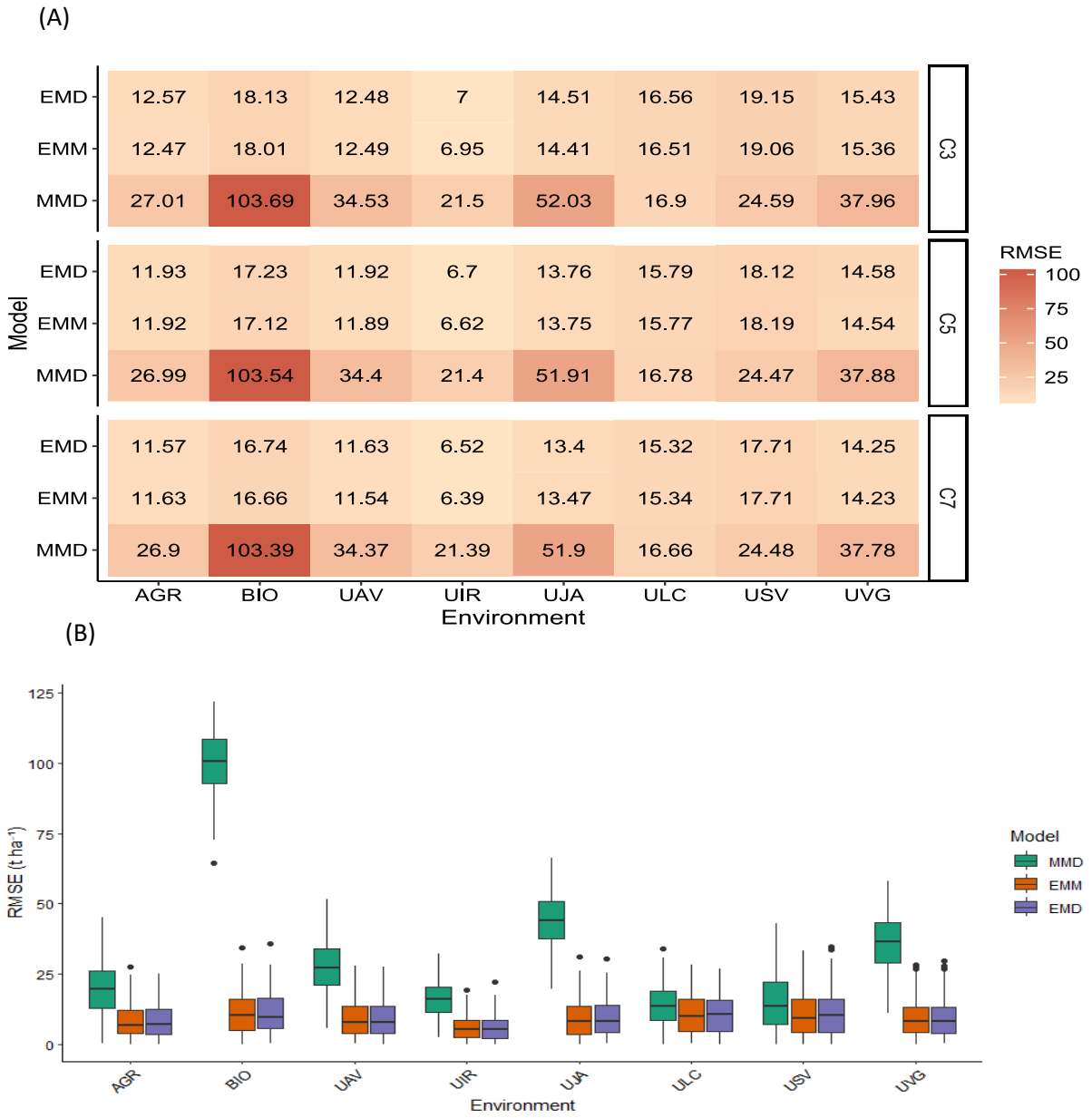

Figure 2. (A) RMSE (Root Mean Square Error) values in t ha^{-1} for each model and validation scenario at each location. Colors range from lighter reddish-brown to darker, representing the lowest and highest RMSE values, respectively. (B) Distribution of RMSE (Root Mean Square Error, t ha^{-1}) values across environments for each model. Boxplots show the variation in prediction error among genotypes. Colors represent the three model structures: MMD (green), EMM (orange), and EMD (purple).

Given that the RMSE values for the MMD model, which does not use environmental covariates, were high, it was decided to exclude it from Figure 3 to improve scale visualization. This decision is justified by the fact that the MMD model, for not considering environmental variables, showed inferior performance, which is why comparisons were made exclusively between the EMM and EMD models, which do.

In the three validation scenarios, RMSE values varied as the number of genotypes used for training increased. In scenario 3G, eight genotypes were removed from the dataset for validation, and the remaining three were used to generate 10 genotypes per class, totaling 183 genotypes in the training set. In the 5G scenario, six genotypes were removed for validation, while five were expanded, resulting in 185 genotypes in the training set. In scenario 7G, four genotypes were removed for validation, leaving 187 genotypes in the training set.

As the number of expanded genotypes used in training increases, the RMSE decreases. Although these values vary across environments, the overall trend indicates a consistent decrease as the number of genotypes in the expansion increases.

