

**TAYS SILVA BATISTA**

**BIOMETRIA DE DADOS AMBIENTAIS APLICADA NA MODELAGEM E  
CARACTERIZAÇÃO DE MEGA-AMBIENTES NO MELHORAMENTO  
GENÉTICO DA CANA-DE-AÇÚCAR**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e melhoramento de plantas, para obtenção do título de *Doctor Scientiae*.

Orientador: Luiz Alexandre Peternelli

**VIÇOSA - MINAS GERAIS**

**2024**

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

T

B333b  
2024

Batista, Tays Silva, 1992-  
Biometria de dados ambientais aplicada na modelagem e  
caracterização de mega-ambientes no melhoramento genético da  
cana-de-açúcar / Tays Silva Batista. – Viçosa, MG, 2024.  
1 tese eletrônica (101 f.): il. (algumas color.).

Texto em português e inglês.

Inclui anexo.

Orientador: Luiz Alexandre Peternelli.

Tese (doutorado) - Universidade Federal de Viçosa,  
Departamento de Estatística, 2024.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2024.515>

Modo de acesso: World Wide Web.

1. Cana-de-açúcar - Melhoramento genético. 2. Interação  
genótipo-ambiente. 3. Clonagem. 4. Análise de regressão.  
5. Cana-de-açúcar - Fatores climáticos. I. Peternelli, Luiz  
Alexandre, 1966-. II. Universidade Federal de Viçosa.  
Departamento de Estatística. Programa de Pós-Graduação em  
Genética e Melhoramento. III. Título.

CDD 22. ed. 633.612

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


**TAYS SILVA BATISTA**

**BIOMETRIA DE DADOS AMBIENTAIS APLICADA NA MODELAGEM E  
CARACTERIZAÇÃO DE MEGA-AMBIENTES NO MELHORAMENTO  
GENÉTICO DA CANA-DE-AÇÚCAR**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e melhoramento de plantas, para obtenção do título de *Doctor Scientiae*.

APROVADA: 21 de maio de 2024.


Assentimento:

Documento assinado digitalmente  
 **TAYS SILVA BATISTA**  
Data: 20/08/2024 17:16:11-0300  
Verifique em <https://validar.iti.gov.br>

---

Tays Silva Batista

Autor

Documento assinado digitalmente  
 **LUIZ ALEXANDRE PETERNELLI**  
Data: 20/08/2024 22:06:41-0300  
Verifique em <https://validar.iti.gov.br>

---

Luiz Alexandre Peternelli

Orientador

## AGRADECIMENTOS

A Deus, por me conceder força, sabedoria e perseverança ao longo desta jornada.

Ao meu orientador, Luiz Alexandre Peternelli, pela orientação, apoio e motivação constantes, que foram fundamentais para a realização deste trabalho.

Aos meus pais, Maria de Fátima Luiz da Silva e Agumair Alves Batista, pelo amor incondicional, e incentivo em todos os momentos. A vocês, dedico esta conquista.

Ao meu irmão, Cristian Silva Batista, pelo companheirismo e suporte, que me motivaram a seguir em frente.

Ao meu companheiro, André Silveira de Assis, pela paciência, compreensão e encorajamento, fundamentais para a conclusão desta etapa.

Aos meus parceiros de laboratório, pela colaboração, discussões enriquecedoras e pelo ambiente de trabalho harmonioso e produtivo.

À Universidade Federal de Viçosa, pela oportunidade de realizar a pós-graduação e pelo apoio institucional essencial para o desenvolvimento deste trabalho.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro e pela concessão da bolsa de estudos, que foram fundamentais para a realização do doutorado.

A todos que, de alguma forma, contribuíram para a realização desta pesquisa, meu sincero agradecimento.

*“A dívida não é uma condição agradável, mas a certeza é absurda.”*

(Voltaire)

## RESUMO

BATISTA, Tays Silva, D.Sc., Universidade Federal de Viçosa, maio de 2024. **Biometria de dados ambientais aplicada na modelagem e caracterização de mega-ambientes no melhoramento genético da cana-de-açúcar.** Orientador: Luiz Alexandre Peternelli.

A identificação de clones com alto potencial produtivo e estabilidade, representa um desafio para os melhoristas diante da complexa interação genótipo  $\times$  ambiente ( $G \times A$ ), onde genótipos produtivos em um ambiente podem não ser em outros. A incorporação de informações ambientais em estudos com cana-de-açúcar é crucial para aumentar a precisão da avaliação dos clones. Ignorar os efeitos do ambiente pode levar a interpretações equivocadas e à exclusão de genótipos com potencial adaptativo. A expressão fenotípica é influenciada pela constituição genética e pelos fatores ambientais como tipo de solo, posição geográfica e fatores climáticos variáveis. Esses fatores podem ser utilizados para agrupar ambientes semelhantes, conhecidos como mega-ambientes. O agrupamento de ambientes com base em limitações geográficas pode ser insuficiente para orientar estratégias de melhoramento genético. Uma abordagem alternativa define mega-ambientes com base nos genótipos, identificando áreas onde eles mostram respostas semelhantes em diferentes ambientes. No entanto, confiar exclusivamente nos genótipos para formar mega-ambientes pode ser problemático, pois os métodos são sensíveis à remoção de genótipos, alterando a configuração dos agrupamentos. Para melhorar a precisão na seleção de genótipos e nas previsões fenotípicas em locais não testados, são fundamentais o procedimento *Envirotyping*, que coleta, armazena e processa dados ambientais, e a abordagem *Enviroomics*, que obtém dados em larga escala. Modelos como os Modelos Mistos e Modelo Aditivo Generalizado são úteis em compreender a relação entre dados ambientais e produtividade, permitindo prever e explorar o comportamento dos genótipos com maior precisão. Além disso, outro fator que afeta a qualidade das previsões é a falta de dados fenotípicos, que pode levar a estimativas tendenciosas. Para contornar isso, a simulação de dados fenotípicos amplia a população de treinamento com novos indivíduos. Este estudo buscou comparar a formação de mega-ambientes com base em características fenotípicas e em variáveis climáticas, avaliando se ambas as abordagens resultariam em grupos de ambientes consistentes. Adicionalmente, investigou-se como a inclusão de covariáveis ambientais em modelos bayesianos e o aumento do número de genótipos na simulação influenciariam na precisão das previsões de produtividade da cana-de-açúcar em locais não testados. Para isso, este estudo foi estruturado em três capítulos para explorar diferentes aspectos da formação de mega-ambientes e

aprimorar a predição da produtividade da cana-de-açúcar em ambientes não testados. O primeiro capítulo investigou como a exclusão de genótipos específicos afeta a formação de mega-ambientes e examinou a correspondência desses mega-ambientes com os locais recomendados para futuros experimentos de melhoramento da cultura. No segundo capítulo, foi realizada uma análise da formação de mega-ambientes usando variáveis climáticas, avaliando a consistência desses grupos ao longo dos estádios fenológicos da cultura. Adicionalmente, identificaram-se as variáveis ambientais mais relevantes na performance dos clones de cana-de-açúcar por meio do Modelo Aditivo Generalizado (GAM) em diferentes grupos de clones (estáveis, instáveis e de alta produtividade). Por fim, o terceiro capítulo abordou o impacto da inclusão de covariáveis ambientais em modelos mistos bayesianos e investigou se o aumento do número de genótipos na simulação reduziria os erros de predição da produtividade.

Palavras-chave: Covariáveis; Remoção de Clones; Interação  $G \times A$ ; GGE biplot; dados climáticos.

## ABSTRACT

BATISTA, Tays Silva, D.Sc., Universidade Federal de Viçosa, May, 2024. **Biometry of environmental data applied in the modeling and characterization of mega-environments in the genetic improvement of sugarcane.** Adviser: Luiz Alexandre Peternelli.

Identifying clones with high productivity and stability potential is a challenge for breeders given the complex genotype  $\times$  environment (G $\times$ E) interaction, where productive genotypes in one environment may not be productive in others. Incorporating environmental information into sugarcane studies is crucial to increase the accuracy of clone assessment. Ignoring environmental effects can lead to misinterpretations and the exclusion of genotypes with adaptive potential. Phenotypic expression is influenced by genetic composition and environmental factors, such as soil type, geographic position and variable climatic factors. These factors can be used to group similar environments together, known as megaenvironments. Grouping environments based on geographic limitations may be insufficient to guide genetic improvement strategies. An alternative approach defines megaenvironments based on genotypes, identifying areas where they show similar responses in different environments. However, relying solely on genotypes for the formation of megaenvironments can be problematic due to sensitivity to removing genotypes through changing group configurations. To improve the accuracy of genotype selection and phenotypic predictions in untested sites, Envirotyping procedures, which collect, store and process environmental data, and Enviromics approaches, which obtain large-scale data, are crucial. Models such as Mixed Models and Generalized Additive Models are useful for understanding the relationship between environmental data and productivity, allowing for more accurate prediction and exploration of genotypic behavior. Furthermore, the lack of phenotypic data affects the quality of the prediction, potentially leading to biased estimates. To overcome this, simulating phenotypic data expands the training population with new individuals. This study aimed to compare the formation of megaenvironments based on phenotypic characteristics and climatic variables, evaluating whether both approaches would result in consistent environmental groups. Furthermore, we investigated how including environmental covariates in Bayesian models and increasing the number of genotypes in the simulation would influence the accuracy of sugarcane yield predictions in untested locations. Therefore, this study was structured into three chapters to explore different aspects of megaenvironment formation and improve prediction of sugarcane productivity in

untested environments. The first chapter explored how the exclusion of specific genotypes affects the formation of megaenvironments and examined the alignment of these megaenvironments with recommended locations for future crop improvement experiments. In the second chapter, an analysis of the formation of megaenvironments using climatic variables evaluated the consistency of these groups throughout the phenological stages of the crop. Additionally, the most relevant environmental variables that influence the performance of sugarcane clones were identified through Generalized Additive Models in different groups of clones (stable, unstable and high productivity). Finally, the third chapter addressed the impact of including environmental covariates in Bayesian mixed models and investigated whether increasing the number of genotypes in the simulation would reduce yield prediction errors.

Keywords: Covariates; Clone Removal; G×E Interaction; GGE biplot; Climatic Data.

## SUMÁRIO

CAPÍTULO 1 INTRODUÇÃO GERAL.....	11
1.1 Descrição geral da cultura da cana-de-açúcar.....	11
1.2 Impacto da interação genótipo × ambiente no melhoramento genético da cana-de-açúcar 13	
1.3 Integração de variáveis ambientais em modelos de predição para a cultura da cana-de-açúcar 15	
1.4 Objetivo geral e específico.....	17
1.5 Referência bibliográfica.....	17
CAPÍTULO 2 IMPACT OF THE REMOVAL OF GENOTYPES ON THE FORMATION OF MEGA-ENVIRONMENTS BASED ON GGE BIPLOTS ANALYSIS.....	25
2.1 INTRODUCTION.....	26
2.2 METHODS.....	29
2.2.1 Experiment Description.....	29
2.2.2 Individual and joint analysis of variance.....	30
2.2.3 GGE analysis.....	32
2.3 RESULTS.....	33
2.3.1. Joint And Individual ANOVA.....	33
2.3.2 Analysis of yield mean vs stability.....	35
3.3.3 Mega-Environments Based On GGE Biplot Analysis.....	38
2.4 DISCUSSION.....	43
2.5 CONCLUSION.....	45
2.6 REFERENCE.....	46
CAPÍTULO 3 INTEGRAÇÃO DE DADOS CLIMÁTICOS NA FORMAÇÃO DE MEGA-AMBIENTES E NA PERFORMANCE DOS CLONES DE CANA-DE-AÇÚCAR .....	51
3.1 INTRODUÇÃO.....	51
3.2 MATERIAL E MÉTODOS.....	55
3.2.1 Descrição dos dados e da área experimental.....	55

3.2.2	Análise de componentes principais e agrupamento hierárquico de Ward para obter a formação de mega-ambientes.....	58
3.2.3	Contribuição das variáveis climáticas na performance do clone.....	59
3.3	RESULTADOS.....	60
3.3.1	Formação de mega-ambientes com o uso de variáveis climáticas.....	60
3.3.2	Análise climática global e seu impacto no rendimento de TCH por localidade.....	63
3.3.3	Influência de variáveis climáticas na adaptação de clones de cana-de-açúcar.....	65
3.4	DISCUSSÃO.....	68
3.5	CONCLUSÃO.....	71
3.6	REFERÊNCIA BIBLIOGRÁFICA.....	72
CAPÍTULO 4 PREDIÇÃO AMBIENTÔMICA DA PRODUTIVIDADE DA CANA-DE-AÇÚCAR USANDO KERNELS GAUSSIANOS NÃO LINEARES E DADOS CLIMÁTICOS.....		78
4.1	INTRODUÇÃO.....	79
4.2	MATERIAL E MÉTODOS.....	81
4.2.1	Descrição da área de estudo.....	81
4.2.2	Coleta e ampliação de dados de TCH da cana-de-açúcar.....	82
4.2.3	Coleta de dados climáticos e ecofisiológicos.....	84
4.2.4	Descrição dos modelos mistos com o uso da abordagem bayesiana.....	85
4.2.5	Análise estatística.....	86
4.3	RESULTADOS.....	86
4.4	DISCUSSÃO.....	89
4.5	CONCLUSÃO.....	92
4.6	SUPPLEMENTARY MATERIAL.....	93
4.7	REFERÊNCIA BIBLIOGRÁFICA.....	95
CAPÍTULO 5 CONCLUSÕES GERAIS.....		101

## CAPÍTULO 1 INTRODUÇÃO GERAL

### 1.1 Descrição geral da cultura da cana-de-açúcar

A cana-de-açúcar desempenha um papel crucial na economia agrícola, tanto no Brasil quanto no mundo, devido à sua ampla utilização na produção de açúcar e bioenergia, o que demonstra sua importância na segurança alimentar e a sustentabilidade energética (FAO, 2023). Sua relevante participação na produção mundial de açúcar pode chegar a até 70% (Babu et al., 2022), em virtude de suas características morfológicas e fisiológicas específicas, como sua altura variável de dois a seis metros e seu caule resistente, flexível e fibroso, capaz de armazenar altas concentrações de sacarose nos entrenós (Som-ard et al, 2021). Originária do Sudeste Asiático, Nova Guiné, a cana-de-açúcar é comumente cultivada em regiões tropicais e subtropicais (Rodrigues, 1995; FAO, 2023).

Devido a sua grande importância, a área cultivada de cana-de-açúcar aumentou de cerca de 9 para 26 milhões de hectares no mundo todo ao longo de um período de 57 anos (1961 - 2018). Essa expansão reflete a relevância da cana-de-açúcar na produção agrícola global. O Brasil e a Índia destacam-se como os maiores produtores de cana-de-açúcar, seguidos pela China e Tailândia (FAO, 2021). No Brasil, segundo o boletim da CONAB (Companhia Nacional de Abastecimento) da safra 2023/24, a produção estimada é de 610,1 milhões de toneladas. No entanto, as condições climáticas desfavoráveis durante a safra 2021/22, afetaram a produção atual (CONAB, 2023). Apesar dessas adversidades, na região centro-sul do país, houve uma recuperação notável na produtividade, que foi impulsionada por um clima favorável, o que resultou em um aumento da área colhida em relação à safra anterior. A produtividade estimada é 6,1% superior à obtida na safra anterior, destacando a resiliência do setor diante das condições climáticas desafiadoras para a produção de cana-de-açúcar (CONAB, 2024).

A cana-de-açúcar é uma cultura semiperene, cujo ciclo de crescimento dura aproximadamente de 12 a 18 meses até seu primeiro corte, denominado cana planta. Após essa etapa inicial, a cana-de-açúcar tem a capacidade de rebrotar, e essa nova brotação é chamada de soqueira. Dessa forma, a colheita da cana-de-açúcar é realizada anualmente, por um período de 5 a 7 anos, a depender da região, da variedade e condições de manejo e tecnologias disponíveis na localidade. A colheita da cana soca é realizada até que a cultura

não gere mais benefícios econômicos, momento em que o canavial é renovado (Rudorff et al., 2010; Shukla et al., 2017; Som-ard et al., 2021).

Durante seu ciclo de crescimento, a cana-de-açúcar passa por diferentes estádios de desenvolvimento, denominados estádios fenológicos. Na primeira fase ocorre a brotação e perfilhamento, seguidos pela fase em que a cana se concentra no crescimento em altura. Por fim, na última fase, ocorre a redução do crescimento, momento em que toda a sacarose é acumulada nos colmos, processo conhecido como amadurecimento (Cardozo and Sentelhas 2013; Gurski et al., 2021; Viana et al., 2023).

As condições de crescimento favorável fazem com que grandes quantidades de sacarose sejam acumuladas nos colmos (Cardozo and Sentelhas, 2013). Procedimentos específicos ao longo das fases fenológicas podem resultar em alta produtividades, pois cada fase requer condições singulares para um melhor desempenho. Além das práticas de manejo, disponibilidade de água e nutrientes adequados, também é fundamental considerar as condições climáticas e ecofisiológicas para o desenvolvimento da cultura (Rudorff et al., 2010; Baghdadi et al., 2009; Lisboa et al., 2018).

É fundamental investigar a influência dos fatores ambientais nas plantas, pois a falta ou excesso dessas variáveis pode inibir ou reduzir sua produtividade. Embora alguns desses fatores estejam fora do controle humano (Singh et al., 2019), o conhecimento dessas variáveis capacita o melhorista a lidar com condições adversas para explorar o máximo potencial das plantas, especialmente em relação às variáveis meteorológicas, que são as principais responsáveis pela produtividade (Cardozo and Sentelhas, 2013).

Ao avaliar cada fase de desenvolvimento, é possível identificar fatores prejudiciais ao crescimento das plantas. A literatura já apresenta estudos sobre as condições climáticas ideais exigidas para a cultura da cana-de-açúcar (Teruel et al., 1997; Simões et al., 2009; Rudorff et al., 2010; Cardozo and Sentelhas 2013). Na fase inicial, é necessário que a precipitação varie entre 1.100 e 1.500 mm, com temperatura média entre 32 e 38°C, radiação solar de 18 a 36 MJ/m<sup>2</sup> e umidade relativa de 80 a 85%. Durante a fase em que o crescimento em altura é prioridade, a temperatura e a precipitação adequadas são essenciais para um desenvolvimento satisfatório (Som-ard et al., 2018). Altos níveis de umidade e intensidade luminosa são necessários para aumentar a taxa fotossintética e de respiração. Na fase final, o acúmulo de sacarose é influenciado pela temperatura do ar, precipitação, umidade do solo e radiação

solar. Essas variações climáticas podem resultar em menor acúmulo de sacarose nos colmos e redução da qualidade do suco (Li and Yang, 2015). Portanto, é fundamental que os níveis de umidade relativa sejam moderados, e com um déficit hídrico controlado, uma radiação solar adequada com um fotoperíodo de 10-14 horas e valores de 31-36 MJ/m<sup>2</sup> são fundamentais para o acúmulo de sacarose. Além disso, temperaturas mais baixas (18 a 30°C) contribuem para o processo de maturação adequado da planta.

Entender as condições edafoclimáticas é essencial, uma vez que exercem um papel fundamental no desenvolvimento das culturas. Esse conhecimento não só permite aprimorar as práticas agrícolas, mas também facilita a escolha de variedades mais adequadas, possibilita prever a produtividade e otimiza as diversas etapas do programa de melhoramento genético de plantas (Scarpari et al., 2004; Scarpari et al., 2009; Scarpari et al., 2012; Caetano et al., 2017).

## 1.2 Impacto da interação genótipo × ambiente no melhoramento genético da cana-de-açúcar

Nos programas de melhoramento genético da cana-de-açúcar, para alcançar maiores produtividades, é crucial selecionar variedades que não apenas demonstrem valores ótimos para as características agronômicas, mas também sejam adaptadas às diferentes regiões (Trenberth et al., 2013). O processo de seleção é dividido em fases. Começa-se identificando um banco de germoplasma com floração intensa e sincronizada entre genitores. Após o cruzamento entre os genitores selecionados, obtêm-se as sementes que darão origem a um grande número de progênies (primeira etapa do programa – T1). O material selecionado é clonado por meio de propagação vegetativa e plantados em delineamentos em blocos aumentados para identificar aqueles com maior potencial agronômico (T2). Em seguida, os clones selecionados são cultivados em diferentes condições ambientais (T3 - Fase experimental) para avaliar a interação genótipo por ambiente (G×A) e selecionar o clone que apresenta características desejáveis para atingir objetivos específicos comerciais. Esse clone será chamado de variedade (Cheavegatti-Gianotto et al., 2011).

A G×A existente está associada ao rendimento e estabilidade dos clones cultivados em diferentes localidades. Contudo, clones que apresentam uma boa performance em determinadas localidades podem não ter um bom desempenho em outros locais (Sibiya et al., 2012). Para isso, é essencial compreender o ambiente da cultura, com o uso de métodos que

explorem a  $G \times A$  para otimizar a alocação de produtos capazes de maximizar essa interação. Dessa forma, é possível aumentar a produtividade e atender às demandas futuras do mercado agrícola (Ramburan et al., 2011).