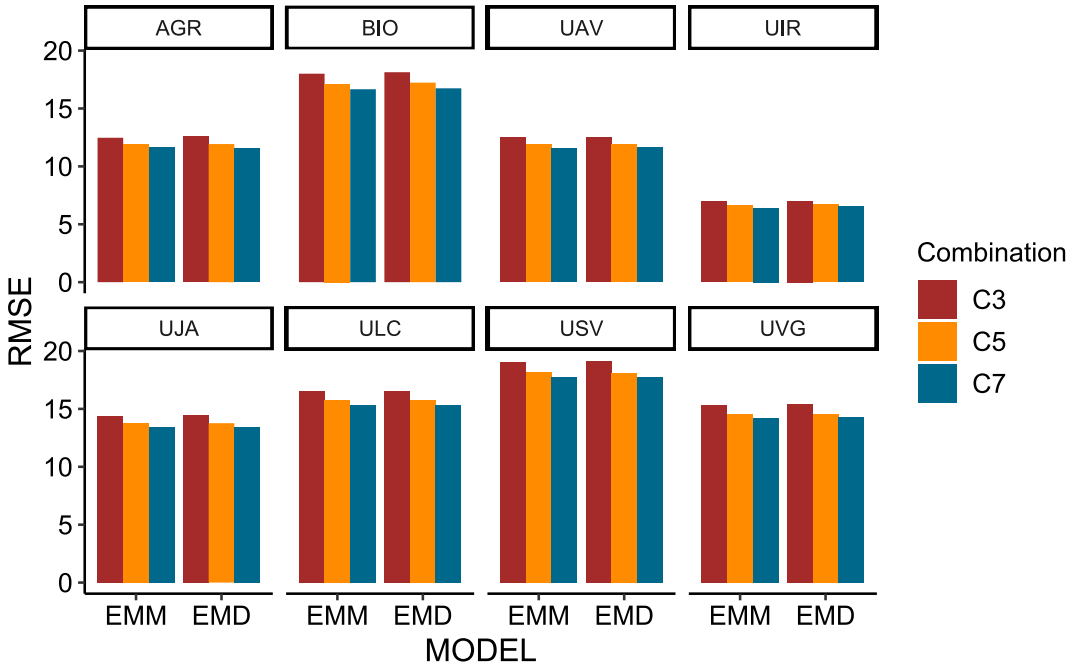

Figure 3. Bar chart showing RMSE values for the 3G, 5G, and 7G validation scenarios using the EMM and EMD models at different locations. In these scenarios, 3G, 5G, and 7G represent the use of genotypes 3, 5, and 7 for simulation and training, while the remaining genotypes were allocated for validation. The colors red, yellow, and blue correspond to scenarios C3, C5, and C7, respectively.

With the inclusion of environmental covariates in the prediction models, the RMSE values at the BIO and UJA locs. Bar chart showing RMSE values for the 3G, 5G, and 7G validation scenarios using the EMM and EMD models at different locations. In these scenarios, 3G, 5G, and 7G represent the use of genotypes 3, 5, and 7 for simulation and training, while the remaining genotypes were allocated for validation. The colors red, yellow, and blue correspond to scenarios C3, C5, and C7, respectively. Thus, when comparing the EMM and EMD models, the errors become more similar across locations, and the variation between them is less evident.

When evaluating the EMM and EMD models, we observe variation in RMSE values across environments. The ULC location showed consistent RMSE values across models, ranging from 15.32 t ha^{-1} to 16.9 t ha^{-1}, depending on the validation scenario. However, the UIR, UAV, and AGR locations showed the lowest errors when environmental covariates were included (Figures 2 and 3).

Figure 5A displays the scatter plot between the observed and predicted values for the EMM and EMD models at each location. The MMD model was removed due to the presence of outliers.

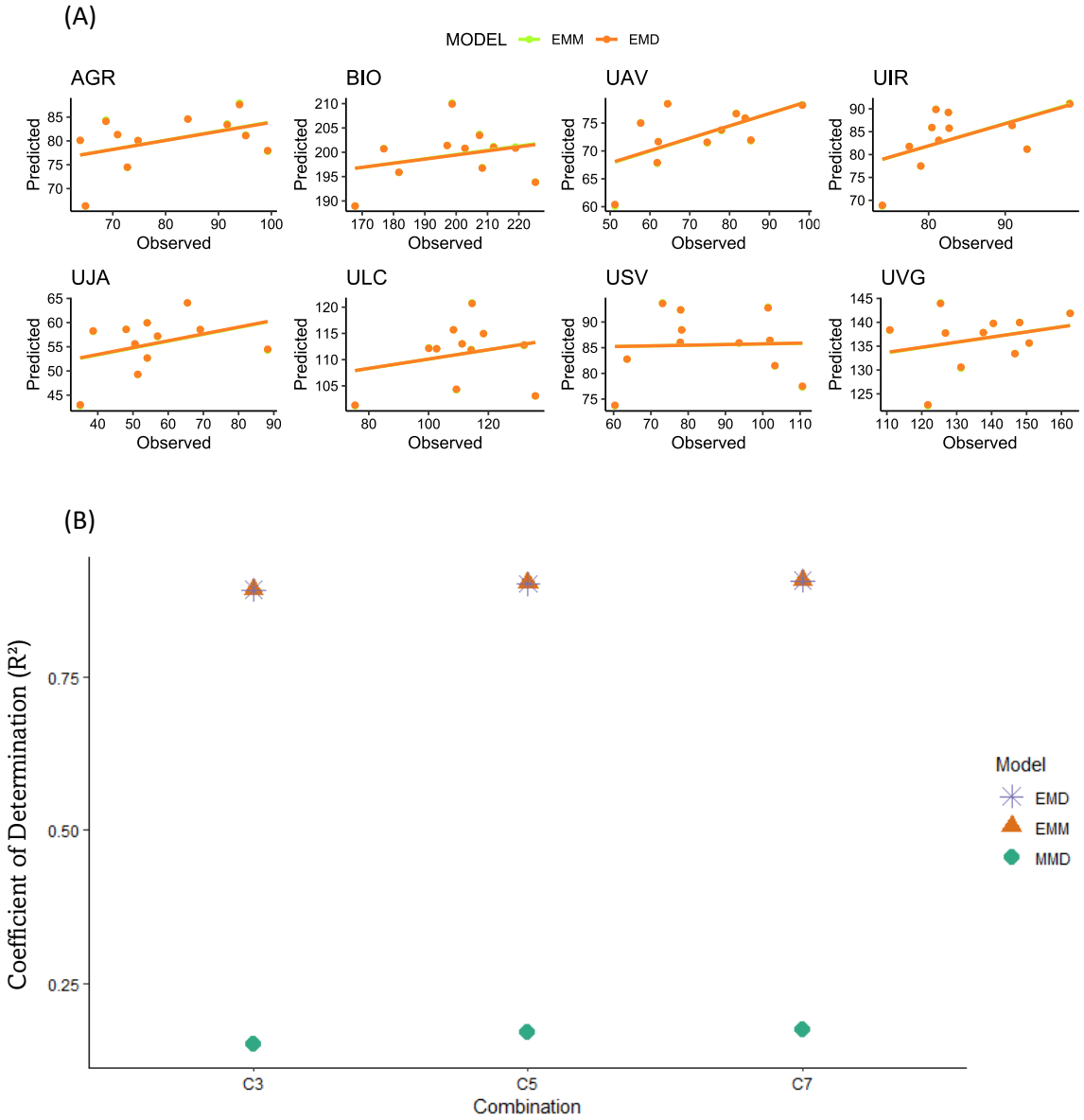

Figure 4. (A) Predicted and observed values of tons of sugarcane per hectare (TSH) at each location, predicted by the EMM and EMD models. The EMM model is represented by the green line and points, while the EMD model is represented by the orange line and points. (B) Coefficient of determination (R^2) for each model across training genotype combinations (C3, C5, C7). Shapes indicate model structures: MMD (circle), EMM (triangle), and EMD (star).

As seen in Figure 5, it is not possible to differentiate the points and trend lines of the models (EMM and EMD) due to the proximity of the estimated values. Therefore, the evaluation of the models will be done collectively. In general, the points are distant from the trend line; however, the low number of points interfered with the evaluation of the scatter plot. Nevertheless, the points show a certain trend in some locations (USV, BIO). In the locations with higher errors (USV and BIO), the trend line is not inclined; in this sense, the model may not capture the overall relationship in the data. Additionally, the points are not distributed uniformly; they are scattered relative to the trend line.

Another important point is the distance of the points from the trend line. In the USV, ULC, and BIO locations, most points are farther from the trend line, indicating a higher residual error. Meanwhile, in the UIR, UAV, and AGR locations, the points are closer to the trend line but still present some outliers, which may be anomalies that the model is not capturing or errors in the original data. It is noticeable that not all locations show a consistent dispersion across all ranges of the graphs, as seen, for example, in the BIO and USV locations. The presence of heteroscedasticity may indicate that the model is not adequately capturing the variability of the data. To complement the analysis of model accuracy, the coefficient of determination (R^2) was evaluated for each model across different genotype combinations (Figure 5B). The MMD model, which does not use environmental covariates, presented consistently lower R^2 values. In contrast, both EMM and EMD models achieved R^2 values above 0.8 in all scenarios, with minimal differences between them.

4. DISCUSSION

The use of prediction models in sugarcane breeding programs, including environmental covariates, remains underexplored. The reliability of such predictions, particularly during the experimental phase of the program, enables the early identification of promising genotypes and the anticipation of their performance under varying environmental conditions. Knowing the behavior of genotypes in advance reduces the need for extensive multi-environment trials, ultimately lowering costs, accelerating selection cycles, and enhancing genetic gain (Neyhart et al., 2022).

Among modeling strategies, Bayesian approaches stand out for their flexibility in incorporating covariance structures derived from phenotypic, genotypic, and environmental sources. Additionally, they accommodate nonlinear genotype responses to environmental variation, more accurately reflecting biological reality (Granato et al., 2018). In this study, three Bayesian mixed model structures were compared across experimental locations. The MMD model relied only on phenotypic data and G×E interaction. In contrast, the EMM and EMD models incorporated an environmental covariance matrix, with EMD also modeling a unique G×E deviation.