Um dos métodos mais utilizados para investigar a  $G \times A$  com base em características agronômicas, como a produtividade, são o GGE biplot e o AMMI (*additive main effects and multiplicative interaction*). Esses métodos são utilizados frequentemente por melhoristas, pela facilidade em obter diferentes respostas com relação a  $G \times A$  graficamente (Yan, 2001). Por meio de biplots, é possível visualizar os padrões de desempenho dos genótipos e a formação de mega-ambientes.

Mega – ambientes são um grupo de localidades distintas onde as plantas apresentam um comportamento semelhante, ou seja, não há mudança no ordenamento dos genótipos (Gauch and Zobel, 1997). No entanto, embora esse agrupamento de localidades já fosse amplamente utilizado por meio das zonas agroecológicas, uma crítica feita por Yan et al. (2023) é que as zonas agroecológicas não são específicas da cultura. O agrupamento de regiões é baseado em fatores ambientais, normalmente definidos por latitude, longitude, altitude, solo, temperatura, precipitação, entre outros.

A análise dos mega - ambientes na seleção de variedades agrícolas é fundamental para maximizar a produtividade e adaptabilidade das culturas às condições locais. No entanto, é crucial reconhecer que essa abordagem não está isenta de desafios e limitações. Ao depender exclusivamente do genótipo para identificar os mega - ambientes, corre-se o risco de subestimar a complexidade das interações entre as características genéticas das plantas e o ambiente em que são cultivadas (Annicchiarico, 2002). Isso pode resultar em uma seleção de clones que não são as mais adequadas para condições locais específicas e mais produtivas.

É importante destacar que os métodos de seleção utilizados são sensíveis à exclusão de determinados genótipos, podendo alterar significativamente o padrão de agrupamento dos ambientes. Além disso, a qualidade das medidas fenotípicas utilizadas na avaliação dos ambientes é crucial. Erros na coleta ou na mensuração dessas variáveis podem distorcer as estimativas e levar à escolha inadequada de genótipos, comprometendo a eficácia do processo seletivo (Annicchiarico, 2002).

Nesse sentido, é essencial considerar qualquer tipo de estresse, tanto biótico quanto abiótico, ao avaliar o desempenho dos genótipos em diferentes regiões. A compreensão das

interações entre as características genéticas das plantas e os fatores ambientais permite uma seleção mais precisa e eficiente, evitando interpretações equivocadas e a eliminação de genótipos adaptativos diante das mudanças climáticas (Zhao & Li, 2015). Dessa forma, é possível otimizar a seleção de variedades agrícolas que sejam capazes de prosperar em diferentes condições ambientais, contribuindo para a sustentabilidade e resiliência dos sistemas agrícolas.

Assim, as análises da interação  $G \times A$  trazem vantagens significativas para o melhoramento de plantas, pois aumenta a eficiência da seleção e, conseqüentemente, aumenta a herdabilidade. Isso possibilita a otimização a seleção de variedades adaptadas a locais específicos, o que resulta em um aumento da produtividade da região (Yan et al., 2023). Além disso, o uso dessas análises de interação  $G \times A$  permite a identificação de mega-ambientes, onde locais com características ambientais semelhantes podem ser agrupados, inferindo-se que os indivíduos terão comportamentos similares. Dessa forma, em vez de realizar mais experimentos ou coletas custosas de dados fenotípicos, uma alternativa para redução de custos é prever os valores de produtividade em ambientes onde os indivíduos não foram cultivados.

### 1.3 Integração de variáveis ambientais em modelos de predição para a cultura da cana-de-açúcar

A inclusão de variáveis ambientais em modelos de predição nos programas de melhoramento genético da cana-de-açúcar ainda é uma área pouco explorada. A confiabilidade dessas predições, especialmente na fase experimental, permite a seleção precoce de genótipos promissores e a predição de seu desempenho em diferentes condições ambientais. Com o comportamento desses genótipos previamente conhecido, torna-se viável evitar a necessidade de cultivo em um número excessivo de ambientes, reduzindo custos de mão de obra e insumos, acelerando o ganho genético e encurtando o tempo do programa de melhoramento (Neyhart et al., 2022).

Inicialmente, o uso de informações ambientais foi integrado em modelos de regressão linear simples, como a regressão fatorial (Basilevsky, 1981; Romy et al., 2010) e mínimos quadrados parciais (Vargas et al., 1999; Monteverde et al., 2019). Esses modelos de regressão têm sido amplamente utilizados, pois facilitam a convergência do modelo ao lidar com grandes conjuntos de dados em diversos ambientes (Krause et al., 2022). No entanto, modelos mais flexíveis como o Modelo Aditivo Generalizado (GAM) (Hastie & Tibshirani, 1987), têm

demonstrado eficiência tanto em modelos preditivos quanto exploratórios, sendo cruciais para capturar relações complexas entre variáveis explicativas e a produtividade agrícola, especialmente em contextos em que essas relações não são lineares (Heinemann et al., 2022).

Dentre os modelos utilizados, os modelos mistos destacam-se pela capacidade de incorporar matrizes de covariância relacionadas às características fenotípicas, genotípicas e ao ambiente. Esses modelos reconhecem que a resposta às covariáveis ambientais não é linear e permitem uma representação mais fiel da variação observada nos genótipos. Isso é fundamental para melhorar a precisão das previsões e para entender como diferentes fatores ambientais influenciam o desempenho das variedades de cana-de-açúcar em diferentes locais e condições climáticas (Granato et al., 2018).

Com os dados ambientais tornando-se menos onerosos e de fácil acesso, foram propostos alguns modelos para integrar as covariáveis ambientais em modelos de predição (Piepho et al., 1998; Annicchiarico et al., 2006; Jarquín et al. 2014; Saint Pierre et al. 2016; Monteverde et al. 2019 ; Millet et al. 2019). A interação entre os ambientes de teste pode ser modelada com o uso de covariâncias lineares (Jarquín et al., 2014) e métodos não lineares (kernel gaussiano) (Costa-Neto et al., 2021). A integração de dados ambientais tem se mostrado particularmente útil na previsão do desempenho das culturas, permitindo uma melhor compreensão das interações  $G \times A$ . Isso é essencial para o melhoramento de plantas, pois possibilita a identificação de genótipos com maior adaptabilidade e estabilidade em diferentes condições ambientais.

Modelos como o GBLUP bayesiano têm incorporado com sucesso covariáveis ambientais (Jarquín et al., 2014), melhorando a precisão da predição ao considerar tanto os efeitos principais genômicos quanto as interações  $G \times A$  específicas do ambiente. Além disso, abordagens baseadas em *kernel*, demonstraram maior eficácia ao capturar efeitos de interação complexos (Pérez-Rodríguez et al., 2012). Estudos de Souza et al. (2017) confirmam que métodos com *kernels* não lineares superam os kernels lineares em termos de precisão preditiva, especialmente em ambientes com correlações positivas. Esses avanços destacam a importância de utilizar dados ambientais detalhados e metodologias avançadas para otimizar a seleção e o desenvolvimento de novas cultivares. A combinação de dados genotípicos, fenotípicos e ambientais, oferece uma abordagem robusta para explorar e utilizar essas interações de maneira eficiente.

Esses métodos têm sido avaliados em diversos estudos, como o realizado por los Campos et al. (2020) que investigou padrões genotípicos e ambientais em locais de teste, que permitiu a predição do desempenho das cultivares em novos ambientes e sob diferentes condições climáticas. Essa abordagem demonstrou ser capaz de acelerar a avaliação de adaptabilidade e estabilidade dos indivíduos. Além disso, o uso de *kernels* não lineares tem sido explorado (Cuevas et al., 2016; Montesinos-López et al., 2018), oferecendo *insights* sobre o desempenho das culturas em diferentes estádios de desenvolvimento (Costa-Neto et al., 2021). A inclusão de informações ambientais tem contribuído significativamente para aprimorar a capacidade preditiva dos modelos, destacando a importância de considerar esses dados na predição do rendimento das culturas (Jarquín et al., 2014; Acosta-Pech et al., 2017). O desenvolvimento contínuo de modelos que integram dados ambientais, genotípicos e fenotípicos, com a utilização de técnicas avançadas de machine learning e análise estatística, é essencial para enfrentar os desafios agrícolas futuros e garantir a segurança alimentar global.

#### 1.4 Objetivo geral e específico

O objetivo geral deste trabalho foi investigar o impacto da inclusão de variáveis ambientais na formação de mega-ambientes e na predição da produtividade da cana-de-açúcar, e sua importância em um programa de melhoramento genético. Para isso, foram realizados três artigos científicos, cada um com seus objetivos específicos. No primeiro estudo, verificou-se como a formação de grupos de localidades foi afetada pela remoção de clones com diferentes comportamentos na análise GGE biplot. Além disso, investigou-se a definição de mega-ambientes com base no genótipo é consistente na indicação de localidades para futuros experimentos em melhoramento de cana-de-açúcar. O segundo estudo teve como objetivos específicos: (i) analisar a formação de mega-ambientes utilizando variáveis climáticas; (ii) avaliar a consistência dos grupos de ambientes em diferentes estádios fenológicos da cultura; e (iii) determinar quais variáveis climáticas contribuem mais para a compreensão da performance das variedades de cana-de-açúcar em cada grupo de clones (mais instáveis, mais estáveis e melhor classificados), utilizando o modelo GAM. No terceiro estudo, o objetivo foi avaliar se a inclusão de variáveis ambientais resultaria em melhorias no ajuste dos modelos na predição da produtividade da cana-de-açúcar (TCH - Tonelada de Cana por Hectare). Também foi analisado se o aumento no número de genótipos usados na ampliação dos dados resultaria em menores erros de predição.

### 1.5 Referência bibliográfica

Akbarian, S., Xu, C., Wang, W., Ginns, S., & Lim, S. (2022). Sugarcane yields prediction at the row level using a novel cross-validation approach to multi-year multispectral images. *Computers and Electronics in Agriculture*, 198, 107024.

Annicchiarico, P. (2002). Genotype x environment interactions: challenges and opportunities for plant breeding and cultivar recommendations.

Annicchiarico, P., Bellah, F., & Chiari, T. (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.

Acosta-Pech, R., Crossa, J., de Los Campos, G., Teysse re, S., Claustres, B., P rez-Elizalde, S., & P rez-Rodr guez, P. (2017). Genomic models with genotype× environment interaction for predicting hybrid performance: an application in maize hybrids. *Theoretical and Applied Genetics*, 130, 1431-1440.

Babu, D. K. S., Janakiraman, V., Palaniswamy, H., Kasirajan, L., Gomathi, R., & Ramkumar, T. R. (2022). A short review on sugarcane: its domestication, molecular manipulations and future perspectives. *Genetic Resources and Crop Evolution*, 69(8), 2623-2643.

Baghdadi, N., Boyer, N., Todoroff, P., El Hajj, M., & B gu , A. (2009). Potential of SAR sensors Terra SAR-X, ASAR/ENVISAT and PALSAR/ALOS for monitoring sugarcane crops on Reunion Island. *Remote sensing of environment*, 113(8), 1724-1738.

Basilevsky, A. (1981). Factor analysis regression. *Canadian Journal of Statistics*, 9(1), 109-117.

Berding, N., Hurney, A. P., Salter, B., & Bonnett, G. D. (2005). Agronomic impact of sucker development in sugarcane under different environmental conditions. *Field crops research*, 92(2-3), 203-217.

Caetano, J. M., & Casaroli, D. (2017). Sugarcane yield estimation for climatic conditions in the state of Goi s. *Revista Ceres*, 64, 298-306.

- Canata, T. F., Wei, M. C. F., Maldaner, L. F., & Molin, J. P. (2021). Sugarcane yield mapping using high-resolution imagery data and machine learning technique. *Remote Sensing*, 13(2), 232.
- Cardozo, N. P., & Sentelhas, P. C. (2013). Climatic effects on sugarcane ripening under the influence of cultivars and crop age. *Scientia Agricola*, 70, 449-456.
- Cavalier, L. (2008). Nonparametric statistical inverse problems. *Inverse Problems*, 24(3), 034004.
- Chea, C., Saengprachatanarug, K., Posom, J., Wongphati, M., & Taira, E. (2020). Sugar yield parameters and fiber prediction in sugarcane fields using a multispectral camera mounted on a small unmanned aerial system (UAS). *Sugar Tech*, 22, 605-621.
- Cheavegatti-Gianotto, A., De Abreu, H. M. C., Arruda, P., Bespalkok Filho, J. C., Burnquist, W. L., Creste, S. & César Ulian, E. (2011). Sugarcane (*Saccharum X officinarum*): a reference study for the regulation of genetically modified cultivars in Brazil. *Tropical plant biology*, 4, 62-89.
- CONAB - COMPANHIA NACIONAL DE ABASTCIMENTO. Acompanhamento da safra brasileira de cana-de-açúcar safra 2022/23. 4º levantamento, Brasília, DF, v. 10, n. 4 abril 2023.
- CONAB - COMPANHIA NACIONAL DE ABASTCIMENTO. Acompanhamento da safra brasileira de cana-de-açúcar safra 2023/24. 3º levantamento, Brasília, DF, v. 11, n. 3 novembro 2023.
- Costa-Neto, G., Galli, G., Carvalho, H. F., Crossa, J., & Fritsche-Neto, R. (2021). EnvRtype: a software to interplay enviromics and quantitative genomics in agriculture. *G3*, 11(4), jkab040.
- Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., De Los Campos, G., & Varshney, R. K. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends in plant science*, 22(11), 961-975.
- Cuevas, J., Pérez-Elizalde, S., Soberanis, V., Pérez-Rodríguez, P., Gianola, D., & Crossa, J. (2014). Bayesian genomic-enabled prediction as an inverse problem. *G3: Genes, Genomes, Genetics*, 4(10), 1991-2001.

Cuevas, J., Crossa, J., Soberanis, V., Pérez-Elizalde, S., Pérez-Rodríguez, P., De los Campos, G., & Burgueño, J. (2016). Genomic prediction of genotype × environment interaction kernel regression models. *Plant Genome* 9: 1–20.

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.

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

Deomano, E., Jackson, P., Wei, X., Aitken, K., Kota, R., & Pérez-Rodríguez, P. (2020). Genomic prediction of sugar content and cane yield in sugar cane clones in different stages of selection in a breeding program, with and without pedigree information. *Molecular Breeding*, 40, 1-12.

Everingham, Y., Sexton, J., Skocaj, D., & Inman-Bamber, G. (2016). Accurate prediction of sugarcane yield using a random forest algorithm. *Agronomy for sustainable development*, 36, 1-9.

FAO. 2021. World Food and Agriculture - Statistical Yearbook 2021. Rome. <https://doi.org/10.4060/cb4477en>

Food and Agriculture Organization of the United Nations (FAO). Sugarcane. 2023. Available online: <http://www.fao.org/land-water/databases-and-software/crop-information/sugarcane/en> (accessed on 16 April 2024).

Fernandes, J. L., Ebecken, N. F. F., & Esquerdo, J. C. D. M. (2017). Sugarcane yield prediction in Brazil using NDVI time series and neural networks ensemble. *International journal of remote sensing*, 38(16), 4631-4644. Gauch Jr, H. G., & Zobel, R. W. (1997). Identifying mega-environments and targeting genotypes. *Crop science*, 37(2), 311-326.

Granato, I., Cuevas, J., Luna-Vázquez, F., Crossa, J., Montesinos-López, O., Burgueño, J., & Fritsche-Neto, R. (2018). BGGE: a new package for genomic-enabled prediction incorporating genotype × environment interaction models. *G3: Genes, Genomes, Genetics*, 8(9), 3039-3047.

- Gurski, B. C., Moretti de Souza, J. L., Gerstemberger, E., & de Oliveira, R. A. (2020). Water requirements and restrictions to sugarcane in cane plants and ratoon cane cycles in Southern Brazil. *Acta Agronómica*, 69(2), 136-144.
- Hastie, T. & Tibshirani, R. (1987). Generalized additive models: some applications. *Journal of the American Statistical Association*, 82(398), 371-386. <http://doi.org/10.1080/01621459.1987.10478440>.
- Hayes, B. J., Wei, X., Joyce, P., Atkin, F., Deomano, E., Yue, J., ... & Voss-Fels, K. P. (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.
- Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., ... & de los Campos, G. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and applied genetics*, 127, 595-607.
- Li, Y. R., & Yang, L. T. (2015). Sugarcane agriculture and sugar industry in China. *Sugar Tech*, 17(1), 1-8.
- Lisboa, I. P., Cherubin, M. R., Lima, R. P., Cerri, C. C., Satiro, L. S., Wienhold, B. J., ... & Cerri, C. E. (2018). Sugarcane straw removal effects on plant growth and stalk yield. *Industrial Crops and Products*, 111, 794-806.
- Mahadevaiah, C., Appunu, C., Aitken, K., Suresha, G. S., Vignesh, P., Mahadeva Swamy, H. K., ... & Ram, B. (2021). Genomic selection in sugarcane: Current status and future prospects. *Frontiers in Plant Science*, 12, 708233.
- Millet, E. J., Kruijer, W., Coupel-Ledru, A., Alvarez Prado, S., Cabrera-Bosquet, L., Lacube, S., & Tardieu, F. (2019). Genomic prediction of maize yield across European environmental conditions. *Nature genetics*, 51(6), 952-956.
- Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Gianola, D., Hernández-Suárez, C. M., & Martín-Vallejo, J. (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., Gutierrez, L., Blanco, P., Pérez de Vida, F., Rosas, J. E., Bonnacarrère, V., ... & McCouch, S. (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.
- Neyhart, J. L., Silverstein, K. A., & Smith, K. P. (2022). Accurate predictions of barley phenotypes using genomewide markers and environmental covariates. *Crop Science*, 62(5), 1821-1833.
- Nybom, H., & Laci, G. (2021). Recent large-scale genotyping and phenotyping of plant genetic resources of vegetatively propagated crops. *Plants* 10: 415.
- Pérez-Rodríguez, P., Gianola, D., González-Camacho, J. M., Crossa, J., Manès, Y., & Dreisigacker, S. (2012). Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. *G3: Genes| Genomes| Genetics*, 2(12), 1595-1605.
- Piepho, H. P., Denis, J. B., & van Eeuwijk, F. A. (1998). Predicting cultivar differences using covariates. *Journal of Agricultural, Biological, and Environmental Statistics*, 151-162.
- Rahman, M. M., & Robson, A. J. (2016). A novel approach for sugarcane yield prediction using landsat time series imagery: A case study on Bundaberg region. *Advances in Remote Sensing*, 5(2), 93-102.
- Ramburan, S., Zhou, M., & Labuschagne, M. (2011). Interpretation of genotype $\times$  environment interactions of sugarcane: Identifying significant environmental factors. *Field Crops Research*, 124(3), 392-399.
- Rodrigues, J. D. (1995). *Fisiologia da cana-de-açúcar*. Botucatu: Unesp, 419-449.
- Robson, A., Abbott, C., Lamb, D., & Bramley, R. O. B. (2012). Developing sugar cane yield prediction algorithms from satellite imagery.
- Rogers, A. R., Dunne, J. C., Romay, C., Bohn, M., Buckler, E. S., Ciampitti, I. A., ... & Holland, J. B. (2021). The importance of dominance and genotype-by-environment interactions on grain yield variation in a large-scale public cooperative maize experiment. *G3 (Bethesda)* 11: jkaa050.
- Romay, M. C., Malvar, R. A., Campo, L., Alvarez, A., Moreno-González, J., Ordás, A., & Revilla, P. (2010). Climatic and genotypic effects for grain yield in maize under stress conditions. *Crop Science*, 50(1), 51-58.

- Rudorff, B. F. T., Aguiar, D. A., Silva, W. F., Sugawara, L. M., Adami, M., & Moreira, M. A. (2010). Studies on the rapid expansion of sugarcane for ethanol production in São Paulo State (Brazil) using Landsat data. *Remote sensing*, 2(4), 1057-1076.
- Saint Pierre, C., Burgueño, J., Crossa, J., Fuentes Dávila, G., Figueroa López, P., Solís Moya, E., & Singh, S. (2016). Genomic prediction models for grain yield of spring bread wheat in diverse agro-ecological zones. *Scientific reports*, 6(1), 27312.
- Scarpate, F. V., van Lier, O. D. J., Corrêa, S. T. R., Hugo, A., Barros, C., Marin, F. R., & Nassif, D. S. P. (2012). modelos de crescimento da cana-de-açúcar e sua parametrização. *Revista de Agricultura*, 87(1), 66-80.
- Scarpari, M. S., & Beauclair, E. G. F. D. (2004). Sugarcane maturity estimation through edaphic-climatic parameters. *Scientia Agricola*, 61, 486-491.
- Simões, M. D. S., Rocha, J. V., & Lamparelli, R. A. C. (2009). Orbital spectral variables, growth analysis and sugarcane yield. *Scientia Agricola*, 66, 451-461.
- Singh, N., & Parida, B. R. (2019). Environmental factors associated with seasonal variations of night-time plant canopy and soil respiration fluxes in deciduous conifer forest, Western Himalaya, India. *Trees*, 33(2), 599-613.
- Shanmugavadivu, R., & Gururaja Rao, P. N. (2009). A comparison of flowering behaviour of sugarcane clones in two different locations. *Sugar Tech*, 11, 401-404.
- Shi, L., Hu, S., & Zha, Y. (2018). Estimation of sugarcane yield by assimilating UAV and ground measurements via ensemble Kalman filter. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 8816-8819).
- Shendryk, Y., Davy, R., & Thorburn, P. (2021). Integrating satellite imagery and environmental data to predict field-level cane and sugar yields in Australia using machine learning. *Field Crops Research*, 260, 107984.
- Sibiya, J., Tongoona, P., Derera, J., & van Rij, N. (2012). Genetic analysis and genotype × environment ( $G \times E$ ) for grey leaf spot disease resistance in elite African maize (*Zea mays* L.) germplasm. *Euphytica*, 185, 349-362.
- Shukla, S. K., Sharma, L., Awasthi, S. K., & Pathak, A. D. (2017). Sugarcane in India. Package of practices for different agro-climatic zones, All Indian Coordinated Research Project on Sugarcane, IISR Lucknow, Uttar Pradesh, 1-64.

- Sofonia, J., Shendryk, Y., Phinn, S., Roelfsema, C., Kendoul, F., & Skocaj, D. (2019). Monitoring sugarcane growth response to varying nitrogen application rates: A comparison of UAV SLAM LiDAR and photogrammetry. *International Journal of Applied Earth Observation and Geoinformation*, 82, 101878.
- Som-ard, J., Hossain, M. D., Ninsawat, S., & Veerachitt, V. (2018). Pre-harvest sugarcane yield estimation using UAV-based RGB images and ground observation. *Sugar Tech*, 20(6), 645-657.
- Som-Ard, J., Atzberger, C., Izquierdo-Verdiguier, E., Vuolo, F., & Immitzer, M. (2021). Remote sensing applications in sugarcane cultivation: A review. *Remote sensing*, 13(20), 4040.
- Sousa, M. B., Cuevas, J., de Oliveira Couto, E. G., Pérez-Rodríguez, P., Jarquín, D., Fritsch-Neto, R., ... & Crossa, J. (2017). Genomic-Enabled Prediction in Maize Using Kernel Models with Genotype  $\times$  Environment Interaction.
- Teruel, D. A., Barbieri, V., & Ferraro Jr, L. A. (1997). Sugarcane leaf area index modeling under different soil water conditions. *Scientia Agricola*, 54, 39-44.
- Todd, J., Hardev Sandhu, H. S., Hale, A., Glaz, B., & Wang JianPing, W. J. (2017). Phenotypic evaluation of a diversity panel selected from the world collection of sugarcane (*Saccharum spp.*) and related grasses.
- Trenberth, K. E., Dai, A., Van Der Schrier, G., Jones, P. D., Barichivich, J., Briffa, K. R., & Sheffield, J. (2014). Global warming and changes in drought. *Nature Climate Change*, 4(1), 17-22.
- Vargas, M., Crossa, J., van Eeuwijk, F. A., Ramírez, M. E., & Sayre, K. (1999). Using partial least squares regression, factorial regression, and AMMI models for interpreting genotype  $\times$  environment interaction. *Crop science*, 39(4), 955-967.
- Viana, J. L., de Souza, J. L. M., Hoshide, A. K., de Oliveira, R. A., de Abreu, D. C., & da Silva, W. M. (2023). Estimating Sugarcane Yield in a Subtropical Climate Using Climatic Variables and Soil Water Storage. *Sustainability*, 15(5), 4360.
- Waldron, J. C., Glasziou, K. T., & Bull, T. A. (1967). The Physiology of Sugar-Cane. IX. Factors Affecting Photosynthesis and Sugar Storage. *Australian Journal of Biological Sciences*, 20(6), 1043-1052.

Wang, X., Xu, Y., Hu, Z., & Xu, C. (2018). Genomic selection methods for crop improvement: Current status and prospects. *The Crop Journal*, 6(4), 330-340.

Yan, W. (2001). GGEbiplot—A windows application for graphical analysis of multi-environment trial data and other types of two-way data. *Agronomy Journal*, 93(5), 1111–1118. <https://doi.org/10.2134/agronj2001.9351111x>

Yan, W., Nilsen, K. T., & Beattie, A. (2023). Mega-environment analysis and breeding for specific adaptation. *Crop Science*, 63(2), 480-494.

Zhao, D. & Li, Y. R. (2015). Climate change and sugarcane production: potential impact and mitigation strategies. *International Journal of Agronomy*, 2015, 1-10. <http://doi.org/10.1155/2015/547386>.

## **CAPÍTULO 2 IMPACT OF THE REMOVAL OF GENOTYPES ON THE FORMATION OF MEGA-ENVIRONMENTS BASED ON GGE BIPLOTS ANALYSIS**

Accepted by the Journal of Agricultural, Biological and Environmental Statistics (JABES).

Tays Silva Batista, Luiz Alexandre Peternelli, Wagner Faria  
Barbosa, Antônio Policarpo Souza Carneiro, and Felipe Lopes da  
Silva

### **ABSTRACT**

The GGE biplot analysis is an essential technique for optimizing genotype selection. However, it is particularly sensitive as its effectiveness depends on the genotypic effect, which can vary greatly according to the cultivation environment. This study thoroughly explores how the formation of mega-environments can be affected by the exclusion of specific genotypes and whether the categorization of mega-environments based on genotypes aligns with recommended locations for future sugarcane breeding experiments. Eight experiments were conducted in Minas Gerais, Brazil 2021, evaluating 11 sugarcane clones planted in 2019, with data collected in 2021. The exclusion of genotypes impacts the formation of mega-environ-

ments: unstable genotypes exhibit variable performance in different environments, while their exclusion may compromise the interpretation of genotype  $\times$  environment interaction; on the other hand, stable genotypes maintain consistent performance, minimizing the impact on mega-environment formation. The study enhances the understanding of GGE biplot analysis and its application in sugarcane genetic improvement, underscoring the relevance of precise genotype selection in diverse environmental contexts.

**Keywords:** Adaptability, Plant selection, Sugarcane, Stability.

## 2.1 INTRODUCTION

Sugarcane is an important crop grown in Brazil and other tropical and subtropical regions (Zhao and Li, 2015). Its main products are sugar and ethanol, which are used by the local industry and for export (UDOP, 2022). In addition, sugarcane has other uses, such as biogas, biomethane, plastic, and other uses in the chemical industry (CONAB, 2023).

Brazil had a 1.9% decrease in the sugarcane cultivated area compared to the previous harvest. However, the average yield increased by 1.3%, reaching 72,609 Kg/ha (CONAB, 2023). The increase in yield results from plant genetic improvement, which develops varieties resistant to pests and diseases, and adapted to different regions (Bezerra et al., 2018). In addition, improvements in management practices, fertilization, and irrigation also contribute to the increase in production (de Moraes et al., 2015).

Sugarcane improvement programs develop new commercial varieties adapted to different regions. The varieties are grown in various locations to assess their performance in different

environmental conditions. The data collected in this experimental phase identify the most promising varieties, which are then launched on the market (Ramburan et al. 2011).

Genotypes often display distinct behaviors in each region, making it challenging to select superior varieties and resulting in genotype  $\times$  environment interaction (G $\times$ E) (Baker, 1988; Ramburan et al., 2011). To address this, conducting multi-environment trials (METs), where the performance of varieties is assessed under various environmental conditions, becomes crucial to identify superior genotypes specifically adapted to certain environments.

METs can be subdivided, and one notable advantage is the reduction in the number of trial locations, consequently lowering implementation costs and enabling the establishment of specific improvement strategies for each region (Ramburan and Zhou, 2011). This subdivision is known as mega-environments, a concept defined by Gauch and Zobel (1997) as part of the cultivation area characterized by a homogeneous environment, where plants exhibit similar behavior even when cultivated in distinct regions.

However, different terms are used in the literature to define mega-environments or regions with distinct agroecological characteristics (Krause et al., 2022). These designations are used to delimit geographical areas with similar environmental characteristics, such as climate, soil, and crop management, which can affect biotic and abiotic stress and, consequently, plant productivity (Clay and Dejaegher, 1987; Fischer et al., 2000).

However, grouping the environments based on geographical limitations may not be sufficient to guide genetic improvement strategies in the face of climate change (Yan et al., 2022). It is necessary to conduct evaluations for several consecutive years to verify the relationship between the test environments and the response of the genotypes. This approach allows us to assess whether the genotypes maintain stable performance in different times and climatic conditions, providing a more robust and reliable assessment (Yan and Kang, 2003).

Thus, another definition of mega-environments is directly related to the genotype. Within heterogeneous regions, it is possible to identify cultivation areas where the genotypes exhibit similar responses (Gauch and Zobel, 1997). This approach considers the similarity in the response of the genotypes as a criterion for grouping environments, which can help identify patterns and direct cultivation recommendations.

However, using statistical models to investigate the  $G \times E$  interaction in selecting high-yielding, stable, and adaptable genotypes is common. Breeders frequently use the GGE biplots method, as it is easy to graphically obtain different responses concerning the  $G \times E$  interaction (Yan and Kang, 2003). With this, through biplots, it is possible to visualize the patterns of performance of the genotypes and the formation of mega-environments.

Biplots are diagrams that show the position of genotypes and environments in a two-dimensional space (Yan et al., 2007). The distances between the genotypes and environments in the biplots reflect the similarity of their phenotypic responses. Biplots can be used to identify groups of genotypes responsive to similar environments and to identify environments favorable to specific genotypes.

GGE biplot analysis is a statistical technique that can be used to improve the efficiency of genotype selection. It is a relatively simple technique to understand and use and can be applied to various  $G \times E$  data. However, as the analysis is based on the genotype, it is essential to consider the quality and precision of phenotypic measurements when performing these analyses. Errors in collecting or measuring phenotypic variables can lead to biased estimates, resulting in inappropriate choices of genotypes and the identification of mega-environments (Annicchiarico, 2002).

In addition, it is common for some genotypes to be discarded throughout the breeding program. The clones are generally evaluated based on several criteria: productivity, disease resis-

tance, and drought tolerance. If a clone does not meet these criteria, it will be discarded. This result can affect GGE analysis, as the environments formed may differ. Therefore, performing the analysis for several years is expected to verify the consistency of the mega-environments formed (Yan et al. 2022).

Therefore, the objective of this study was to verify how the formation of groups of localities would be affected by removing clones with different types of behaviors in the GGE biplot analysis and if the definition of mega-environments based on the genotype is consistent in the indication of the locations for future experiments in a sugarcane breeding program.

## 2.2 METHODS

### 2.2.1 Experiment Description

The experimental data were obtained from experiments developed in the experimental phase of the sugarcane breeding program at the Universidade Federal de Viçosa (PMGCA-UFV) (Barbosa and da Silveira, 2015). The 11 clones evaluated were planted in 2019 (Table 1). However, only the measurements obtained in 2021 (cane ratoon) from eight experiments located in eight municipalities or localities in Minas Gerais, Brazil (Fig. 1) were used. In all experiments, the experimental plots were four furrows of five meters in length, with a spacing of 1.4 m between rows.

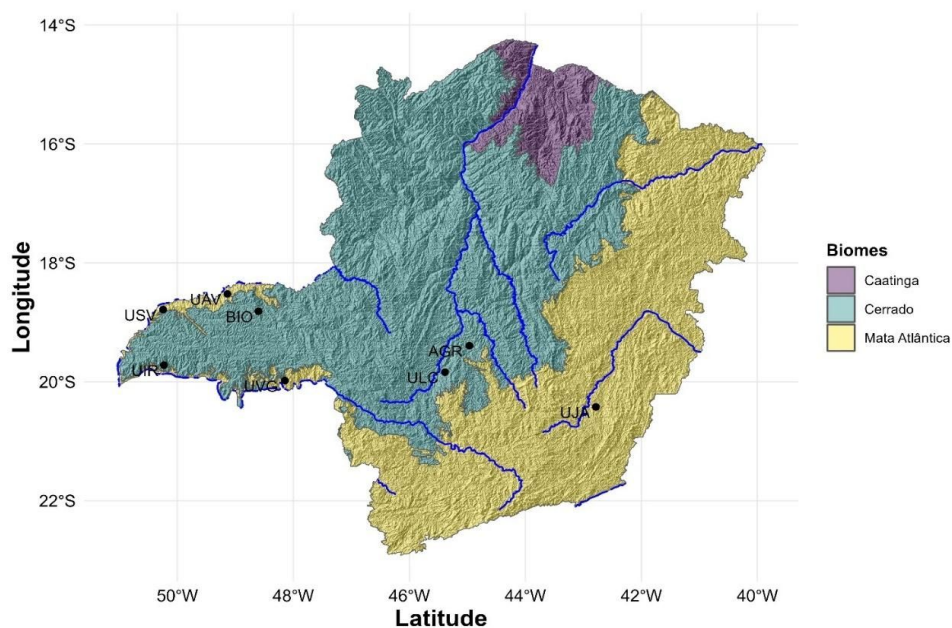


Figure 1 Locations where the experiments were implemented in Minas Gerais state, Brazil.

Table 1 (a) Geographic positions and municipalities of each experimental location in the state of Minas Gerais and (b) coding of clones used in each location.

(a)

Locality	Latitude	Longitude	County	Altitude
ULC	19°50'13.691"S	45°23'5.577"W	Moema	739
UJA	20°25'18.190"S	42°47'0.764"W	Oratórios	524
AGR	19°23'35.510"S	44°57'54.840"W	Pompéu Velho	638
USV	18°47'14.843"S	50°14'36.817"W	Perdilândia	490
BIO	18°48'45.257"S	48°36'4.312"W	Xapetuba	903
UIR	19°43'7.080"S	50°13'43.624"W	Iturama	466
UAV	18°31'20.310"S	49°8'7.069"W	Alceu Garcia	513

UVG	19°58'45.875"S	48°8'43.437"W	Comprida	556	
(b) Varieties (code)					
G1	RB037059	G5	RB107060	G9	RB107277
G2	RB057310	G6	RB107070	G10	RB107414
G3	RB097021	G7	RB107210	G11	RB966928
G4	RB107020	G8	RB107264		

The phenotypic data collected were Recoverable total sugar (RTS), fiber content (FIBER), and tons of stems per hectare (TSH), calculated according to the CONSECANA (2012) manual. The method of collecting these data is defined in the Program's protocol manual.

### 2.2.2 Individual and joint analysis of variance

The experimental design adopted at each locality (environment) was in randomized complete blocks (RCBD) with four replications. Four blocks were used in each environment, with no repetition of the same genotype within a single block. The statistical model considered fixed effects for environments and genotypes and random effects for blocks (Equation 1). This model allows independent evaluation of the effect of each locality and genotype on the variables of interest (Kutner et al. 2004).

$$Y_{ij} = \mu + b_j + g_i + \varepsilon_{ij} \quad (1)$$

Where  $Y_{ij}$  is the observed value for the plot with genotype  $i$  ( $i=1,2,\dots,g$ ) in block  $j$  ( $j=1,2,\dots,b$ );  $\mu$  is the model constant;  $b_j$  is the random effect of block  $j$   $b_j \sim N(0, \sigma_b^2)$ ;  $g_i$  is the fixed effect of genotype  $i$ ; and  $\varepsilon_{ij} \sim N(0, \sigma^2)$  is the random error.

In addition, a joint analysis (Equation 2) was performed. This model allows for combining and analyzing data from all locations, considering the differences between the locations and the environmental effects. Environmental effects are due to the different climatic, soil, and management conditions that affect plant productivity. The genotype is also considered in the model (Baker, 1969).

$$Y_{ijk} = \mu + g_i + e_j + (g \times e)_{ij} + b_k + e_j + \varepsilon_{ijk} \quad (2)$$

Where  $Y_{ij}$  is the mean of genotype  $i$  in environment  $j$  in the  $k^{\text{th}}$  block;  $\mu$  is the model constant;  $g_i$  is the effect of genotype  $i$ ;  $e_j$  is the environment effect  $j$  ( $j=1,2,\dots,e$ );  $(g \times e)_{ij}$  is the interaction between genotype  $i$  and environment  $j$ ;  $\varepsilon_{ijk}$  is the random experimental error, assuming that  $\varepsilon_{ijk} \sim N(0, \sigma^2)$ . The broad-sense heritability for each location was estimated using the formula  $H^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_{g \times e}^2 + \sigma_\varepsilon^2)$ , where  $\sigma_g^2$  is the genotypic variance,  $\sigma_{g \times e}^2$  is the variance of the G×E interaction and  $\sigma_\varepsilon^2$  is the residual variance (Sampaio Filho et al., 2023). The individual and joint analysis of variance and the estimation of heritability were performed using the *metan* package in the R programming environment (Olivoto and Lúcio, 2020).

### 2.2.3 GGE analysis

The previous analyses were functional in verifying the significance of the effects present in the ANOVA model. They assess the overall accuracy of the experiment and genetic parameters. However, through GGE analysis, it is possible to represent the G×E interaction patterns visually.

Various biplots are generated according to the specific objectives of each study (Faheem et al., 2023). In this study, stability versus yield graphs were used to identify, in each trait, the clone with the highest yield and the most unstable and stable. This identification of clones is essential for evaluating the graph called *Who Won Where* (who won where), which allows observing the changes in the formation of mega-environments with the removal of each type of clone.

However, in obtaining the biplots of the GGE analysis, the dataset of phenotypic traits (TCH, FIBER, and ATR) of sugarcane varieties across eight locations underwent singular value decomposition (SVD) to generate the principal components (PC). The method details are described by Yan and Kang (2003); however, the basic model used in this study is  $Y_{ij} - \mu - \beta_j = \lambda_1 \xi_{i1} \eta_{1j} + \lambda_2 \xi_{i2} \eta_{2j} + \varepsilon_{ij}$ . In this equation,  $Y_{ij}$  represents the average yield of genotype  $i$  in  $j$  environments.  $\mu$  stands for the overall mean,  $\beta_j$  is the main effect of environment  $j$ ,  $\lambda_1$  and  $\lambda_2$  are the singular values for PC1 e PC2,  $\xi_{i1}$  and  $\xi_{i2}$  are the eigenvectors of genotype  $i$  for PC1 e PC2,  $\eta_{1j}$  e  $\eta_{2j}$  are the eigenvectors of environment  $j$  for PC1 and PC2, and  $\varepsilon_{ij}$  is the residual associated with genotype  $i$  in environment  $j$ .

This analysis used the `gge` function from the `metan` package in the R programming environment (Olivoto and Lúcio, 2020).

## 2.3 RESULTS

The results are presented in three topics. In the first topic, the results of the variance analysis are presented to highlight significant differences between genotypes and environments, as well as the  $G \times A$  interaction. Coefficients of variation are also discussed to demonstrate the experiment's reliability and heritability, providing an analysis of one of the reasons causing differentiation in forming mega-environments for each characteristic. Before detailing the results, it is crucial to present the criteria for selecting genotypes to be excluded from the GGE

analysis, as presented in topic 3.2. Finally, in topic 3.3, we examine the impact of genotype removal on mega-environment formation.

### 2.3.1. Joint And Individual ANOVA

The individual variance analysis presents the coefficient of variation (CV %) values for each location and phenotypic trait (Table 2). RTS and FIBER variables presented the lowest CV values, generally below 7%, except for the USV location, where the CV was close to 10%. This result indicates a lower dispersion of errors relative to the mean for these variables, which suggests greater experimental precision. On the other hand, the TSH variable presented higher CV values, above 20%, in the AGR and USV locations. For this variable, the CV in the range of 15 to 20% is considered high according to the classification of Couto et al. (2013).

The heritability values for the different phenotypic traits in each location of the experiment are presented in Table 2. It is observed that the RTS and FIBER variables presented heritability values above 60% in all locations. For the TSH variable, heritability varied from 27 to 87%, depending on the location of the experiment. This variation in the values can be attributed to the different environmental conditions in each location, which can influence gene expression and, consequently, phenotypic variation. These results are similar to the studies by Barreto et al. (2019) and Tomaz et al. (2020), which also found variable heritability values depending on the trait under study and the locations of the experiment. These results reinforce the understanding that a trait's heritability is conditioned to the study population and the environment in which the experiment is conducted (Tena et al., 2019).

Table 2 Original means and genetic parameters for the characteristics of RTS (Kg), FIBER (%), and TSH (t/ha) of the 11 clones at each location, evaluated in 2021.