When examining RMSE values for the BIO and UJA environments under the MMD model, markedly high errors were observed. In UJA, these errors can be explained by a cultivation setting marked by lower average temperatures and uneven rainfall distribution—conditions opposite to those of other environments. This finding highlights the relevance of environmental similarity in improving predictive reliability for untested sites, supporting results from Jarquin et al. (2021) and Neyhart et al. (2022).

In the BIO location, although environmental conditions resembled those of other locations, RMSE remained high. This was attributed to inconsistencies in the observed TSH data, including abnormally high values and variation among replicates within the same genotype. Since the training set was generated by simulating data from a small original sample, potential measurement errors and outliers could have amplified prediction inaccuracies (Ly et al., 2018).

Despite these challenges, the inclusion of environmental covariates significantly improved model performance in most locations. In environments such as AGR, UAV, and UIR, the models showed markedly lower RMSE values with covariates. This reinforces previous findings suggesting that even a limited number of relevant covariates can be sufficient in heterogeneous environments (Li et al., 2018; Millet et al., 2019). In our analysis, preliminary tests indicated that solar radiation, vapor pressure deficit (VPD), and early-stage temperature were the most influential covariates. This highlights the importance of variable selection based on crop physiology. In contrast, in homogeneous environments such as ULC, where G×E interaction is minimal, the inclusion of environmental covariates had little impact on prediction accuracy (Neyhart et al., 2022).

According to Jarquin et al. (2013), predictions tend to be more accurate when environmental conditions across sites are similar. However, given the small number of genotypes (n = 11) and environments (n = 8) in our study, additional strategies were needed. Thus, we employed genotype expansion via simulation, which led to consistent improvements in model fit as the number of genotypes increased.

To minimize the risk of overfitting to synthetic data, we validated models using genotypes excluded from the simulation. These individuals were reserved for testing to ensure that model evaluation was based solely on observed genotypes. The consistent reduction in RMSE across validation scenarios (C3, C5, C7) indicates that the simulation strategy contributed to generalization rather than distortion. Moreover, the gain in consistency, even with modest reductions in RMSE (about 0.10 t ha^{-1}), supports the utility of data augmentation in sugarcane prediction.

Some authors suggest that environmental variables are better leveraged when organized by phenological stages, which improves interpretability and accuracy (Heslot et al., 2014; Monteverde et al., 2019; Rincent et al., 2019). While this approach was not feasible in our study due to variable cutting schedules in ratoon cane across locations, future work could benefit from integrating phenological information into the modeling process.

Overall, our results demonstrate that environmental covariates are valuable tools in predictive modeling, particularly when genomic and pedigree data are unavailable. Given that environmental data are accessible and free, they offer a practical and scalable alternative for improving predictions.

If detailed genotypic or pedigree information were available, model performance could likely be further improved (Hayes et al., 2021). However, the high cost of genotyping, especially for crops such as sugarcane, limits its feasibility. In this context, environmental data remain a cost-effective solution, helping guide breeding programs by identifying representative sites and reducing the number of field evaluations required.

This study has several limitations. The reduced number of observed genotypes limited genetic diversity, although partially offset by the simulation strategy. The lack of genomic data precluded explicit modeling of genetic relationships. Additionally, the inability to align phenological stages across locations due to variability in cutting schedules may have introduced noise in covariate interpretation. Future studies combining genomic data with phenology-aligned environmental summaries may achieve even greater predictive accuracy.

These findings highlight important implications for future sugarcane breeding strategies. The demonstrated utility of environmental covariates in predictive models suggests that enviromic approaches can be incorporated into the final stages of

selection, especially in contexts where genomic data are limited or unavailable. In large-scale breeding programs, this strategy could help breeders prioritize representative environments, reduce the number of field trials, and accelerate the selection cycle. Moreover, by integrating readily available environmental data, the methodology offers a scalable, low-cost solution that optimizes resource allocation and improves genotype targeting to suitable environments.

5. CONCLUSION

The inclusion of environmental covariates proved crucial for improving the model's accuracy in predicting sugarcane yield (TSH). Furthermore, increasing the number of genotypes used to expand the dataset significantly reduced prediction errors.

These results highlight the potential of integrating environmental information into the final stages of breeding programs, offering a low-cost and scalable strategy to improve predictive performance in the absence of genomic data. The methodology adopted here can be extended to cover a wider range of environments and genotypes, allowing breeders to prioritize experimental efforts in more representative locations. As environmental data are freely available and increasingly accurate through remote sensing platforms, this approach has great potential for application in large-scale breeding programs, supporting the selection of more productive and stable genotypes in diverse agroecological conditions.

6. REFERENCE

Acosta-Pech, R., J. Crossa, G. De Los Campos, S. Teyssèdre, B. Claustres, S. Pérez-Elizalde e P. Pérez-Rodríguez, 2017. Genomic models with genotype × environment interaction for predicting hybrid performance: an application in maize hybrids. Theoretical and Applied Genetics, 130: 1431-1440.

Annicchiarico, P., F. Bellah e T. Chiari, 2006. Repeatable genotype × location interaction and its exploitation by conventional and GIS-based cultivar recommendation for durum wheat in Algeria. European Journal of Agronomy, 24(1): 70-81.

Barbosa, M.H.P. e L.C.I. da Silveira, 2015. Breeding program and cultivar recommendations. Em Sugarcane, Ed. Academic Press, pp: 241-255.

Barbosa, W.F., A.N. Silva, J.P.M. Silva, C.M.C. Manhães e A.R. Gallo, 2023. Artificial neural networks based on segmented model for adaptability and stability evaluation of soybean genotypes. Australian Journal of Crop Science, 17(9): 735-740.

Caetano, J.M. e D. Casaroli, 2017. Sugarcane yield estimation for climatic conditions in the state of Goiás. Revista Ceres, 64: 298-306.

CONAB – Companhia Nacional de Abastecimento, 2023. Acompanhamento da safra brasileira de cana-de-açúcar safra 2022/23. 4º levantamento. Brasília, DF, 10(4), abril.

CONAB – Companhia Nacional de Abastecimento, 2023. Acompanhamento da safra brasileira de cana-de-açúcar safra 2023/24. 3º levantamento. Brasília, DF, 11(3), novembro.

Conselho dos Produtores de Cana-de-Açúcar, Açúcar e Álcool do Estado do Paraná – CONSECANA, 2012. Manual de Instruções, 3. ed. Curitiba-PR: 118p.

Costa-Neto, G., G. Galli, H.F. Carvalho, J. Crossa e R. Fritsche-Neto, 2021a. EnvRtype: a software to interplay enviromics and quantitative genomics in agriculture. G3, 11(4): jkab040.

Costa-Neto, G., R. Fritsche-Neto e J. Crossa, 2021b. Nonlinear kernels, dominance, and envirotyping data increase the accuracy of genome-based prediction in multi-environment trials. Heredity, 126(1): 92-106.

Cursi, D. E., Gazaffi, R., Hoffmann, H. P., Brasco, T. L., do Amaral, L. R., & Dourado Neto, D. (2021). Novel tools for adjusting spatial variability in the early sugarcane breeding stage. *Frontiers in Plant Science*, *12*, 749533.

Cuevas, J., Crossa, J., Montesinos-López, O. A., Burgueño, J., Pérez-Rodríguez, P., & de Los Campos, G. (2017). Bayesian genomic prediction with genotype× environment interaction kernel models. *G3: Genes, Genomes, Genetics*, *7*(1), 41-53.

Crossa, J., O.A. Montesinos-Lopez, P. Pérez-Rodríguez, G. Costa-Neto, R. Fritsche-Neto, R. Ortiz, ... e R. Rincent, 2022. Genome and environment based prediction models and methods of complex traits incorporating genotype × environment interaction. Em Genomic Prediction of Complex Traits: Methods and Protocols, pp: 245-283.

Crossa, J., R. Fritsche-Neto, O.A. Montesinos-Lopez, G. Costa-Neto, S. Dreisigacker, A. Montesinos-Lopez e A.R. Bentley, 2021. The modern plant breeding triangle: optimizing the use of genomics, phenomics, and enviromics data. Frontiers in Plant Science, 12: 651480.

Cruz, C.D., R.D.A. Torres e R. Vencovsky, 1989. An alternative approach to the stability analysis proposed by Silva and Barreto.

Cruz, C.D., A.J. Regazzi e P.C.S. Carneiro, 2004. Modelos biométricos aplicados ao melhoramento genético, 3. ed. Viçosa: 480p.

De Los Campos, G., H. Naya, D. Gianola, J. Crossa, A. Legarra, E. Manfredi, ... e J.M. Cotes, 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics, 182(1): 375-385.

De Los Campos, G., P. Pérez-Rodríguez, M. Bogard, D. Gouache e J. Crossa, 2020. A data-driven simulation platform to predict cultivars' performances under uncertain weather conditions. Nature Communications, 11(1): 4876.

des Marais, D.L., K.M. Hernandez e T.E. Juenger, 2013. Genotype-by-environment interaction and plasticity: exploring genomic responses of plants to the abiotic environment. Annual Review of Ecology, Evolution, and Systematics, 44: 5-29.