---

ULC	UJA	AGR	USV	BIO	UIR	UAV	UVG
-----	-----	-----	-----	-----	-----	-----	-----

---

	Mean	132.64	127.77	133.88	133.02	162.48	131.02	129.11	147.42
RTS	CV	5.66	8.28	7.15	9.82	3.89	7.12	6.71	4.95
	h <sup>2</sup>	84.95	73.89	72.81	62.98	76.72	85.3	80.55	87.35
	Mean	12.48	14.85	13.11	14	11.66	14.87	13.02	10.96
FIBER	CV	5.39	5.9	4.9	9.3	5.5	4.97	3.59	3.92
	h <sup>2</sup>	72.3	71.44	52.09	57.64	72.48	87.38	92.91	88.89
	Mean	111.08	55.59	80	85.64	199.72	83.68	72.62	136.55
TSH	CV	13.71	18.7	20.61	24.01	11.93	11.51	16.14	18.47
	h <sup>2</sup>	77.63	87.51	60.79	64.41	55.5	56.62	83.16	29.38

CV = coefficient of variation (%); h<sup>2</sup> = broad-sense heritability (%); RTS = recoverable total sugar; TSH = ton of stalk per hectare.

The results of the joint analysis of variance are presented in Table 3. The effect of the environments for all the traits was responsible for the most extensive total variation of the sum of squares, about 70% to 90%, concerning the effect of genotypes and the G×E interaction, as also seen in some studies (Badu-Apraku et al. 2011; Tena et al. 2019; Kumar et al. 2023). Generally, most of the total variation is due to environmental causes, so the importance of evaluating the environment more closely and understanding its behavior in each location and its influence on plant development is emphasized.

The results were significant ( $p < 0.01$ ) for all sources of variation in all traits. Therefore, it is possible to identify superior and divergent genotypes and select individuals according to the Program's purpose. The significance of the G×E interaction effects allows for identifying highly productive genotypes in a specific environment. In each environment, genotypes can show distinct behaviors, which causes changes in ranking. Defining which genotypes are more adapted and stable in a given environment is possible in this case.

Table 3 Summary of the joint analysis of variance for RTS, FIBER, and TSH evaluated in eight locations in 2021, plus the mean and coefficient of variation (CV%) for each variable.

Source of variation	DF	RTS		FIBER		TSH	
		MSE	p-value	MSE	p-value	MSE	p-value
Block/A	24	136.01	4.45E-02	0.49	6.82E-01	706.88	8.55E-04
Genotype (G)	10	2122.69*	7.35E-32	12.1*	2.03E-27	2897.57*	3.98E-13
Environment (E)	7	6195.02*	1.52E-55	88.24*	4.73E-84	93774.92*	2.00E-115
G × E	70	149.84*	1.05E-03	1.09*	2.87E-04	584.73*	2.03E-04
Residue	240	85.73		0.59		308.64	
Mean	-	137.17		13.12		103.11	
CV (%)	-	12.24		12.66		46.8	

\* Significant at 1% probability by the F test.

### 2.3.2 Analysis of yield mean vs stability

The selection of genotypes for removal in the GGE analysis was made by visualizing the stability versus yield mean graph. The average environment coordinate (AEC) is obtained to evaluate these phenotypic responses. The AEC is calculated from the scores of the PC1 and PC2 coordinates of the environments, representing the average yield between the locations (blue line horizontally – Figs. 2, 3, and 4).

The visual distance between the genotypes in the graph indicates their magnitude difference. The line perpendicular to the x-axis, which passes through the origin of the biplots, represents the AEC ordinate. The projection of the genotypes on this ordinate indicates the measure of variability or instability (blue line vertically). The greater the projection along the AEC ordinate, the greater the instability of the genotype (each dotted line refers to the instability measure of a genotype) (Yan et al. 2007).

Based on this analysis, the genotypes that showed the highest yield for the TSH variable were G11, G5, and G6, with G9 being the most unstable (Fig. 2). Regarding the RTS variable,

the clones with the highest averages were G1, G3, G11, and G7, with clone G7 being the most unstable (Fig. 3). In the case of the FIBER variable, the clones did not show significant variations in yield and instability. However, based on the coordinates, it was possible to identify the best classified, emphasizing G10, followed by G2 and G11 (Fig. 4).

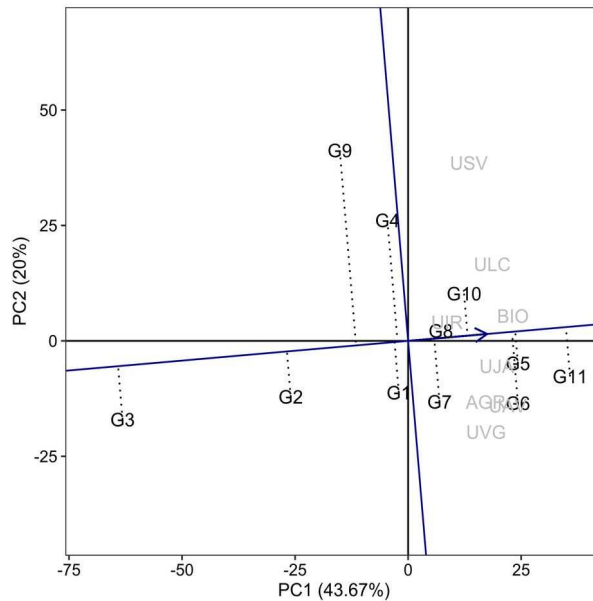


Figure 2 Stability of the average yield of the 11 clones cultivated in eight locations for the TSH characteristic.

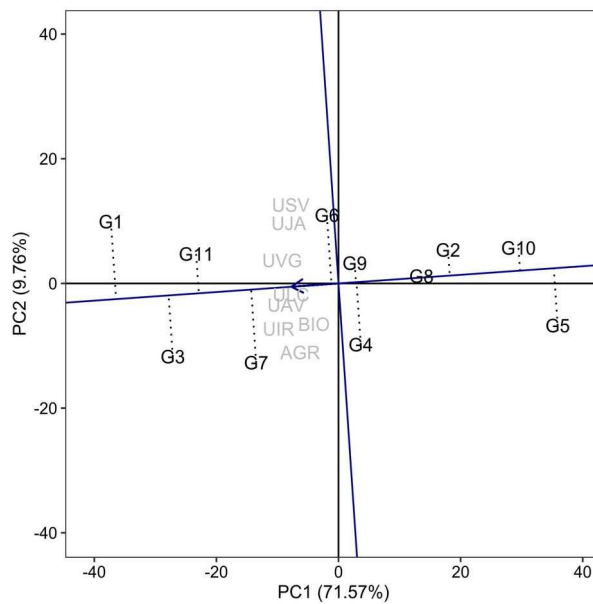


Figure 3 Stability of the average yield of the 11 clones cultivated in eight locations for the RTS characteristic.



without environments and, consequently, without forming groups, the genotypes have lower values than those in those with environments.

In Fig. 5a, it can be observed that G11 performed better in a specific sector for the TSH variable, while the locations UJA, UIR, BIO, AGR, ULC, UAV, and UVG formed a mega-environment, which indicates that these areas share similar characteristics that favor the performance of this clone. On the other hand, the clone G9 presented higher averages in a specific sector but only in one location (USV). This result indicates that clone G9 had a superior performance only in that region, while it may have presented lower results in other locations.

When the clone with the highest average TSH in most locations (G11) was removed, there was a change in the formation of mega-environments (Fig. 5a). Three environmental groups were formed: Group 1 with the locations USV and UIR, Group 2 with ULC, BIO, and UJA, and Group 3 with AGR, UAV, and UVG. When the best-ranked clone is removed, other clones will take the winner's position in specific environments, and these new clones are likely to be unsuccessful in the same environments. For this reason, the separation of the environments will not be the same.

The change in mega-environments' formation also occurred after removing the most unstable clone, coincidentally the clone G9, which was also the winner in one location. Three locations were formed: Group 1 with UJA and UAV, Group 2 with ULC, AGR, and UVG, and Group 3 with UIR, USV, and BIO (Fig. 5b). Its yield varies depending on the cultivation location. Thus, this clone presents different behaviors depending on the environment.

When evaluating only the RTS variable (Fig. 6a), the formation of mega-environments showed different groupings, as expected, but with the same effect on removing specific clones. Initially, with all clones present, two mega-environments were identified. The first

comprised the USV, UJA, UVG, ULC, and UAV locations, while the UIR, BIO, and AGR locations formed the second group.

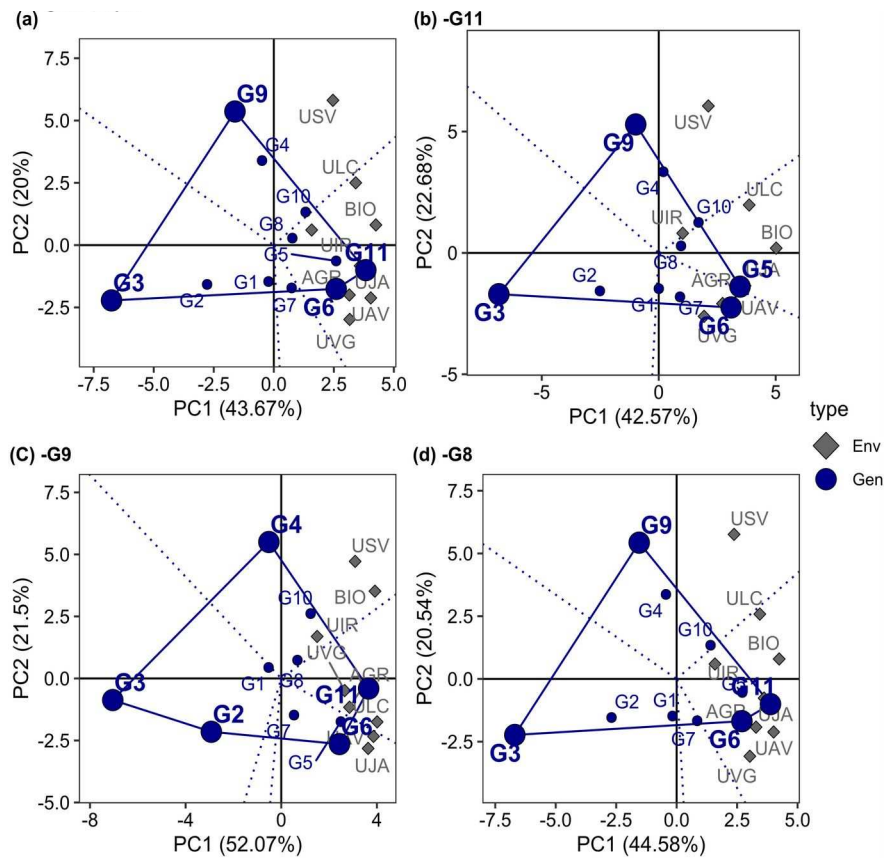


Figure 5 GGE biplot polygon for the TSH variable in sugarcane at the experimental phase, representing the "Who won where" graph. (A) all genotypes present in the GGE biplot analysis; (B) removal of clone G11, best ranked in most locations; (C) removal of the most unstable clone, G9; (D) removal of the least unstable clone, G8.

When clone G1 was removed, the two groups were maintained, but one location changed to the second group (Fig. 6b). Clone G7, the most unstable and the second with the highest yield in group 2, was removed. Only the BIO and UIR locations remained in this sector with

this removal. Similarly, the environmental groups were unaffected when the most stable genotypes were removed.

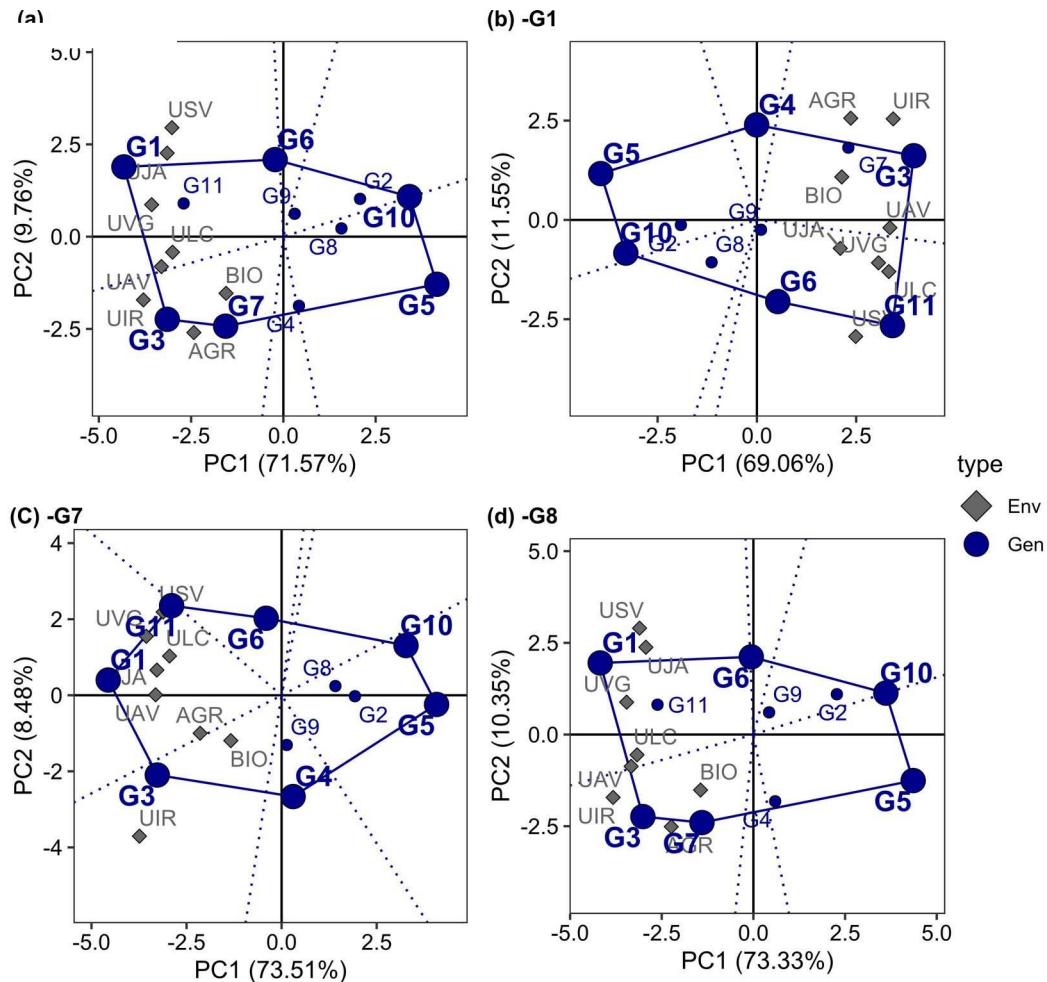


Figure 6 GGE biplot polygon for the RTS variable in sugarcane at the experimental phase, representing the "Who won where" graph. (A) all genotypes present in the GGE biplot analysis; (B) removal of clone G1, best ranked in most locations; (C) removal of the most unstable clone, G7; (D) removal of the least unstable clone, G8.

For the FIBER characteristic, all environments are grouped in a single sector, and removing clones in the analysis had little impact on forming mega-environments. Only by removing the clone with the highest averages in the localities (G10) two groups of environments were formed: Group 1 (UIR and UAV) and Group 2 (BIO, UVG, AGR, UJA, and ULC) (Fig. 7). It

is possible that the other clones responsible for the formation of the mega-environments were more unstable than it.

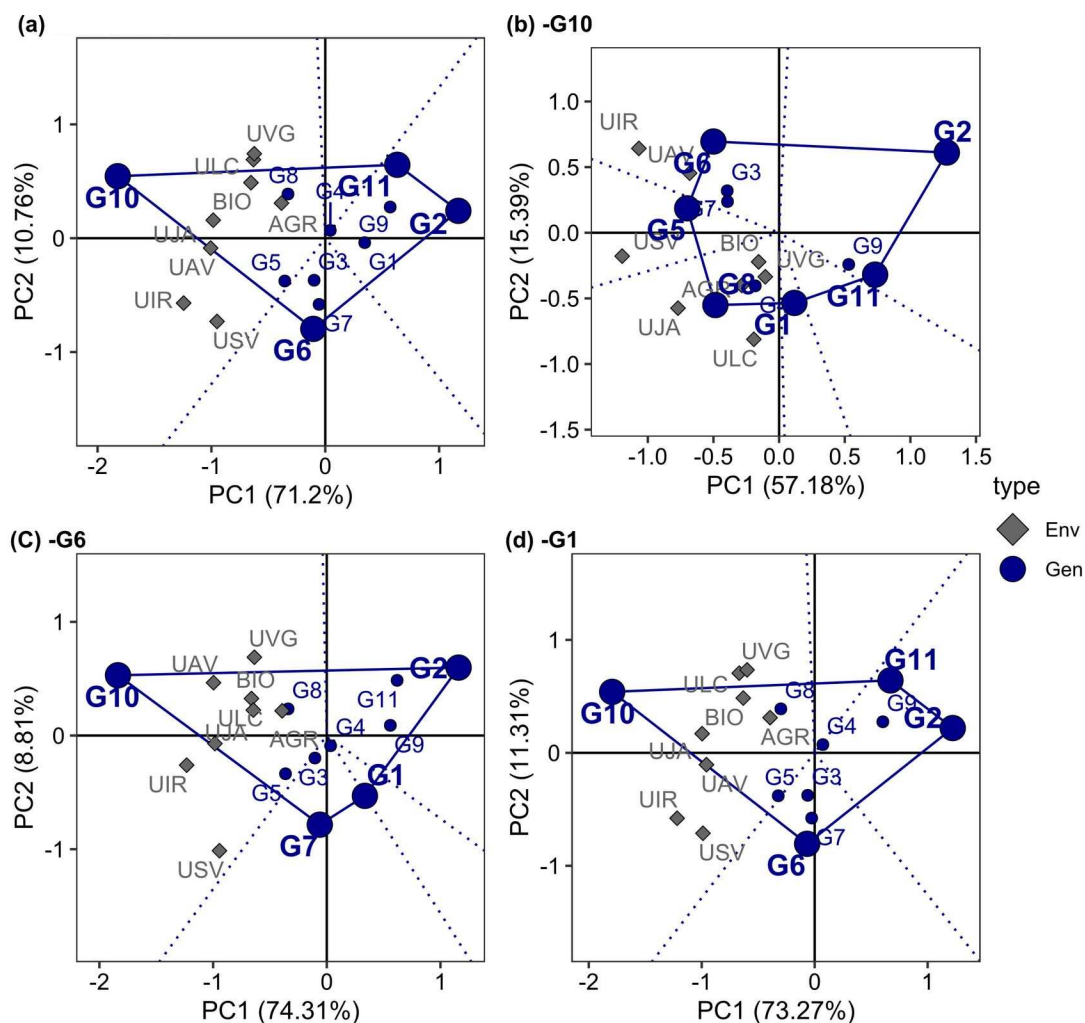


Figure 7 GGE biplot polygon for the FIBER variable in sugarcane at the experimental phase, representing the "Who won where" graph. (A) all genotypes present in the GGE biplot analysis; (B) removal of clone G10, best ranked in most locations; (C) removal of the most unstable clone, G6; (D) removal of the least unstable clone, G1.

The most stable clones in each characteristic were removed from the GGE analysis. Clone G8 was removed in the TSH and RTS characteristics, while clone G1 was removed in the FIBER characteristic. Despite these removals, there was no change in the formation of mega-environments. Even though they were considered, these clones did not show superior performance or specific characteristics that could affect the formation of the locality groups.

## 2.4 DISCUSSION

In this work, removing clones for different traits resulted in distinct mega-environment configurations. One of the reasons for this may be the very nature of the traits being evaluated. TSH is directly related to productivity, i.e., the amount of plant material produced per unit area (Poudyal et al., 2022). Several factors, such as soil fertility, water availability, agronomic management, and the sugarcane varieties' genetics, influence this productivity. Therefore, any change in these variables can directly impact sugarcane productivity (Cardozo and Sentelhas, 2013).

On the other hand, the FIBER trait is related to the quantity and quality of fibers present in the plant, which are influenced by genetic and environmental factors. However, these influences may have a less significant effect on the formation of mega-environments, as fiber is more related to the composition and internal structure of the plant (Marques et al. 2021), while TSH is related to productivity. RTS, on the other hand, measures the amount of sugar in sugarcane, which includes reducing sugars and sucrose (CONSECANA, 2012), and is more closely linked to the chemical composition of the plant (Cardozo and Sentelhas, 2013). Therefore, removing clones for different traits can form distinct mega-environment configurations, as different factors influence these traits.

The formation of mega-environments can be done for different traits, depending on the purpose of the genetic improvement program. As the measurement of these variables is expensive, the collection can be carried out in some environments by defining the groups of environments, reducing the cost and time of this stage of the Program. In addition, by evaluating the genetic parameters by mega-environments, possibly the areas are more homogeneous,

there may be a reduction in  $G \times E$  interaction, and heritability may increase. By altering heritability, the analysis can lead to the formation of mega-environments that are more or less predictive of the performance of genotypes in new environments.

However, for the breeder to have these advantages, the mega-environments must be consistent over the years. This result means that the groups of environments must be defined to reflect the climatic and soil characteristics of the different regions. In addition, the genotypes evaluated in different mega-environments must represent the population's genetic diversity.

Considering that in a breeding program, some genotypes are discarded, this formation of groups of environments may vary, as the GGE biplots analysis is sensitive to the genotype. This result means that the removal of some genotypes can alter the structure of the mega-environments, which can affect the accuracy of the analysis. Considering these limitations when using the GGE biplots analysis to form mega-environments is essential. However, this tool can be a valuable ally for breeders to improve the efficiency of breeding programs.

Another reason for the change in mega-environment formation is the type of genotype to be removed from the analysis. Unstable clones are more susceptible to environmental variations, and their performance can vary considerably in different environments (Azevedo et al., 2018). Removing unstable and high-yielding genotypes from the GGE analysis excludes essential information about the genotype-environment interaction, resulting in changes in the distances between the points that represent the environments in the biplots (Yan et al., 2007) and, consequently, alter the formation of the groups of environments.

On the other hand, stable genotypes show consistent performance in different environments (Mohammadi and Amri, 2008), i.e., they are not significantly affected by environmental variations. Therefore, removing a stable genotype is unlikely to cause significant changes

in the group environments, as these genotypes have relatively consistent and predictable responses.

The data quality can influence the accuracy of the GGE analysis results. Therefore, selecting genotypes representative of the population's genetic diversity and growing in various environments is crucial. The inclusion or exclusion of genotypes in a specific location can affect the formation of mega-environments, depending on the performance of those genotypes.

The GGE biplots graph may not be the best option for visualizing the grouping of environments, as it does not provide detailed information about the environments due to the sources of variation present in the analysis, such as genotype and G×E interaction. In future studies, to obtain a more reliable evaluation of the GGE analysis, it is possible to include information about the environments, such as climate, soil, and management.

This statement is because the characteristics of a genotype can vary depending on the environment in which it is grown. By including information about the environments in the GGE analysis, it is possible to identify the more adaptable genotypes to different environmental conditions. This information can be helpful for plant breeders, who can use it to develop new varieties of plants that are more resistant to climate change and soil conditions.

By including information about the environments in the GGE analysis, it is possible to obtain a more reliable evaluation of the genotypes' characteristics and select the more suitable genotypes for different environmental conditions. As a result, it can help plant breeders develop new varieties of more productive, resistant, and adaptable plants.

## 2.5 CONCLUSION

The removal of genotypes in the GGE biplots analysis affects mega-environments' formation. Unstable genotypes, sensitive to environmental variations, have variable performance in different environments. By removing these genotypes, meaningful information about the

genotype-environment interaction is lost, impacting the formation of the groups of environments.

On the other hand, stable genotypes have a minor effect on forming mega-environments when removed, as they show consistent performance in different conditions. The definition of mega-environments based solely on the varieties' performance seems inappropriate and may generate errors in the case of recommendations for the allocation of experiments for future work in a breeding program.

## 2.6 REFERENCE

- Annicchiarico P (2002) Genotype x environment interactions: challenges and opportunities for plant breeding and cultivar recommendations.
- Azevedo PLD, Teodoro PE, Silva LA, Rodrigues EV, Laviola BG, Bhering LL (2018) Jatropha half-sib family selection with high adaptability and genotypic stability. *Plos One*, 13(7): e0199880. <https://doi.org/10.1371/journal.pone.0199880>
- Badu-Apraku B, Akinwale RO, Ajala SO, Menkir A, Fakorede MAB, Oyekunle M (2011) Relationships among traits of tropical early maize cultivars in contrasting environments. *Agronomy journal*, 103(3): 717–729. <https://doi.org/10.2134/agronj2010.0484>
- Baker, R. J (1969) Genotype-environment interactions in yield of wheat. *Canadian Journal of Plant Science*, 49(6), 743-751. <https://doi.org/10.4141/cjps69-127>.
- Baker, R. J. (1988). Tests for crossover genotype-environmental interactions. *Canadian Journal of Plant Science*, 68(2), 405-410. <http://doi.org/10.4141/cjps88-051>.
- Barbosa MHP, da Silveira LCI (2015) Breeding program and cultivar recommendations. In: Sugarcane Academic Press, pp. 241-255.

- Bezerra JDC, Ferreira GDG, de Oliveira MW, Campos JMS, de Andrade AP, do Nascimento Júnior JRS (2018) Cana-de-açúcar: Melhoramento genético e suas finalidades forrageiras. *Nucleus Animalium*, 10(2): 131-147.
- Cardozo NP, Sentelhas PC (2013) Climatic effects on sugarcane ripening under the influence of cultivars and crop age. *Scientia Agricola*, 70: 449-456. <http://doi.org/10.1590/S0103-90162013000600011>.
- CONAB. Companhia Nacional de Abastecimento. Disponível em: [http://www.conab.gov.br/conabweb/download/safra/1graos\\_08.09.pdf](http://www.conab.gov.br/conabweb/download/safra/1graos_08.09.pdf). Accessed on: January 15, 2023.
- Conselho dos Produtores de Cana-de-Açúcar, Açúcar e Álcool do Estado do Paraná - CONSECANA (2012) Manual de Instruções. 3. ed. Curitiba-PR. 118p.
- Clay DC, Dejaegher YMJ (1987) Agroecological zones: the development of a regional classification. *Tropicultura*, 5(4): 153-159.
- de Moraes LK, de Aguiar MS, de Albuquerque SP, Câmara TMM, Cursi DE, Júnior ARF, Bessalho Filho JC (2015) Breeding of sugarcane. *Industrial crops: breeding for bioenergy and bioproducts*, 29-42.
- Faheem M, Sial MA, Arain S, Laghari KA (2023) Comparison of yield performance of wheat genotypes over environments by gge-biplots analysis. *Journal of Animal and Plant Sciences*, 33(1): 85–94. <http://doi.org/10.36899/JAPS.2023.1.0597>.
- Fischer G, van Velthuisen HT, Nachtergaele FO (2000) Global agroecological zones assessment: methodology and results.

- Gauch Jr HG, Zobel, RW (1997) Identifying mega-environments and targeting genotypes. *Crop science*, 37(2): 311-326. <https://doi.org/10.2135/crop-sci1997.0011183X003700020002x>
- Kumar R, Dhansu P, Kulshreshtha N, Meena MR, Kumaraswamy MH, Appunu C, Pandey SK (2023) Identification of Salinity Tolerant Stable Sugarcane Cultivars Using AMMI, GGE and Some Other Stability Parameters Under Multi Environments of Salinity Stress. *Sustainability*, 15(2): 1119. <https://doi.org/10.3390/su15021119>
- Kutner M, Nachtsheim C, Neter J (2004). *Applied Linear Regression Models*.
- Krause MD, Dias KO, Singh AK, Beavis WD (2022) Using large soybean historical data to study genotype by environment variation and identify mega-environments with the integration of genetic and non-genetic factors. *bioRxiv*, 2022-04. <https://doi.org/10.1101/2022.04.11.487885>
- Marques, J. P. R., Aferrri, G., Montanha, G. S., Guedes, F. T. P., Soares, M. M., Muniz, L. F., Tomazello Filho, M., Xavier, M. A., de Carvalho, H. W. P. (2021). Sugarcane as a forage plant: Structural and chemical traits that affect fiber quality. *BioResources*, 16(4), 7623.
- Mohammadi R, Amri A (2008) Comparison of parametric and non-parametric methods for selecting stable and adapted durum wheat genotypes in variable environments. *Euphytica*, 159: 419-432.
- Olivoto T, Lúcio AD (2020) Metan: An R package for multi-environment trial analysis. *Methods in Ecology and Evolution*, 11(6): 783-789.
- Poudyal C, Costa LF, Sandhu H, Ampatzidis Y, Odero DC, Arbelo OC, Cherry RH (2022) Sugarcane yield prediction and genotype selection using unmanned aerial vehicle-based

hyperspectral imaging and machine learning. *Agronomy Journal*, 114(4): 2320-2333.  
<https://doi.org/10.1002/agj2.21133>

R CORE TEAM. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2019.

Ramburan S, Zhou M, Labuschagne, M (2011) Interpretation of genotype  $\times$  environment interactions of sugarcane: Identifying significant environmental factors. *Field Crops Research*, 124(3): 392-399. <https://doi.org/10.1016/j.fcr.2011.07.008>

Sampaio Filho, J. S., Olivoto, T., Campos, M. D. S., & Oliveira, E. J. (2023). Multi-trait selection in multi-environments for performance and stability in cassava genotypes. *Frontiers in Plant Science*, 14, 1282221. <https://doi.org/10.3389/fpls.2023.1282221>

Tena E, Goshu F, Mohamad H, Tesfa M, Tesfaye D, Seife A (2019) Genotype  $\times$  environment interaction by AMMI and GGE-biplots analysis for sugar yield in three crop cycles of sugarcane (*Saccharum officinarum* L.) clones in Ethiopia. *Cogent Food and Agriculture*, 5(1): 1651925. <http://doi.org/10.1080/23311932.2019.1651925>.

UDOP. União Nacional da Bioenergia. Disponível em: < [https://www.udop.com.br/index.php?item=unidadesandcn=amandid\\_pais=1#topo](https://www.udop.com.br/index.php?item=unidadesandcn=amandid_pais=1#topo)>. Accessed on: January 15, 2023.

Yan W, Kang MS (2003) GGE biplots analysis: a graphical tool for breeders, geneticists, and agronomists. CRC Press.

Yan W, Kang MS, Ma B, Woods S, Cornelius PL (2007) GGE biplots vs. AMMI analysis of genotype-by-environment data. *Crop Science*, 47(2): 643-653. <https://doi.org/10.2135/cropsci2006.06.0374>

- Yan W, Frégeau-Reid J, Pageau D, Martin R, Mitchell-Fetch J, Etienne M, Sparry E (2010) Identifying essential test locations for oat breeding in Eastern, Canada. *Crop Science*, 50(2): 504-515. <https://doi.org/10.2135/cropsci2009.03.0133>
- Yan W, Nilsen KT, Beattie A (2022) Mega-environment analysis and breeding for specific adaptation. *Crop Science*, 63(2): 480-494. <https://doi.org/10.1002/csc2.20895>
- Zhao D, Li YR (2015) Climate change and sugarcane production: potential impact and mitigation strategies. *International Journal of Agronomy*, 2015: 1-10. <https://doi.org/10.1155/2015/547386>

**Declarations****Conflict of interest**

The authors declare that they have no conflict of interest to declare.

**Data availability**

The current study's datasets generated and/or analyzed are available at <https://github.com/Tayssb/GGEbiplot.git>.

### **CAPÍTULO 3 INTEGRAÇÃO DE DADOS CLIMÁTICOS NA FORMAÇÃO DE MEGA-AMBIENTES E NA PERFORMANCE DOS CLONES DE CANA-DE-AÇÚCAR**

#### **RESUMO**

A seleção de clones enfrenta desafios devido à interação genótipo  $\times$  ambiente, a falta de representatividade dos ambientes pode comprometer a escolha dos clones. A integração de dados ambientais pode ajudar a entender as respostas fisiológicas dos genótipos frente à variação dos padrões ambientais, permitindo a análise dos materiais em termos de normas de reação. Este estudo visa avaliar a formação de mega-ambientes usando variáveis climáticas, analisar a consistência desses grupos ao longo dos estádios fenológicos da cultura e identificar as variáveis ambientais mais relevantes na performance dos clones de cana-de-açúcar, por meio do Modelo Aditivo Generalizado (GAM) em cada grupo de clones (estáveis, instáveis e de alta produtividade). A relação entre os dados climáticos e produtividade é complexa, porém, o GAM é uma ferramenta flexível, capaz de ajustar funções para cada variável climática. Para cada localidade, foram coletados dados ambientais, submetidos a uma análise de agrupamento, visando identificar padrões de similaridade ambiental utilizados como mega-ambientes na recomendação de clones. Além disso, o modelo GAM foi aplicado aos dados de TCH (Tonelada de Cana por Hectare) de onze genótipos. Foram identificados padrões climáticos e correlações entre as variáveis em diferentes localidades, o que permitiu a identificação de mega-ambientes. A consistência desses mega-ambientes foi estável ao longo do estágio fenológico dos genótipos, com fatores ambientais influenciando sua formação. As variáveis climáticas afetaram o desempenho dos clones de cana-de-açúcar de forma diversa, sendo a temperatura e umidade relativa do ar os principais contribuintes para o ajuste do modelo. Identificar as variáveis climáticas mais influentes no desempenho dos clones é essencial para aprimorar estratégias de seleção e manejo das culturas.

Palavras-chaves: Modelo Aditivo Generalizado. Interação  $G \times A$ . Similaridade ambiental. Variáveis ambientais.

### 3.1 INTRODUÇÃO

Na fase experimental de um programa de melhoramento genético da cana-de-açúcar, os clones selecionados em etapas anteriores são cultivados em diferentes localidades, com cortes realizados por até três anos consecutivos (Rudorff et al., 2010). Dessa forma, os genótipos podem apresentar comportamentos distintos em cada local de cultivo, o que torna o processo de seleção dos melhores clones mais complexo devido à interação entre genótipo e ambiente ( $G \times A$ ).

A identificação do tipo de interação  $G \times A$  (*crossover* – em que o ranqueamento dos genótipos varia entre os ambientes, e *non-crossover* – onde os genótipos mantêm suas posições no ranqueamento, mas a magnitude das diferenças no desempenho varia entre os ambientes) é crucial para compreender o comportamento dos clones. Esse conhecimento permite a seleção de genótipos com adaptabilidade geral ou específica e a avaliação da representatividade dos ambientes de teste. Além disso, possibilita a escolha de locais que melhor capturam as diferenças genotípicas e a determinação do número de locais com diferentes condições de cultivo necessários para uma avaliação precisa (Fox et al., 1997).

No entanto, a falta de representatividade nos ambientes de seleção pode comprometer a escolha dos clones promissores ao final do programa, devido à falta de conhecimento sobre os fatores ambientais responsáveis pela interação  $G \times A$  (Ramburan et al., 2011). Portanto, é fundamental que os programas de melhoramento integrem uma ampla gama de ambientes de seleção que capturem a diversidade de condições climáticas e edáficas. Isso permitirá uma avaliação mais precisa e confiável dos genótipos, e garantirá que os clones selecionados possuam não apenas alto potencial produtivo, mas também estabilidade e adaptabilidade às variações ambientais.

A adaptabilidade é definida como a capacidade do genótipo de explorar o ambiente de cultivo, mesmo em condições adversas ao seu desenvolvimento, e pode ser classificada como geral ou específica. Quando o objetivo do programa de melhoramento é obter genótipos com adaptabilidade geral e alta estabilidade, é necessário que eles apresentem bom desempenho na maioria das localidades. Por outro lado, selecionar genótipos com adaptabilidade específica

sugere a intenção de obter genótipos que tenham bom desempenho em uma região de cultivo particular (Annicchiarico, 2002; Borém et al., 2013).

A uniformidade entre os locais de teste pode distorcer as avaliações e não refletir a verdadeira adaptabilidade às diferentes condições ambientais. Isso inclui o desempenho dos genótipos ao longo das colheitas. Genótipos podem apresentar comportamentos semelhantes em diferentes ambientes não por suas características intrínsecas de estabilidade ou adaptabilidade, mas devido à similaridade das condições edafoclimáticas dos ambientes. Por outro lado, alguns ambientes de teste apresentam variações nas condições edafoclimáticas (Landell et al., 2003), o que também torna a seleção dos clones mais complexa. Nesse contexto, no Brasil, foi desenvolvido um sistema de classificação de ambientes de produção de cana-de-açúcar. Esse sistema leva em consideração as características físico-químicas dos solos e as épocas de corte, considerando especialmente o tempo de maturação da cana-de-açúcar (Joaquim, 1997).

O período de colheita da cana-de-açúcar é variado, geralmente ocorrendo entre os meses de abril a novembro (Landell e Bressiani, 2008). Durante esse período, ocorrem variações climáticas que, combinadas com as características do solo, podem influenciar significativamente o potencial de produção de um determinado local, modificando assim o ambiente de produção. Além disso, as condições climáticas afetam diretamente o cultivo dos genótipos, pois cada estágio fenológico requer condições climáticas específicas para um desenvolvimento adequado.

A cana-de-açúcar é uma cultura semiperene, com seu ciclo que varia de um a um ano e meio, conforme as características do clone e as condições de cultivo (Rudorff et al., 2010). Seu desenvolvimento passa por três estádios distintos: o estágio I, que compreende a brotação e o perfilhamento; o estágio II, marcado pelo crescimento máximo em altura; e, por fim, o estágio III, caracterizado pela redução do crescimento e pelo acúmulo de sacarose (Viana et al., 2023).

As pesquisas sobre a influência das variáveis climáticas no cultivo de cana-de-açúcar oferecem uma contribuição valiosa para a compreensão do comportamento dos clones em diferentes ambientes. Por exemplo, Ramburan et al. (2011) analisaram o impacto das variáveis climáticas na interação  $G \times A$ , enquanto Pipitpukdee et al. (2020) investigaram o efeito das mudanças climáticas na produtividade. Viana et al. (2023) utilizaram dados

climáticos para prever a produtividade da cana-de-açúcar. Além disso, Cardozo & Sentelhas (2013) examinaram a influência das condições climáticas na maturação da cana-de-açúcar, considerando fatores como o genótipo e a idade da cultura, e concluíram que a complexa combinação de variáveis climáticas afeta a maturação da cana-de-açúcar. Essas pesquisas em conjunto fornecem uma visão ampla sobre como as condições climáticas afetam a cultura da cana-de-açúcar e destacam a importância da integração desses aspectos no desenvolvimento de estratégias para aprimorar o manejo agrícola da cultura.

A incorporação de informações climáticas em estudos com cana-de-açúcar torna a avaliação dos clones consideravelmente mais precisa (Cardozo & Sentelhas 2013). Portanto, negligenciar os efeitos do clima em modelos de análise pode levar a interpretações equivocadas e à exclusão de genótipos com potencial adaptativo. Como a expressão fenotípica dos genótipos é influenciada tanto por sua constituição genética quanto pelos fatores ambientais que regulam o ambiente de crescimento, a interação  $G \times A$  é influenciada por fatores ambientais previsíveis, como o tipo de solo e a posição geográfica (latitude e longitude), bem como por elementos imprevisíveis, pois variam em sua intensidade e frequência ao longo do tempo, bem como as variações sazonais, como os fatores climáticos. Estes são elementos que não seguem um padrão fixo e podem ser considerados variáveis estocásticas (Mirzawan et al., 1994; Cullis et al., 2000; Costa Neto et al., 2020). Esses fatores podem ser usados para agrupar ambientes semelhantes (Clay & Dejaegher, 1987; Fischer et al., 2000; Huggins et al., 2011). Cada grupo de ambiente formado é conhecido por diversos termos, dentre os quais o termo mega-ambientes (Krause et al., 2022).

No entanto, segundo Yan et al. (2023) o agrupamento de ambientes baseado apenas em limitações geográficas pode não ser suficiente para orientar estratégias de melhoramento genético. Uma definição alternativa de mega-ambientes está relacionada aos genótipos, onde é possível a identificação de áreas onde eles mostram respostas semelhantes dentro de diferentes regiões (Gauch & Zobel, 1997). Contudo, basear-se apenas no genótipo para obter mega-ambientes pode não ser confiável, pois os métodos usados são sensíveis a remoção do genótipo, mudando completamente a configuração do agrupamento de ambientes. Além disso, é preciso considerar a qualidade das medidas fenotípicas ao agrupar ambientes, já que erros na coleta ou mensuração dessas variáveis podem levar a estimativas viesadas e escolhas inadequadas de genótipos (Annicchiarico, 2002). Portanto, é importante considerar os efeitos climáticos para avaliar corretamente o desempenho dos genótipos em diferentes regiões, e

com isso evitar interpretações equivocadas e a eliminação de genótipos adaptativos diante das mudanças climáticas (Zhao & Li, 2015).

Atualmente a coleta de dados climáticos está cada vez mais acessível e pode ser obtida gratuitamente por meio de portais como a NASA (<https://power.larc.nasa.gov>), o Instituto Nacional de Meteorologia (<https://portal.inmet.gov.br>) e o Centro Europeu de Previsão de Tempo de Médio Prazo (<https://www.ecmwf.int>), além de outras instituições meteorológicas e órgãos governamentais que disponibilizam informações climáticas cada vez mais abrangentes em todo o mundo.

A complexa variabilidade dos dados climáticos, especialmente no melhoramento de culturas de ciclo longo como a cana-de-açúcar (Luo et al., 2023) pode trazer desafios na explicação do desempenho dos clones nas etapas finais do programa de melhoramento genético. Nesse cenário, o Modelo Aditivo Generalizado (GAM) (Hastie & Tibshirani, 1987) pode ser uma ferramenta útil e mais flexível em comparação a outros modelos como o Modelo Linear Generalizado (GLM) Dobson (2001). O GAM ajusta funções para cada covariável ambiental, lidando com relações complexas e não lineares entre as variáveis independentes e a variável dependente. Ele incorpora funções suaves, como splines, para modelar padrões não lineares nos dados, capturando relações mais flexíveis e complexas (Wood, 2017). Essa flexibilidade torna o GAM valioso na avaliação da influência do clima na resposta dos clones de cana-de-açúcar ao ambiente de crescimento a que estão submetidas, especialmente quando padrões complexos e não lineares estão presentes. Sua aplicação em estudos em feijão demonstrou a habilidade de elucidar os possíveis “limites ambientais” para a adaptação através de diferentes condições ambientais (Heinemann et al., 2022).