Dlamini, N.E. e M. Zhou, 2024. Predicting Ratooning Ability of Sugarcane Varieties in Selection Trials. Sugar Tech, 26(1): 52-62.
Eberhart, S.T. e W. Russell, 1966. Stability parameters for comparing varieties 1. Crop Science, 6(1): 36-40.
Fradgley, N. S., Bacon, J., Bentley, A. R., Costa-Neto, G., Cottrell, A., Crossa, J., ... & Gardner, K. A. (2023). Prediction of near-term climate change impacts on UK wheat quality and the potential for adaptation through plant breeding. Global Change Biology, 29(5), 1296-1313.
Gauch, H.J., 1992. Statistical analysis of regional yield trials: AMMI analysis of factorial designs, xi+-278.
Hayes, B.J., X. Wei, P. Joyce, F. Atkin, E. Deomano, J. Yue, ... e K.P. Voss-Fels, 2021. Accuracy of genomic prediction of complex traits in sugarcane. Theoretical and Applied Genetics, 134: 1455-1462.
Heinemann, A. B., Costa-Neto, G., Fritsche-Neto, R., da Matta, D. H., & Fernandes, I. K. (2022). Enviromic prediction is useful to define the limits of climate adaptation: a case study of common bean in Brazil. *Field Crops Research*, *286*, 108628.
Heinemann, A. B., Costa-Neto, G., da Matta, D. H., Fernandes, I. K., & Stone, L. F. (2024). Harnessing crop models and machine learning for a spatial-temporal characterization of irrigated rice breeding environments in Brazil. *Field Crops Research*, *315*, 109452.
Heslot, N., H.P. Yang, M.E. Sorrells e J.L. Jannink, 2012. Genomic selection in plant breeding: a comparison of models. Crop Science, 52(1): 146-160.
Heslot, N., D. Akdemir, M.E. Sorrells e J.L. Jannink, 2014. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theoretical and Applied Genetics, 127: 463-480.
Jarquín, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt, J. Lorgeou, ... e G. De Los Campos, 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theoretical and Applied Genetics, 127: 595-607.
Li, X., T. Guo, Q. Mu, X. Li e J. Yu, 2018. Genomic and environmental determinants and their interplay underlying phenotypic plasticity. Proceedings of the National Academy of Sciences, 115(26): 6679-6684.
los Campos, G., 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theoretical and Applied Genetics, 127: 595-607.
Mahadevaiah, C., C. Appunu, K. Aitken, G.S. Suresha, P. Vignesh, H.K. Mahadeva Swamy, ... e B. Ram, 2021. Genomic selection in sugarcane: Current status and future prospects. Frontiers in Plant Science, 12: 708233.
Manhães, C.M.C., R.F. Garcia, F.M.A. Francelino, H. de Oliveira Francelino e F.C. Coelho, 2015. Fatores que afetam a brotação e o perfilhamento da cana-de-açúcar. Revista Vértices, 17(1): 163-181.
Mangiafico, S.S., 2016. Summary and Analysis of Extension Program Evaluation in R, Version 1.18.1. Disponível em: https://rcompanion.org/handbook/. Acesso em 10 jan. 2021.
Medar, R.A., V.S. Rajpurohit e A.M. Ambekar, 2019. Sugarcane crop yield forecasting model using supervised machine learning. International Journal of Intelligent Systems and Applications, 11(8): 11.
Millet, E.J., W. Kruijer, A. Coupel-Ledru, S. Alvarez Prado, L. Cabrera-Bosquet, S. Lacube, ... e F. Tardieu, 2019. Genomic prediction of maize yield across European environmental conditions. Nature Genetics, 51(6): 952-956.
Miphokasap, P. e W. Wannasiri, 2018. Estimations of nitrogen concentration in sugarcane using hyperspectral imagery. Sustainability, 10(4): 1266.
Montesinos-López, O.A., A. Montesinos-López, J. Crossa, D. Gianola, C.M. Hernández-Suárez e J. Martín-Vallejo, 2018. Multi-trait, multi-environment deep learning modeling for genomic-enabled prediction of plant traits. G3: Genes, Genomes, Genetics, 8(12): 3829-3840.
Monteverde, E., L. Gutierrez, P. Blanco, F. Pérez de Vida, J.E. Rosas, V. Bonnecarrère, ... e S. McCouch, 2019. Integrating molecular markers and environmental covariates to interpret genotype by environment interaction in rice (Oryza sativa L.) grown in subtropical areas. G3: Genes, Genomes, Genetics, 9(5): 1519-1531.
Muetanene, B.A., L.A. Peternelli, P. Carneiro, F.L. da Silva, D.P. Barbosa e J.I.R. Júnior, 2023. Selection indices and support vector machines in the selection of sugarcane families. Brazilian Journal of Agriculture – Revista de Agricultura, 98(1): 23-37.
Pagani, V., T. Stella, T. Guarneri, G. Finotto, M. Van den Berg, F.R. Marin, ... e R. Confalonieri, 2017. Forecasting sugarcane yields using agro-climatic indicators and Canegro model: A case study in the main production region in Brazil. Agricultural Systems, 154: 45-52.
Pérez-Rodríguez, P., J. Crossa, K. Bondalapati, G. De Meyer, F. Pita e G.D.L. Campos, 2015. A pedigree-based reaction norm model for prediction of cotton yield in multienvironment trials. Crop Science, 55(3): 1143-1151.
Poudyal, C., L.F. Costa, H. Sandhu, Y. Ampatzidis, D.C. Odero, O.C. Arbelo e R.H. Cherry, 2022. Sugarcane yield prediction and genotype selection using unmanned aerial vehicle-based hyperspectral imaging and machine learning. Agronomy Journal, 114(4): 2320-2333.
Ramburan, S., Zhou, M., & Labuschagne, M. (2011). Interpretation of genotype× environment interactions of sugarcane: Identifying significant environmental factors. *Field Crops Research*, *124*(3), 392-399.
Ramburan, S., Zhou, M., & Labuschagne, M. T. (2012). Investigating test site similarity, trait relations and causes of genotype× environment interactions of sugarcane in the Midlands region of South Africa. *Field crops research*, *129*, 71-80.
Resende, R.T., H.P. Piepho, G.J. Rosa, O.B. Silva-Junior, F.F. e Silva, M.D.V. de Resende e D. Grattapaglia, 2021. Enviromics in breeding: applications and perspectives on envirotypic-assisted selection. Theoretical and Applied Genetics, 134: 95-112.
Resende, R. T., Hickey, L., Amaral, C. H., Peixoto, L. L., Marcatti, G. E., & Xu, Y. (2024). Satellite-enabled enviromics to enhance crop improvement. *Molecular Plant*, *17*(6), 848-866.
Rincent, R., M. Malosetti, B. Ababaei, G. Touzy, A. Mini, M. Bogard, ... e F. Van Eeuwijk, 2019. Using crop growth model stress covariates and AMMI decomposition to better predict genotype-by-environment interactions. Theoretical and Applied Genetics, 132: 3399-3411.