Neste contexto, o objetivo deste estudo foi: (i) avaliar a formação de mega-ambientes por meio do uso de variáveis climáticas; (ii) avaliar a consistência de grupos de ambientes em cada estágio fenológico da cultura; (iii) e avaliar, por meio do modelo GAM, quais variáveis climáticas têm maior contribuição na compreensão da performance dos diferentes grupos de genótipos cana-de-açúcar (mais instáveis, mais estáveis e melhor classificados).

## 3.2 MATERIAL E MÉTODOS

### 3.2.1 Descrição dos dados e da área experimental

Os dados foram obtidos a partir de experimentos conduzidos na fase experimental do programa de melhoramento da cana-de-açúcar da Universidade Federal de Viçosa (PMGCA-

UFV) (Barbosa & da Silveira, 2015). Em 2019, foram avaliados 11 genótipos, incluindo a variedade RB 966928 e dez clones. Estes genótipos foram plantados em oito localidades distintas (Figura 1), com um delineamento experimental de blocos completos casualizados com 4 repetições. Na análise foram considerados os dados de TCH referentes ao ano 2 (junho de 2020 a agosto de 2021), correspondente à safra de cana soca, calculada de acordo com o Manual de Instruções do Conselho dos Produtores de Cana-de-Açúcar, Açúcar e Álcool do Estado de São Paulo (CONSECANA, 2012). Cada parcela experimental foi constituída de quatro sulcos de cinco metros de comprimento, com espaçamento de 1,4 m entre linhas.

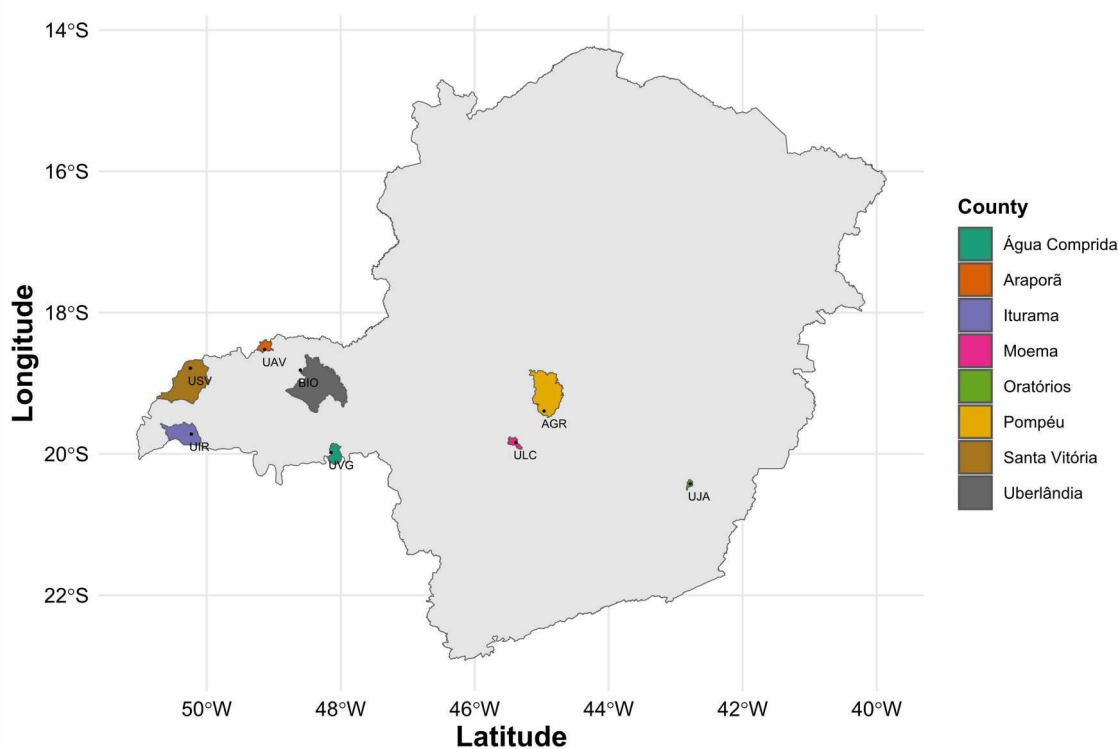


Figura 1 Locais onde os experimentos foram realizados no estado de Minas Gerais, Brasil. UAV - Usina Alvorada, UIR - Usina Coruripe, AGR - Usina Agropéu, BIO - Usina Bioaroeira, USV - Usina Santa Vitória, ULC - Usina Santa Lúcia, UJA - Usina Jatiboca, UVG - Usina Delta Sucroenergia (Unidade Volta Grande).

Cada localidade teve sua coordenada geográfica coletada com a finalidade de se obter os dados climáticos para uso nessa pesquisa (Tabela 1). Essas coordenadas são fundamentais para a identificação da localização precisa de cada área de estudo e permitir a obtenção dos dados climáticos a serem associados a cada experimento.

Tabela 1 Informações geográficas da fase experimental do programa de melhoramento da cana-de-açúcar: latitude, longitude, município e altitude (m).

Código	Localidade	Latitude	Longitude	Cidade	Altitude
1	ULC	19°50'13.691"S	45°23'5.577"O	Moema	739
2	UJA	20°25'18.190"S	42°47'0.764"O	Oratórios	524
3	AGR	19°23'35.510"S	44°57'54.840"O	Pompéu Velho	638
4	USV	18°47'14.843"S	50°14'36.817"O	Perdilândia	490
5	BIO	18°48'45.257"S	48°36'4.312"O	Uberlândia	903
6	UIR	19°43'7.080"S	50°13'43.624"O	Iturama	466
7	UAV	18°31'20.310"S	49°8'7.069"O	Alceu Garcia	513
8	UVG	19°58'45.875"S	48°8'43.437"O	Conceição das Alagoas	556

Os dados climáticos diários foram coletados durante o período de cultivo da cana soca, de junho de 2020 a agosto de 2021, doravante chamado de ano 2, com o uso da função `get_weather` do pacote `EnvRtype` (Costa-Neto et al., 2021) no software R (R Core Team, 2019). Algumas das variáveis que descrevem os processos ecofisiológicos (evapotranspiração, déficit de pressão de vapor, inclinação da curva de déficit de pressão de vapor e efeito da temperatura na eficiência do uso da radiação) foram calculadas conforme a descrição feita por Soltani & Sinclair (2012). Diversas variáveis são adquiridas por meio da função `get_weather`. Entretanto, optou-se por selecionar aquelas que já foram analisadas em estudos anteriores, dado o reconhecimento de sua importância e influência na cultura da cana-de-açúcar (Viana et al., 2023; Zhang et al., 2017; Zhao & Li 2015; Cardozo & Sentelhas 2013; Ramburan et al., 2011). Um resumo das variáveis usadas nas análises e suas descrições estão apresentadas na Tabela 2.

Tabela 2 Descrição das variáveis climáticas e ecofisiológicas obtidas de duas fontes diferentes. (Nasa Power, <https://power.larc.nasa.gov>; e calculadas conforme necessário).

Fonte	Variáveis	Descrição	Unidade
Nasa Power	Tmed	Temperatura média do ar a 2 metros acima da superfície da Terra	°C
	Tmax	Temperatura máxima do ar a 2 metros acima da superfície da Terra	°C
	Tmin	Temperatura mínima do ar a 2 metros acima da superfície da Terra	°C
	Prec	Precipitação de chuva	mm
	RH	Umidade relativa a 2 metros acima da superfície da Terra	%
	DEWPT	Temperatura do ponto de orvalho a 2 metros acima da superfície da Terra	°C
	rad1	Insolação total incidente no céu em uma superfície horizontal	MJ m <sup>-2</sup>
rad2	Fluxo radiativo de ondas longas infravermelhas térmicas	MJ m <sup>-3</sup>	
Calculadas	ETP	Evapotranspiração	mm
	VPD	Déficit de pressão de vapor	Kpa
	SPV	Inclinação da curva de déficit de pressão de vapor	Kpa °C
	FRUE	Efeito da temperatura na eficiência do uso da radiação	-

### 3.2.2 Análise de componentes principais e agrupamento hierárquico de Ward para obter a formação de mega-ambientes

A organização dos dados de clima para a subsequente aplicação das análises de componentes principais (ACP) e de agrupamento foi realizada de duas formas distintas. Na primeira, foi utilizada a média das variáveis climáticas do ano 2 para cada local de cultivo. Na segunda abordagem, com o intuito de assegurar a representatividade dos dados climáticos e identificar possíveis variações na formação dos mega-ambientes em cada estágio fenológico, o ano 2 foi dividido em três períodos distintos, correspondentes aos estágios fenológicos da cana-de-açúcar. Para cada estágio, foi calculada a média das variáveis climáticas. Devido às variações nas datas de colheita da cana planta e, conseqüentemente, às diferenças no início da brotação da cana soca (com cinco locais iniciando em junho e o restante em julho) entre as localidades, decidiu-se adotar uma data única para representar o início do estágio fenológico em todos os locais. Essa data escolhida coincidiu com o início da brotação da cana soca na maioria das localidades, que ocorreu no mês de junho.

Posteriormente, foi realizada a ACP para obter a formação dos mega-ambientes. A análise de componentes principais (ACP) corresponde a uma técnica multivariada utilizada para reduzir a dimensionalidade dos dados e compreender como as diferentes variáveis estão relacionadas entre si (Aamer et al., 2018). Esse tipo de análise possibilita identificar padrões ou estruturas que podem não ser prontamente visíveis devido à multicolinearidade entre as variáveis originais, possibilitando a observação visual das diferenças entre grupos de amostras (Wang et al. 2020). A ACP foi realizada em cada estágio usando a função `prcomp` (R Core Team, 2019) do pacote `stats` (R Core Team 2023).

A análise de agrupamento de Ward complementou os resultados obtidos pela ACP, pois oferece uma visão clara dos grupos de localidades formados. Dessa forma, seguiu-se o mesmo procedimento utilizado na análise anterior para os dados climáticos, que consistiu na obtenção da média dos três anos e para cada estágio fenológico. A análise de agrupamento foi realizada por meio do método de variância mínima de Ward (Khattree & Naik, 2000), através do pacote `factoextra` e da função `fviz_dend`. Para esta função, é necessário fornecer uma matriz de dissimilaridade, que pode ser facilmente obtida pela função `dist` no ambiente de programação R. O método de agrupamento hierárquico de Ward busca minimizar a variância dentro de cada grupo formado (Kassambara, 2017). Essa análise agrupa as localidades com

base na similaridade das variáveis e ajuda a identificar padrões de agrupamento relacionados aos mega-ambientes.

### 3.2.3 Contribuição das variáveis climáticas na performance do clone.

Os clones foram escolhidos com base na análise da estabilidade *versus* a média de rendimento, usando um gráfico biplot criado a partir da análise GGE biplot. Esta análise foi realizada por meio da função "gge" do pacote "metan" (Olivoto e Lúcio, 2020) e resultou na seleção de três grupos de clones: aqueles com comportamento estável, os instáveis e os que apresentaram as maiores médias de TCH.

As variáveis climáticas consideradas na formação dos mega-ambientes (Tabela 2) também foram utilizadas para explicar o desempenho dos clones em relação à variável TCH. Aplicou-se o GAM para determinar quais variáveis climáticas eram necessárias para explicar a performance de cada grupo de clones. Para melhor ajustar o GAM, foi aplicado o logaritmo aos dados de TCH, pois eles não seguiam uma distribuição normal. A função de suavização  $t_2$  (Wood, 2017) foi a que melhor se adequou aos dados. A equação utilizada para avaliar a relação entre a variável dependente e as variáveis explicativas é a seguinte (Equação 1):

$$\log TCH = \sum_{i=1}^a t_2(V_i) + \varepsilon, \quad (1)$$

em que:  $\log TCH$ , é a variável resposta transformada;  $V_i$  ( $i=1, 2, \dots, a$ ), são as variáveis climáticas,  $t_2$  é a função de suavização e  $\varepsilon$  é o erro aleatório, com  $\varepsilon \sim N(0, \sigma^2)$ .

Foram escolhidas um total de oito variáveis ambientais, uma vez que esse foi o número máximo aceito pelo modelo, devido à quantidade de observações (oito localidades). A seleção das variáveis foi realizada com base na sua ordem de importância obtida por meio do algoritmo Random Forest, por meio dos pacotes randomForest (Liaw & Wiener, 2002) e varImp (Probst 2020). Foram exploradas 256 combinações possíveis, variando de 1 a 8 variáveis ambientais como fatores explicativos no modelo GAM.

Para selecionar o melhor modelo em cada grupo de clones, foi utilizado o critério de informação de Akaike (AIC) (Bozdogan, 1987). Foram comparados os valores de AIC de todos os modelos ajustados e escolhido aquele com o menor valor em cada combinação de variáveis. Após selecionar os modelos, vários gráficos foram comparados, como o histograma e QQ plots dos resíduos, além de gráficos de resíduo versus preditor linear e o observado

versus estimado. Essas análises permitiram identificar a qualidade de ajuste, possíveis relações não lineares nos resíduos e avaliar a precisão dos modelos (Wood, 2017).

Foram também examinados o erro padrão e a significância estatística dos termos de suavização, incluindo os gráficos que mostram a relação entre cada variável ambiental e a variável resposta log de TCH. Essa análise tem como objetivo compreender a influência das variáveis explicativas e determinar quais variáveis climáticas apresentam significância estatística. Por fim, foram avaliados o coeficiente de determinação ( $R^2$ ) e a deviance em alguns modelos nos quais os gráficos foram semelhantes. Esses passos permitiram uma análise completa dos modelos GAM, onde foram selecionados os melhores modelos para cada grupo de clones.

### 3.3 RESULTADOS

#### 3.3.1 Formação de mega-ambientes com o uso de variáveis climáticas.

As análises de componentes principais e agrupamento de Ward, com o uso de variáveis climáticas coletadas no ano 2, produziram resultados semelhantes na identificação de grupos de localidades. Foram identificados dois mega-ambientes (Figura 3A-B). As localidades USV, UIR, UVG e UAV, apresentam padrões e comportamentos similares na variação das temperaturas e suas variáveis correlacionadas (Figura 2). Portanto, essas localidades agrupadas apresentaram uma correlação positiva com o CP1 (Figura 3A).

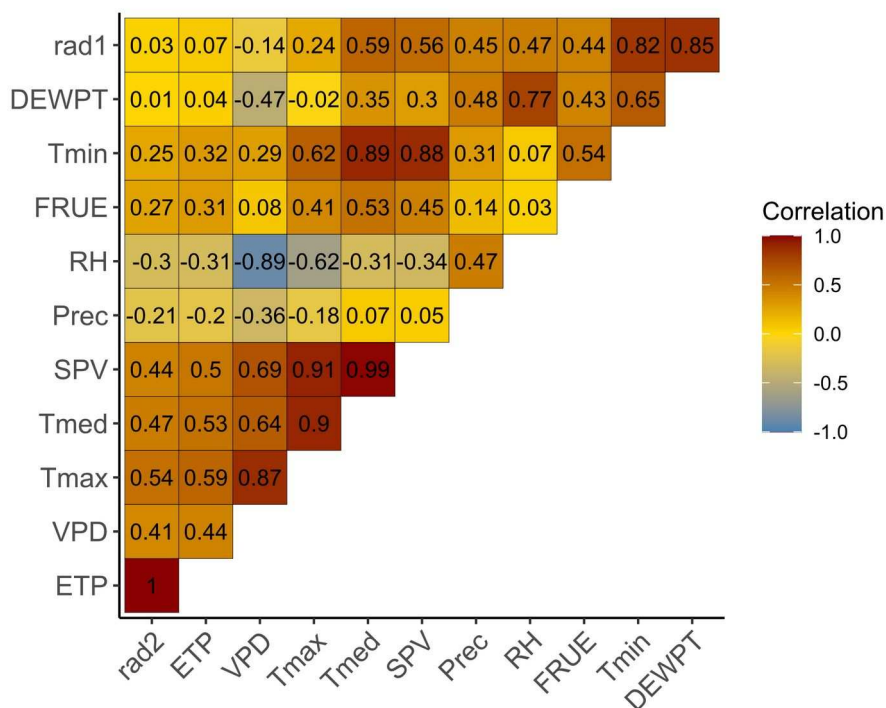


Figura 2 Correlação de Pearson entre as variáveis climáticas obtidas no ano 2 (cana-soca) para todas as localidades.

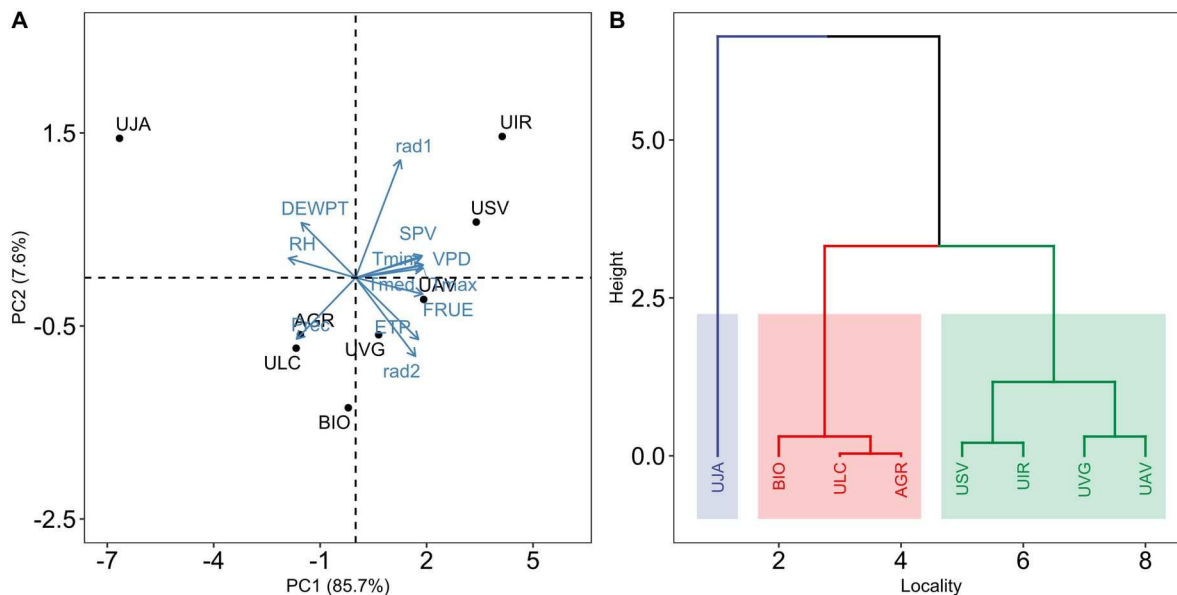


Figura 3 (A) Gráfico biplot para variáveis climáticas e ecofisiológicas. (B) Análise de agrupamento Ward das localidades.

No segundo grupo, formado pelas localidades ULC, AGR e BIO (Figura 3B), as regiões apresentam um comportamento oposto em relação a radiação 1 e correlação positiva com a precipitação (Figura 3A). Isso indica que essas localidades têm valores mais baixos de radiação 1 quando comparados aos outros ambientes. Além disso, a localidade UJA se destaca

por ter valores mais elevados de umidade relativa (RH) e temperatura no ponto de orvalho (DEWPT), além de temperaturas mais baixas (Figura 3A). Portanto, ela foi separada dos demais grupos por causa do seu comportamento singular em relação às outras localidades.

Ao avaliar apenas o ano de cultivo da cana soca, dividido em estádios fenológicos, os resultados obtidos por meio da análise de agrupamento foram semelhantes quando analisados os três anos em conjunto (Figura 4a). Nos estádios fenológicos inicial e intermediário (Figuras 4b-4c) não houve diferenças no agrupamento das localidades em relação aos dados dos três anos. A única exceção foi a localidade UVG, que por estar mais semelhante aos ambientes ULC, AGR e BIO, nos estádios inicial e intermediário passou a fazer parte deste grupo. No terceiro estágio fenológico, o ambiente UAV também pode apresentar alguma semelhança com esse grupo mencionado anteriormente (Figuras 4d). Por outro lado, a localidade UJA continua separada dos demais grupos, exibindo um comportamento distinto em relação aos outros ambientes.

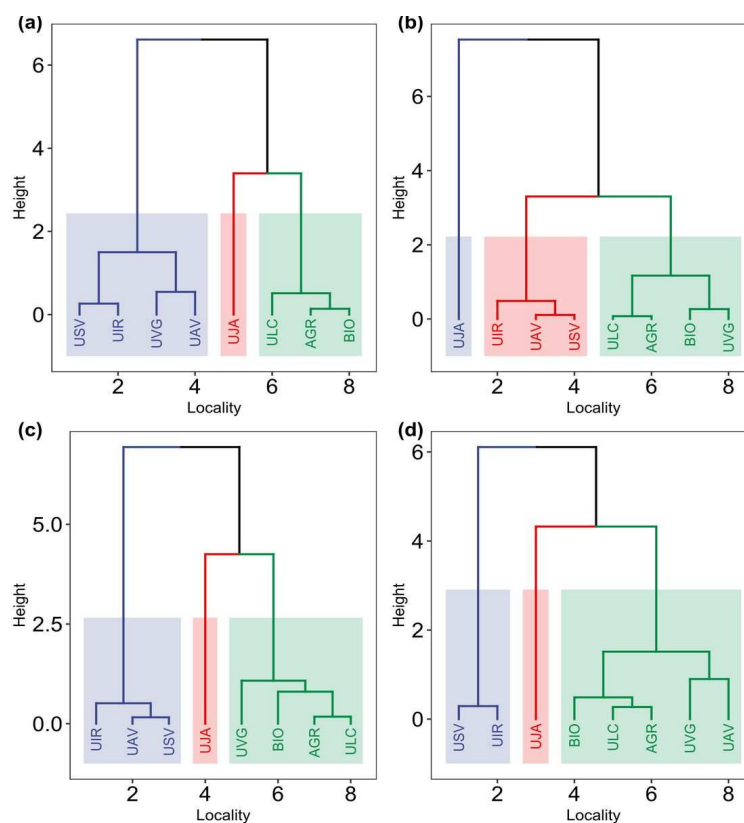


Figura 4 Agrupamento de localidades com base em variáveis climáticas ao longo dos três estágios fenológicos da cana-de-açúcar. (a) Média das variáveis climáticas para os três anos (2019, 2020,

2021); (b) Média das variáveis climáticas para o primeiro estágio: 06-01-2020 to 09-05-2020; (c) Média das variáveis climáticas para o segundo estágio: 09-06-2020 to 02-12-2021; (d) Média das variáveis climáticas para o segundo estágio: 02-13-2021 to 06-04-2021.

### 3.3.2 Análise climática global e seu impacto no rendimento de TCH por localidade.

A avaliação climática das localidades e o desempenho d foi focada exclusivamente na cana soca, dadas as distintas datas de corte da cana planta (junho e julho) e cana soca (maio, junho, julho e agosto), o que resultou em variações no início do estágio fenológico. No entanto, a análise gráfica foi conduzida agrupando as localidades com coincidência nas datas de corte da cana planta. Localidades como BIO e UVG se destacaram com médias mais elevadas de TCH (Figura 5) (corte cana planta e cana soca - julho e agosto, respectivamente). Ao examinar os padrões de precipitação e temperatura, é possível observar uma correspondência entre os acumulados de chuva e os registros de temperatura. No estágio inicial (brotação e perfilhamento), as chuvas foram escassas, no entanto, no estágio subsequente, o acúmulo de chuvas atingiu 90 mm em ambas as localidades (Figuras 6). Este período corresponde à fase de crescimento em altura da planta e exige um suprimento adequado de água. A localidade UIR registrou menor volume de chuvas durante o estágio de crescimento em altura, que reflete em médias menores de TCH.

Por outro lado, embora a localidade UJA tenha apresentado as menores médias de TCH, seu período de cultivo foi similar ao da localidade ULC (corte cana planta e cana soca - junho e maio, respectivamente), que obteve a terceira maior média. A principal diferença entre esses ambientes foi a quantidade de chuva acumulada, na ULC, o volume foi maior e mais constante durante o estágio de desenvolvimento da cultura, enquanto na UJA, foram observados alguns picos de precipitação ao longo desse período. Em relação à temperatura, todas as localidades apresentaram padrões similares, exceto a UJA, que registrou médias mais baixas.

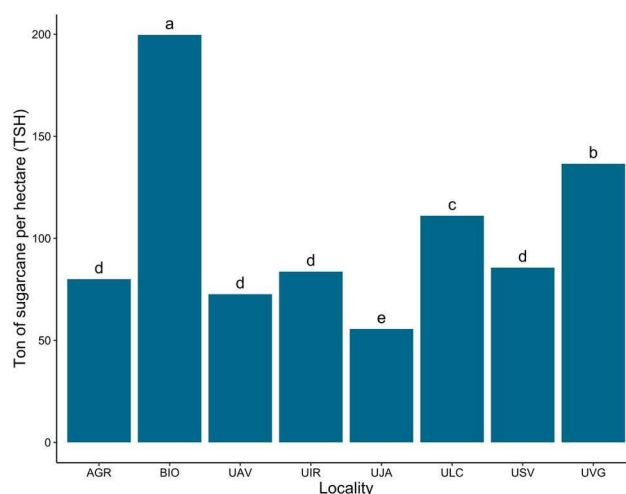


Figura 5 Teste de média de TCH entre localidades na fase experimental do programa de melhoramento da cana-de-açúcar, com dados de cana soca (2021). As letras associadas (quando diferentes) indicam significância estatística ( $p < 0,05$ ) entre as médias de TCH entre as localidades (teste de Tukey).

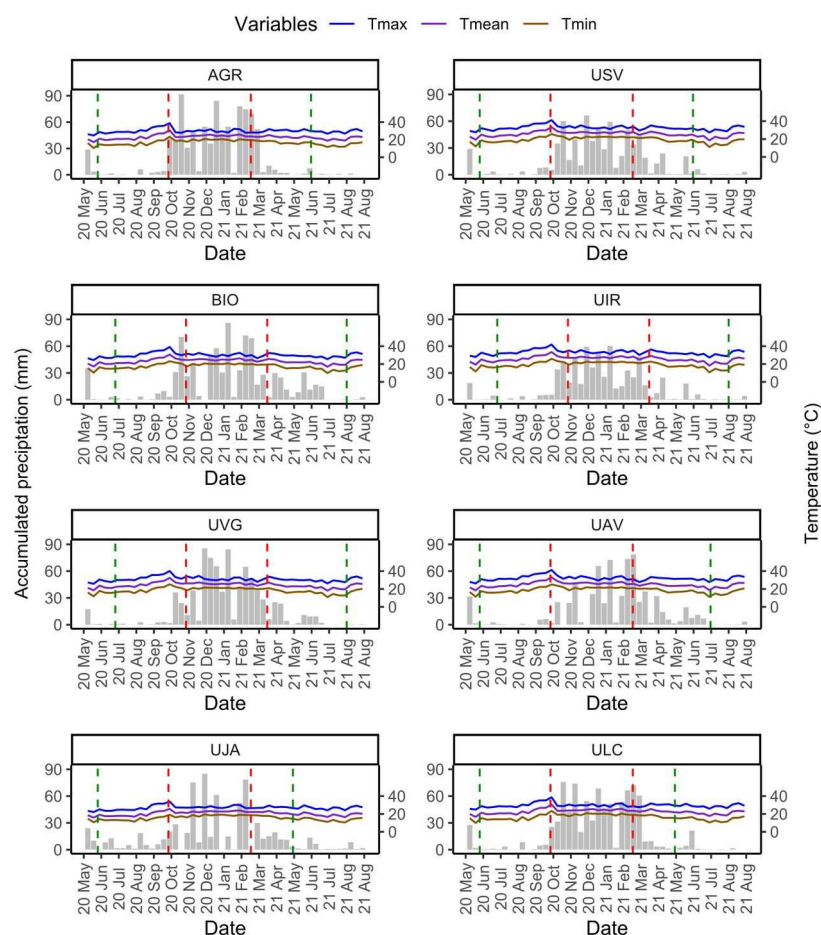


Figura 6 Médias de temperatura média, máxima e mínima, e precipitação acumulada em intervalos de 10 dias ao longo do desenvolvimento da cultura (cana soca). As duas linhas verdes representam as datas de corte da cana planta e da cana soca, respectivamente, e as linhas pontilhadas vermelhas representam a segmentação do gráfico de acordo com as fases fenológicas específicas da cana-de-açúcar para cada localidade.

### 3.3.3 Influência de variáveis climáticas na adaptação de clones de cana-de-açúcar

Inicialmente, os clones de cana-de-açúcar selecionadas foram divididos em três grupos. Cada grupo apresenta características de ser pouco ou muito influenciado pelas variáveis climáticas. A escolha dos clones foi feita através da análise do gráfico de estabilidade versus média de rendimento (Yan et al., 2007). Para avaliar essas respostas fenotípicas, é calculada a coordenada média do ambiente (*Average environment coordinate* - AEC). A AEC é calculada a partir dos escores das coordenadas CP1 e PC2 dos ambientes, representando a média dos rendimentos entre os locais.

A distância visual entre os genótipos no gráfico indica a diferença de magnitude entre eles. A linha perpendicular à abscissa, que passa pela origem do biplot, representa a ordenada da AEC. A projeção dos genótipos nessa ordenada indica a medida de variabilidade ou instabilidade. Quanto maior a projeção ao longo da ordena AEC, maior é a instabilidade do genótipo (Yan et al., 2007). Com base nessa análise, foram selecionados os três genótipos com os maiores rendimentos para a variável TCH, que foram: G11, G5 e G6, formando o primeiro grupo de clones. Além disso, foi identificado os três genótipos com maior instabilidade, que foram G9, G4 e G3, formando outro grupo. Por outro lado, apenas dois clones foram selecionados (G8 e G5) para o terceiro grupo, composto pelos mais estáveis, uma vez que apresentam respostas semelhantes nos diferentes locais (Figura 7).

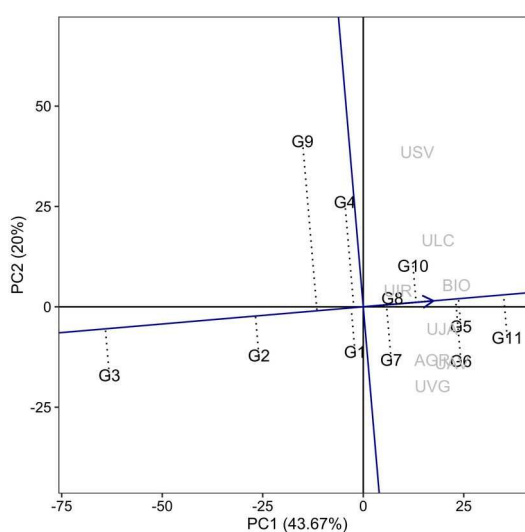


Figura 7 Estabilidade da produtividade média dos 11 clones cultivados em oito localidades para o traço fenotípico TCH.

A avaliação da influência das variáveis ambientais nos genótipos foi realizada no ano dois, uma vez que os dados fenotípicos obtidos correspondem à cana soca. Os resultados apresentados referem-se aos modelos selecionados em cada grupo de genótipos. O número de variáveis explicativas foi variável em cada grupo de clones.

O ajuste do modelo melhorou até o uso de quatro variáveis explicativas para o grupo de clones mais estáveis, mas a partir desse ponto, a melhora tornou-se menos significativa, com base nos parâmetros usados para avaliar os modelos (Tabela 3). À medida que mais variáveis eram incluídas, o desempenho do modelo mostrava uma redução notável (dados não mostrados). Por outro lado, o grupo de clones vencedoras exigiu um maior número de variáveis para otimizar o ajuste do modelo. Um total de sete variáveis foi necessário para atingir um  $R^2$  de 77,5 (Tabela 3). Enquanto isso, para o grupo de clones instáveis, os resultados das análises foram semelhantes, com a adição de uma variável para explicar o desempenho desses clones.

Tabela 3 Parâmetros para avaliação do ajuste do modelo: significância dos termos, AIC (*Akaike Information Criterion*),  $R^2$  (coeficiente de determinação), erro padrão e deviance para cada grupo de clones de cana-de-açúcar.

Group								
Stable			Higler mean			Unstable		
variables	GL ref.	P – value	variables	GL ref.	P – value	variables	GL ref.	P – value
t2(Tmax)	4	7.61e-07 ***	t2(Tmed )	2	4.47e-07 ***	t2(Tmed )	4	<2e-16 ***
t2(RH)	4	< 2e-16 ***	t2(Tmax )	4	5.90e-07 ***	t2(Tmax )	9	<2e-16 ***
t2(SPV)	4	4.80e-07 ***	t2(Prec)	4	4.15e-07 ***	t2(Prec)	4	0.804
t2(FRUE)	3	0.0581	t2(RH)	9	0.00121 **	t2(RH)	4	2.52e-06 ***
			t2(rad1)	2	3.34e-06 ***	t2(rad1)	4	1.66e-06 ***
			t2(VPD)	4	0.14027	t2(VPD)	4	<2e-16 ***
			t2(FRUE )	4	0.00298 **	t2(SPV)	4	<2e-16 ***
						t2(FRUE )	4	0.00298 **
AIC	6,129		36,035			28,397		
$R^2$ (%)	68,6		77,5			73,8		

Standart error	0,02838	0,01836	0,02658
Deviance (%)	75,1	82,2	77,2

\*\* , \*\*\* significant at 1% and 0.1% probability by the F-test

Além disso, a Tabela 3 apresenta a significância das variáveis climáticas combinadas com a função de suavização (t2) para cada grupo de clones. No grupo estável, as variáveis Tmax, RH e SPV foram significativas ( $p < 0,001$ ) e contribuíram para explicar o comportamento dos clones. Ao avaliar o grupo vencedor, houve um aumento no número de variáveis, com a inclusão das variáveis RH, Tmed, Prec, rad1, que foram significativas com uma probabilidade de 0,1 % e 1%, respectivamente, e a variável VPD.

Para o grupo instável, a inclusão da variável SPV melhorou o ajuste do modelo. Foi observado que tanto a temperatura máxima quanto a umidade relativa surgiram como os principais fatores climáticos que tiveram uma contribuição significativa para o ajuste do modelo em todas os clones de cana-de-açúcar. Embora as variáveis FRUE, VPD e Prec não tenham alcançado significância estatística, sua presença ainda demonstrou influenciar positivamente cada modelo, pois sua exclusão resultou em uma piora do ajuste.

### 3.4 DISCUSSÃO

Este estudo apresenta uma abordagem para agrupar ambientes com base em variáveis que afetam o meio de crescimento da cultura, com foco exclusivo nas variáveis climáticas, uma vez que as informações sobre o solo não estavam disponíveis. Dessa forma, a finalidade foi avaliar se apenas com dados climáticos poderíamos obter resultados satisfatórios. Essa estratégia aprimora a precisão e confiabilidade da análise, pois mantém a consistência dos grupos de ambientes mesmo em estádios fenológicos distintos. Assim, a utilização de mega-ambientes pode resultar em benefícios econômicos, pois permite a redução dos custos associados à implementação futura desses experimentos, o que torna crucial a otimização dos recursos (Das et al., 2019).

A instalação de experimentos na fase experimental pode ser mais eficaz ao considerar as condições climáticas dentro de um grupo de ambientes, pois o clima exerce grande influência na performance das plantas, e cada estágio fenológico exige uma condição climática específica para o desenvolvimento das plantas. Com condições climáticas semelhantes, é

possível padronizar as datas de corte dentro de um mega-ambiente. Como por exemplo, as localidades UVG, UAV, USV e UIR apresentaram médias mais altas de temperatura, radiação e outras variáveis correlacionadas, enquanto precipitação e umidade relativa apresentaram correlação negativa. Com base nessa similaridade entre os ambientes, é possível planejar datas semelhantes de corte para um grupo de locais, em vez de ter datas distintas para cada um. Isso possibilita focar apenas na melhor data de corte para clones ou variedades com diferentes tempos de maturação (precoce e tardia). Ou seja, dentro de um mega-ambiente, o corte dos clones precoces seria no início da safra, enquanto o corte dos tardios seria no final da safra.

A vantagem de definir épocas de corte da cana-de-açúcar em cada mega-ambiente é a otimização da logística. Colher em duas épocas diferentes no mesmo local é muito trabalhoso, pois demanda mais mão de obra e eleva os custos. Com a padronização das datas de corte, esses desafios logísticos são reduzidos, resultando em maior eficiência operacional. Dessa forma, basear a definição dos grupos de ambientes nas condições climáticas de cultivo é uma abordagem complementar à simples avaliação fenotípica dos genótipos. No entanto, o uso isolado do clima para o planejamento de futuros experimentos considerando mega-ambientes pode não ser confiável, já que existem outros fatores importantes para o desenvolvimento das plantas, como a fertilidade do solo e a produtividade do local.

Como observado nos resultados, o ambiente BIO apresentou uma média de TCH superior em comparação aos demais ambientes (199,72 t ha<sup>-1</sup>). No entanto, foi agrupado com outras duas localidades, AGR e ULC, que obtiveram médias de 80 a 111 t ha<sup>-1</sup>, respectivamente. Conhecendo-se as características do ambiente de cultivo do local BIO, observa-se que a fertilidade e as características do solo nessa área contribuem para uma maior produtividade. Em contrapartida, os solos de AGR e ULC apresentam mais limitações, refletindo em uma produtividade menor.

Ao otimizar a seleção dos locais de teste na fase experimental e considerar as peculiaridades de cada região, é possível maximizar o aproveitamento tanto dos locais quanto dos genótipos (Yan et al. 2023). Por exemplo, na localidade UJA, onde a irregularidade das chuvas durante o estágio de crescimento varia entre picos elevados em determinados períodos e precipitação próxima a zero em outros, juntamente com temperaturas baixas e radiação reduzida devido à constante presença de nuvens, apresentou menores médias de TCH.

A baixa produtividade em UJA pode ser explicada pela importância da radiação para a fotossíntese, pois a energia necessária para transformar o CO<sub>2</sub> atmosférico em energia metabólica vem da radiação solar (Taiz e Zeiger, 2007). Com menor incidência de radiação, é provável que a taxa fotossintética tenha diminuído, resultando em menor crescimento das plantas e, conseqüentemente, em menor produtividade nesse local. Além disso, a baixa fertilidade do solo no ambiente UJA também contribuiu para a menor produtividade. Portanto, é importante considerar o local UJA no planejamento da fase experimental de forma individual, especialmente no que se refere ao manejo, e identificar estratégias para melhor aproveitar esse ambiente, mesmo com condições de crescimento adversas.

Para uma análise mais detalhada sobre a influência climática nos genótipos de cana-de-açúcar, foi empregado o modelo GAM para identificar quais variáveis mais influenciam na performance dos genótipos. Os dados utilizados foram de cana soca, pois havia poucos dados disponíveis da cana planta devido à pandemia. Em vários locais, não foi possível coletar dados, já que o plantio foi executado em 2019 e a colheita estava prevista para 2020.

Dado que apenas um genótipo era uma variedade e os demais eram clones, nosso conhecimento sobre o comportamento em diferentes ambientes é mais desenvolvido para a variedade. A variedade G11 (RB 966928) é conhecida por sua maturação precoce, com colheita prevista entre abril e maio. Ela é a segunda variedade mais cultivada no Brasil, cobrindo 9,7% da área de plantio (RIDESA, 2010). G11 destaca-se pela alta produtividade e capacidade de rebrota nos ambientes de média a alto potencial produtivo. No entanto, essa variedade apresenta alta instabilidade e não é recomendada para ambientes restritivos.

Na fase experimental, os genótipos geralmente exibem grande variabilidade em relação à produtividade. Por isso, é crucial entender o quão estável e adaptável é cada clone e seu tempo de maturação. As condições climáticas podem fornecer *insights* valiosos para determinar o manejo adequado de cada clone. Compreender como os diferentes fatores climáticos influenciam o desempenho de cada genótipo permite otimizar as práticas agrícolas e selecionar as melhores estratégias de cultivo para maximizar a produtividade e a resiliência das plantas em diferentes ambientes.

Dessa forma, nossos resultados mostraram que o número de variáveis climáticas influentes no modelo depende do grupo ao qual o genótipo pertence. Os genótipos que se destacaram com as maiores médias de TCH demonstraram maior sensibilidade ao ambiente de cultivo e,

responde a uma ampla gama de condições climáticas. A inclusão de um número maior de variáveis explicativas no modelo permite capturar melhor essa relação complexa entre as variáveis climáticas e o rendimento dos clones. É importante considerar que, embora a inclusão de mais variáveis possa proporcionar um ajuste mais preciso, isso também pode aumentar a complexidade e a dificuldade de interpretação do modelo (Wood, 2017).

Por outro lado, os genótipos estáveis, que já apresentam um desempenho consistente em diferentes condições, tendem a ser menos sensíveis às variações climáticas (Mohammadi and Amri, 2008). Para esses genótipos, um menor número de variáveis explicativas pode ser suficiente para explicar a maior parte da variação no TCH. Isso indica que esses clones são menos afetados por mudanças específicas nas condições climáticas e exibem um comportamento mais estável em relação ao TCH.