Rodrigues, M., E. Cezar, G.L.A.A. dos Santos, A.S. Reis, R.H. Furlanetto, R.B. de Oliveira, ... e M.R. Nanni, 2022. Estimating technological parameters and stem productivity of sugarcane treated with rock powder using a proximal spectroradiometer Vis-NIR-SWIR. Industrial Crops and Products, 186: 115278.

Rodrigues, J.D., 1995. Fisiologia da cana-de-açúcar. Botucatu: UNESP, pp: 419-449.

Rudorff, B.F.T., D.A. Aguiar, W.F. Silva, L.M. Sugawara, M. Adami e M.A. Moreira, 2010. Estudos sobre a rápida expansão da cana-de-açúcar para produção de etanol no Estado de São Paulo (Brasil) utilizando dados Landsat. Sensoriamento Remoto, 2(4): 1057-1076.

Singels, A., M. Jones, F. Marin, A. Ruane e P. Thorburn, 2014. Predicting climate change impacts on sugarcane production at sites in Australia, Brazil and South Africa using the Canegro model. Sugar Tech, 16: 347-355.

Scarpare, F.V., O.D.J. van Lier, S.T.R. Corrêa, A. Hugo, C. Barros, F.R. Marin e D.S.P. Nassif, 2012. Modelos de crescimento da cana-de-açúcar e sua parametrização. Revista de Agricultura, 87(1): 66-80.

Scarpari, M.S. e E.G.F.D. Beauclair, 2004. Sugarcane maturity estimation through edaphic-climatic parameters. Scientia Agricola, 61: 486-491.

Schmidt, P., J. Möhring, R.J. Koch e H.P. Piepho, 2018. More, larger, simpler: How comparable are on-farm and on-station trials for cultivar evaluation?. Crop Science, 58(4): 1508-1518.

Shukla, S.K., L. Sharma, S.K. Awasthi e A.D. Pathak, 2017. Sugarcane in India. Package of practices for different agro–climatic zones. All Indian Coordinated Research Project on Sugarcane, IISR Lucknow, Uttar Pradesh, pp: 1-64.

Silva, F.L.D., M.H.P. Barbosa, M.D.V. de Resende, L.A. Peternelli e C.Á. Pedrozo, 2015. Efficiency of selection within sugarcane families via simulated individual BLUP. Crop Breeding and Applied Biotechnology, 15: 1-9.

Sukumaran, S., J. Crossa, D. Jarquín e M. Reynolds, 2017. Pedigree-based prediction models with genotype × environment interaction in multienvironment trials of CIMMYT wheat. Crop Science, 57(4): 1865-1880.

Tena, E., F. Goshu, H. Mohamad, M. Tesfa, D. Tesfaye e A. Seife, 2019. Genotype × environment interaction by AMMI and GGE-biplot analysis for sugar yield in three crop cycles of sugarcane (Saccharum officinarum L.) clones in Ethiopia. Cogent Food & Agriculture, 5(1): 1651925.

Toppa, E.V.B., C.J. Jadoski, A. Julianetti, T. Hulshof e E.O. Ono, 2010. Physiology development in the vegetative stage of sugarcane. Applied Research & Agrotechnology, 3(2): 169-186.

Verma, M.M., G.S. Chahal e B.R. Murty, 1978. Limitations of conventional regression analysis: a proposed modification. Theoretical and Applied Genetics, 53: 89–91.

Van Eeuwijk, F.A., D. Bustos-Korts, E.J. Millet, M.P. Boer, W. Kruijer, A. Thompson, ... e S.C. Chapman, 2019. Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. Plant Science, 282: 23-39.

Viana, J.L., J.L.M. de Souza, A.K. Hoshide, R.A. de Oliveira, D.C. de Abreu e W.M. da Silva, 2023. Estimating sugarcane yield in a subtropical climate using climatic variables and soil water storage. Sustainability, 15(5): 4360.

Xu, Y., 2016. Envirotyping for deciphering environmental impacts on crop plants. Theoretical and Applied Genetics, 129: 653-673.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Acknowledgment

The first author acknowledges the financial support provided by the Coordination for the Improvement of Higher Education Personnel (CAPES) through a scholarship granted for 48 months, from March 2020 to March 2024. The authors also thank the National Council for Scientific and Technological Development (CNPq) and the Research Support Foundation of the State of Minas Gerais (FAPEMIG) for the institutional support and indirect funding that contributed to the development of this work.

Authors' contributions

All authors contributed equally to this work.

Conflict of interest

The authors declare that they have no conflicts of interest.

Tays Silva Batista holds a degree in Forest Engineering and has developed experience in plant pathology, forest inventory, and biometric measurements of native and eucalyptus species. During her master's studies, she worked with artificial neural networks to predict productivity in eucalyptus. In her Ph.D., she joined a statistical research lab and deepened her skills in R programming and enviromics—integrating climatic data into genotype-by-environment interaction studies to improve genotype selection. Her doctoral research focused on optimizing crop performance predictions across diverse environments. She is currently a postdoctoral researcher at the Federal University of Viçosa (UFV), continuing her research in the field. She received research grants from CAPES during her master's and Ph.D. programs, and from CNPq during her postdoctoral work.

7. Appendix

Appendix A – Descriptive statistics of climatic variables used in the prediction models

Variable	Stat	AGR	BIO	UAV	UIR	UJA	ULC	USV	UVG
All Sky Insolation Incident on a Horizontal Surface (MJ m-2 d-1)	mean	19.58	20.15	20.19	20.14	17.68	19.34	20.10	19.75
	var	24.44	20.58	19.38	24.21	31.25	24.23	20.77	22.06
	max	31.44	30.83	30.80	31.94	31.76	30.85	31.17	31.47
	min	3.76	4.67	3.77	4.33	3.78	2.66	4.24	3.01
Downward Thermal Infrared (Longwave) Radiative Flux (MJ m-2 d-1)	mean	32.18	32.54	32.90	33.43	32.34	32.24	33.06	32.13
	var	6.77	7.86	7.99	8.19	5.26	7.49	8.10	8.04
	max	36.97	38.03	38.66	39.32	36.54	37.08	38.81	37.72
	min	24.59	22.48	22.65	22.74	24.57	23.87	22.94	22.32
Wind Speed (m/s)	mean	2.09	2.06	2.04	1.97	2.15	1.92	2.01	2.15
	var	0.53	0.54	0.61	0.49	0.52	0.42	0.52	0.54
	max	5.58	5.24	5.45	5.58	5.69	4.98	5.71	5.22
	min	0.52	0.61	0.63	0.61	0.60	0.59	0.61	0.63
Minimum Temperature (°C)	mean	17.67	18.34	19.65	19.94	16.25	17.43	20.16	19.16
	var	9.23	9.79	9.76	11.34	8.22	10.75	10.70	10.51
	max	26.25	25.57	28.00	28.93	22.11	26.89	29.22	27.77
	min	6.13	4.53	5.13	3.65	6.22	5.70	3.80	4.84
Maximum Temperature (°C)	mean	29.92	29.80	31.54	32.38	27.38	29.43	32.03	30.72
	var	9.68	10.25	11.77	12.57	9.37	10.03	11.46	10.89
	max	41.58	41.10	43.10	43.77	38.43	42.08	42.96	42.09
	min	19.95	19.02	21.02	17.98	17.05	19.37	18.70	17.88
Dew/Frost Point (°C)	mean	15.69	15.57	16.32	15.76	16.06	15.57	16.21	16.10
	var	19.37	21.91	26.53	28.05	10.51	20.28	26.96	21.54
	max	22.20	22.20	23.27	23.17	21.98	22.21	23.24	22.94
	min	-3.36	-7.47	-9.61	-7.29	3.53	-4.31	-8.02	-7.02
Relative Humidity (%)	mean	66.47	64.59	62.86	58.81	76.10	66.95	60.19	63.32
	var	217.78	248.94	304.36	277.47	93.41	226.29	277.94	232.90
	max	91.88	91.97	92.73	93.43	94.04	92.17	93.80	91.52
	min	25.62	24.11	19.46	20.97	45.33	22.60	20.39	22.49
Precipitation (mm)	mean	2.98	3.03	3.16	2.52	3.22	2.98	2.68	2.90
	var	43.44	34.47	39.93	27.96	44.33	37.85	27.81	28.91
	max	62.58	67.87	90.81	66.69	63.82	49.98	54.69	51.35
	min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Evapotranspiration (mm d-1)	mean	2.50	3.64	2.98	2.42	2.98	2.83	2.47	2.92
	var	8.45	8.64	11.25	8.49	8.00	8.10	9.31	9.53
	max	11.92	11.93	12.58	10.93	11.23	10.87	11.97	12.11
	min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Vapour pressure deficit (KPa d-1)	mean	1.33	1.36	1.58	1.78	0.93	1.27	1.69	1.48
	var	0.45	0.52	0.78	0.75	0.20	0.45	0.72	0.54
	max	4.22	4.23	4.78	5.15	2.86	4.16	4.84	4.46
	min	0.20	0.25	0.27	0.20	0.16	0.27	0.19	0.26
Radiation on the top of the atmosphere (MJ m-2 d-1)	mean	34.31	34.41	34.46	34.25	34.12	34.23	34.42	34.20
	var	40.70	38.75	37.79	41.81	44.26	42.22	38.67	42.71
	max	42.03	41.90	41.83	42.11	42.27	42.14	41.89	42.17
	min	24.30	24.60	24.76	24.12	23.75	24.06	24.62	23.99