Portanto, a diferença na quantidade de variáveis climáticas necessárias para um melhor ajuste dos modelos está relacionada às características intrínsecas dos grupos de genótipos. Clones e/ou variedades de maior rendimento podem exigir uma consideração mais abrangente das variáveis climáticas para capturar todos os fatores que influenciam seu desempenho. Em contraste, clones estáveis podem ser menos sensíveis a essas variações e requerer menos variáveis para explicar seu desempenho consistente. Isso sugere que a abordagem para modelagem e análise deve ser adaptada às especificidades de cada grupo de genótipos, garantindo um equilíbrio entre precisão e simplicidade na interpretação dos resultados.

Essas descobertas destacam a complexidade das interações entre as variáveis climáticas e o desempenho dos genótipos de cana-de-açúcar, resumidas na conhecida interação  $G \times A$  (Ramburan et al., 2011). Além disso, nossos resultados ressaltam a importância de considerar o número adequado de variáveis explicativas, levando em conta as características agrônômicas e a estabilidade dos grupos de genótipos, visando obter os melhores ajustes de modelo. Na prática, identificar as variáveis climáticas de maior influência no desempenho dos clones é crucial para aprimorar estratégias eficazes de seleção e manejo das culturas (Liu et al., 2021). Esse entendimento possibilita a adaptação e o desenvolvimento dos clones mais resistentes e adaptadas a condições ambientais específicas, permitindo que agricultores e pesquisadores concentrem esforços para melhorar a produção agrícola, aumentando assim a eficiência e a sustentabilidade dos sistemas de cultivo.

### 3.5 CONCLUSÃO

Foi possível identificar padrões de comportamento e correlações entre diferentes variáveis climáticas em diversas localidades. Ao analisar essas interações, notamos a formação de grupos distintos de localidades que compartilham características climáticas semelhantes, delineando assim os chamados mega-ambientes. No entanto, ao observar o desempenho dos genótipos dentro desses mega-ambientes, torna-se evidente a discrepância na produtividade entre eles. Em estudos futuros, além do clima, será essencial considerar também fatores do solo, para que os mega-ambientes sejam mais representativos e expliquem de forma mais adequada a variabilidade de produção dos genótipos.

A consistência na formação dos mega-ambientes foi notável ao longo do desenvolvimento da cultura. Os fatores ambientais que influenciam a formação desses grupos se mantiveram estáveis ao longo das diferentes fases fenológicas da planta. Esta estabilidade sugere que os mega-ambientes são uma ferramenta promissora para categorizar locais com base em suas características climáticas.

As variáveis climáticas influenciam de maneira diversa o desempenho dos clones de cana-de-açúcar. Os clones que apresentaram maiores médias de TCH e maior instabilidade exigiram um maior número de variáveis independentes para explicar seu desempenho. Em contraste, os clones com comportamento mais estável foram menos dependentes de múltiplas variáveis. Em geral, a temperatura e a umidade relativa do ar foram os principais fatores climáticos que contribuíram significativamente para o ajuste do modelo em todas os clones de cana-de-açúcar.

### 3.6 REFERÊNCIA BIBLIOGRÁFICA

Annicchiarico, P. (2002). Genotype x environment interactions: challenges and opportunities for plant breeding and cultivar recommendations.

Barbosa, M. H. P., & da Silveira, L. C. I. (2015). Breeding program and cultivar recommendations. In Sugarcane (pp. 241-255). Academic Press.

Bezerra, J. D. C., Ferreira, G. D. G., de Oliveira, M. W., de Souza Campos, J. M., de Andrade, A. P. & do Nascimento Júnior, J. R. S. (2018). Cana-de-açúcar: Melhoramento genético e suas finalidades forrageiras. *Nucleus Animalium*, 10(2), 131-147.

- Borém, A., Miranda, G. V., & Fritsche-Neto, R. (2021). Melhoramento de plantas. Oficina de Textos.
- Cardozo, N. P. & Sentelhas, P. C. (2013). Climatic effects on sugarcane ripening under the influence of cultivars and crop age. *Scientia Agricola*, 70, 449-456. <http://doi.org/10.1590/S0103-90162013000600011>.
- Cheavegatti-Gianotto, A., De Abreu, H. M. C., Arruda, P., Bessalho Filho, J. C., Burnquist, W. L., Creste, S. & César Ulian, E. (2011). Sugarcane (*Saccharum X officinarum*): a reference study for the regulation of genetically modified cultivars in Brazil. *Tropical plant biology*, 4, 62-89.
- CONAB. Companhia Nacional de Abastecimento. Available in: <[http://www.conab.gov.br/conabweb/download/safra/1graos\\_08.09.pdf](http://www.conab.gov.br/conabweb/download/safra/1graos_08.09.pdf)>. accessed in January 15, 2023.
- Conselho dos Produtores de Cana-de-Açúcar, Açúcar e Alcool do Estado do Paraná - CONSECANA (2012) Manual de Instruções. 3. ed. Curitiba-PR. 118p.
- Costa-Neto, G.M.F., J.B. Duarte, A.P. de Castro, & Heinemann, A.B. (2020). Uso de informações ambientais na modelagem e interpretação da interação genótipo x ambiente: revisão bibliográfica.
- Costa-Neto, G., Fritsche-Neto, R., & Crossa, J. (2021). Nonlinear kernels, dominance, and environment data increase the accuracy of genome-based prediction in multi-environment trials. *Heredity*, 126(1), 92-106.
- Cullis, B. R., Smith, A., Hunt, C., & Gilmour, A. (2000). An examination of the efficiency of Australian crop variety evaluation programmes. *The Journal of Agricultural Science*, 135(3), 213-222. <http://doi.org/10.1017/S0021859699008163>.
- Clay, D. C., & Dejaegher, Y. M. J. (1987). Agro-ecological zones: the development of a regional classification. *Tropicultura*, 5(4), 153-159.
- Das, A., Parihar, A. K., Saxena, D., Singh, D., Singha, K. D., Kushwaha, K. P. S., ... & Gupta, S. (2019). Deciphering genotype-by-environment interaction for targeting test environments and rust resistant genotypes in field pea (*Pisum sativum* L.). *Frontiers in plant science*, 10, 825.

- Djidonou, D., Leskovar, D. I., Joshi, M., Jifon, J., Avila, C. A., Masabni, J., ... & Crosby, K. (2020). Stability of yield and its components in grafted tomato tested across multiple environments in Texas. *Scientific Reports*, 10(1), 13535.
- DeMorais, L. K., de Aguiar, M. S., de Albuquerque e Silva, P., Câmara, T. M. M., Cursi, D. E., Júnior, A. R. F., ... & Bespalhok Filho, J. C. (2015). Breeding of sugarcane. *Industrial crops: breeding for bioenergy and bioproducts*, 29-42.
- Dobson, A. J. (2001). *An Introduction to Generalized Linear Models* (2nd ed.). Chapman & Hall/CRC.
- Eberhart, S. T., & Russell, W. A. (1966). Stability parameters for comparing varieties 1. *Crop science*, 6(1), 36-40.
- Finlay, K. W., & Wilkinson, G. N. (1963). The analysis of adaptation in a plant-breeding programme. *Australian journal of agricultural research*, 14(6), 742-754.
- Fischer, G., van Velthuisen, H. T., & Nachtergaele, F. O. (2000). Global agro-ecological zones assessment: methodology and results.
- Gauch Jr, H. G., & Zobel, R. W. (1997). Identifying mega-environments and targeting genotypes. *Crop science*, 37(2), 311-326.
- Hastie, T. & Tibshirani, R. (1987). Generalized additive models: some applications. *Journal of the American Statistical Association*, 82(398), 371-386. <http://doi.org/10.1080/01621459.1987.10478440>.
- Huggins, D. R., Rupp, R., Gessler, P., Pan, W., Brown, D. J., Machado, S., ... & Abatzoglou, J. T. (2011). Dynamic agroecological zones for the inland Pacific Northwest, USA. In *AGU Fall Meeting Abstracts* (Vol. 2011, pp. B12A-02).
- Holland, S. M. (2019). *Principal components analysis (PCA)*. Department of Geology, University of Georgia, Athens, GA, 30602-2501.
- 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.
- Hyman, G., Hodson, D., & Jones, P. (2013). Spatial analysis to support geographic targeting of genotypes to environments. *Frontiers in physiology*, 4, 40.

- Kassambara, A. (2017). Practical guide to cluster analysis in R: Unsupervised machine learning (Vol. 1). Sthda.
- Khattree, R., & Naik, D. N. (2000). Multivariate data reduction and discrimination. *SAS Institute, Cary, North Carolina*.
- Krause, M. D., Dias, K. O., Singh, A. K., & Beavis, W. D. (2022). Using large soybean historical data to study genotype by environment variation and identify mega-environments with the integration of genetic and non-genetic factors. *bioRxiv*, 2022-04.
- Joaquim, A. C., Bellinaso, I. F., Donzelli, J. L., Quadros, A. D., & Barata, M. Q. S. (1994). Potencial e manejo de solos cultivados com cana-de-açúcar. *Seminário Copersucar de Tecnologia Agrônômica*, 6, 1-10.
- Landell, M. G. D. A., Prado, H. D., Vasconcelos, A. C. M. D., Perecin, D., Rossetto, R., Bidoia, M. A. P., ... & Xavier, M. A. (2003). Oxisol subsurface chemical attributes related to sugarcane productivity. *Scientia Agricola*, 60, 741-745.
- Landell, M. D. A., & Bressiani, J. A. (2008). Melhoramento genético, caracterização e manejo varietal. *Cana-de-açúcar*. Campinas: Instituto Agrônômico, 101-155.
- Liaw, A. and Wiener, M. (2002). Classification and Regression by randomForest. *R News* 2(3), 18-22.
- Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: an R package for multivariate analysis. *Journal of statistical software*, 25, 1-18.
- Lin, C. S., & Binns, M. R. (1989). Comparison of unpredictable environmental variation generated by year and by seeding-time factors for measuring type 4 stability. *Theoretical and applied genetics*, 78, 61-64.
- Liu, Y., Zhang, J., & Ge, Q. (2021). The optimization of wheat yield through adaptive crop management in a changing climate: evidence from China. *Journal of the Science of Food and Agriculture*, 101(9), 3644-3653.
- Lobell, D. B., Hammer, G. L., McLean, G., Messina, C., Roberts, M. J., & Schlenker, W. (2013). The critical role of extreme heat for maize production in the United States. *Nature climate change*, 3(5), 497-501. <http://doi.org/10.1038/nclimate1832>.

- Luo, J., Pan, Y. B., Que, Y., Zhang, H., Grisham, M. P., & Xu, L. (2015). Biplot evaluation of test environments and identification of mega-environment for sugarcane cultivars in China. *Scientific reports*, 5(1), 15505. <http://doi.gov/10.1038/srep15505>.
- Luo, T., Liu, X., & Lakshmanan, P. (2023). A combined genomics and phenomics approach is needed to boost breeding in sugarcane. *Plant Phenomics*, 5, 0074.
- Mehareb, E. M. & El-Mansoub, M. M. A. (2020). Genetic parameters and principal components analysis biplot for agronomical, insect and pathological traits in some sugarcane genotypes. *SVU-International Journal of Agricultural Sciences*, 2(2), 60-77. <http://doi.gov/10.21608/svuijas.2020.35913.1017>.
- Mirzawan, P. D. N., Cooper, M., DeLacy, I. H., & Hogarth, D. M. (1994). Retrospective analysis of the relationships among the test environments of the Southern Queensland sugarcane breeding programme. *Theoretical and applied Genetics*, 88, 707-716.
- Mohammadi, R., & Amri, A. (2008). Comparison of parametric and non-parametric methods for selecting stable and adapted durum wheat genotypes in variable environments. *Euphytica*, 159, 419-432.
- DeMorais, L. K., de Aguiar, M. S., de Albuquerque e Silva, P., Câmara, T. M. M., Cursi, D. E., Júnior, A. R. F., & Besspalhok Filho, J. C. (2015). Breeding of sugarcane. *Industrial crops: breeding for bioenergy and bioproducts*, 29-42.
- 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.
- Pipitpukdee, S., Attavanich, W., & Bejranonda, S. (2020). Climate change impacts on sugarcane production in Thailand. *Atmosphere*, 11(4), 408. <http://doi.gov/10.3390/ATMOS11040408>.
- Poudyal, C., Costa, L. F., Sandhu, H., Ampatzidis, Y., Odero, D. C., Arbelo, O. C., & Cherry, R. H. (2022). Sugarcane yield prediction and genotype selection using unmanned aerial vehicle-based hyperspectral imaging and machine learning. *Agronomy Journal*, 114(4), 2320-2333. <https://doi.org/10.1002/agj2.21133>.
- Probst P (2020). *\_varImp: RF Variable Importance for Arbitrary Measures\_*. R package version 0.4, <https://CRAN.R-project.org/package=varImp>.

- R CORE TEAM. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2019.
- R Core Team (2023). *\_R: A Language and Environment for Statistical Computing\_*. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>.
- 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. <https://doi.org/10.1016/j.fcr.2011.07.008>.
- Ridesa, B. (2010). Catálogo nacional de variedades “RB” de cana-de-açúcar. Curitiba: Rede Interuniversitária para o Desenvolvimento do Setor Sucroalcooleiro, 136.
- Soltani, A. (2012). *Modeling physiology of crop development, growth and yield*. CABi.
- Taiz, L., & Zeiger, E. (2007). *Fisiologia vegetal* (Vol. 10). Universitat Jaume I.
- Tena, E., Goshu, F., Mohamad, H., Tesfa, M., Tesfaye, D., & Seife, A. (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. <http://doi.org/10.1080/23311932.2019.1651925>.
- Viana, J. L., de Souza, J. L. M., Hoshide, A. K., de Oliveira, R. A., de Abreu, D. C., & da Silva, W. M. (2023). Estimating Sugarcane Yield in a Subtropical Climate Using Climatic Variables and Soil Water Storage. *Sustainability*, 15(5), 4360. <http://doi.org/10.3390/su15054360>.
- Yan, W., Kang, M. S., Ma, B., Woods, S., & Cornelius, P. L. (2007). GGE biplot vs. AMMI analysis of genotype-by-environment data. *Crop science*, 47(2), 643-653.
- Yan, W., Nilsen, K. T., & Beattie, A. (2022). Mega-environment analysis and breeding for specific adaptation. *Crop Science*, 63(2), 480-494. <https://doi.org/10.1002/csc2.20895>.
- Yan, W., Nilsen, K. T., & Beattie, A. (2023). Mega-environment analysis and breeding for specific adaptation. *Crop Science*, 63(2), 480-494.
- Zamski, E., & Schaffer, A. A. (2017). *Photoassimilate distribution plants and crops source-sink relationships*. Routledge.
- Zhang, P., Zhang, J. & Chen, M. (2017). Economic impacts of climate change on agriculture: The importance of additional climatic variables other than temperature and precipitation.

Journal of Environmental Economics and Management, 83, 8-31. <https://doi.org/10.1016/j.jeem.2016.12.001>.

Zhao, D. & Li, Y. R. (2015). Climate change and sugarcane production: potential impact and mitigation strategies. *International Journal of Agronomy*, 2015, 1-10. <http://doi.org/10.1155/2015/547386>.

Wang, F., Zhao, H., Yu, C., Tang, J., Wu, W., & Yang, Q. (2020). Determination of the geographical origin of maize (*Zea mays* L.) using mineral element fingerprints. *Journal of the Science of Food and Agriculture*, 100(3), 1294-1300.

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. CRC press.

## **CAPÍTULO 4 PREDIÇÃO AMBIENTÔMICA DA PRODUTIVIDADE DA CANA-DE-AÇÚCAR USANDO KERNELS GAUSSIANOS NÃO LINEARES E DADOS CLIMÁTICOS**

### **RESUMO**

A interação genótipo  $\times$  ambiente complexa, é um desafio para os melhoristas, pois genótipos produtivos em um ambiente podem não ser em outros, dificultando a identificação de variedades consistentes em diferentes condições ambientais. Para maximizar os ganhos genéticos, é necessário coletar dados ambientais que influenciam a performance das plantas em diferentes locais de cultivo. Dessa forma, o procedimento *Envirotyping*, que coleta, armazena e processa dados ambientais, e a abordagem *Enviroomics*, que obtém dados em larga escala, são fundamentais para melhorar a precisão na seleção de genótipos e previsões fenotípicas em locais não testados. Outro fator que afeta a qualidade das previsões é a falta de dados fenotípicos, que pode levar a estimativas tendenciosas. Para contornar isso, a simulação de dados fenotípicos amplia a população de treinamento com novos indivíduos. Os objetivos

deste estudo foram avaliar se as previsões seriam afetadas pela inclusão de covariáveis ambientais nos modelos bayesianos e investigar o impacto do aumento do número de genótipos na simulação na redução dos erros de previsão. Dados de toneladas de cana por hectare da fase experimental do programa de melhoramento genético foram usados. A simulação gerou 10 indivíduos em cada uma das 18 classes. As covariáveis ambientais foram obtidas por meio de banco de dados de sensoriamento remoto do Nasa Power, e kernels gaussianos foram posteriormente elaborados. Três estruturas de modelos foram obtidas: MMD sem informação ambiental; EMM e EMD com inclusão de kernels ambientais e sem e com inclusão da  $G \times E$ , respectivamente. Os resultados revelaram que a inclusão de covariáveis ambientais melhorou significativamente a precisão dos modelos, e o aumento do número de genótipos contribuiu para a redução dos erros.

Palavras-chaves: Modelos Bayesianos. Covariáveis ambientais. *Envirotyping*. Simulação.

#### 4.1 INTRODUÇÃO

As safras anteriores da cana-de-açúcar foram marcadas por condições climáticas adversas, como geadas e estiagens, que prejudicaram a colheita em algumas áreas em 2021, além de baixa precipitação e temperaturas reduzidas na região centro-sul em 2022 (CONAB - Companhia Nacional do Abastecimento, 2022/23). No entanto, na terceira estimativa da safra 2023/24, de acordo com o CONAB, as condições climáticas favoráveis, aliadas a maiores investimentos no setor, contribuirão para um aumento significativo da produção. Com isso, o Brasil se destaca como o maior produtor mundial de cana-de-açúcar. Apesar do início tardio da colheita devido às frequentes chuvas, estima-se um aumento de 10,9% na produção em relação à safra anterior, com um total de 677,6 milhões de toneladas (CONAB, 2023/24).

A cana-de-açúcar é uma cultura perene (Toppa et al., 2010), cultivada nas regiões tropicais e subtropicais (Rodrigues, 1995). Seu ciclo vegetativo geralmente se estende por cerca de cinco anos. No primeiro ano, ocorre o plantio, e nos anos subsequentes, é realizada a colheita da rebrota enquanto a produtividade for economicamente viável (Rudorff et al., 2010; Shukla et al., 2017). Durante todo o ciclo vegetativo e estágio fenológico da cultura, a cana-de-açúcar é influenciada por fatores edafoclimáticos e práticas agrícolas (Manhães et al., 2015). Compreender as condições do ambiente de cultivo é crucial, pois elas têm um impacto direto

no desempenho da cultura. Esse conhecimento possibilita aprimorar o manejo agrícola, a seleção adequada de variedades, a predição da produtividade e a otimização das etapas do programa de melhoramento genético de plantas (Scarpari et al., 2004; Scarpari et al., 2009; Scarpare; Caetano et al., 2017).

Uma etapa crucial do programa de melhoramento genético é a fase experimental, na qual os clones de cana-de-açúcar selecionados em etapas anteriores são plantados em diferentes ambientes. Os testes em múltiplos ambientes são fundamentais para avaliar o comportamento das plantas diante de condições ambientais favoráveis ou desfavoráveis ao seu desenvolvimento (Gauch, 1992). No entanto, o rendimento dos clones pode variar significativamente de um local para o outro, evidenciando uma forte interação entre genótipo e ambiente ( $G \times E$ ) (Tena et al., 2019).

A interação  $G \times A$ , altamente complexa, apresenta um desafio significativo para os melhoristas. Os ganhos de seleção são limitados, uma vez que as plantas que se destacam em um ambiente podem não ser as mais produtivas em outros, o que pode dificultar a identificação de variedades superiores que demonstrem uma produtividade consistente em diferentes condições ambientais (Nowosad et al., 2016). Para maximizar os ganhos genéticos em locais diversos, é necessário coletar dados ambientais que influenciam na performance das plantas em diferentes locais de cultivo (Des Marais et al., 2013).

Nesse contexto, *Envirotyping* é uma técnica capaz de coletar, armazenar e processar dados ambientais (Xu et al., 2016; Van Eeuwijk et al., 2018). Assim, a obtenção desses dados em larga escala, sejam eles históricos e/ou em diferentes localizações geográficas, é chamado *Enviromics* (Crossa et al., 2022). O uso dessas técnicas possibilita aumentar a precisão na seleção de genótipos em locais de interesse, e na predição fenotípica onde os ensaios experimentais não foram conduzidos (Resende et al., 2021).

A escolha dos modelos a serem usados na predição da produtividade é fundamental. O uso de modelos de regressão linear e modelos mistos bayesiano tem demonstrado melhorias significativas na precisão das predições para seleção genômica (de los campos et al., 2009; Heslot et al., 2012). Nas análises envolvendo múltiplas características e ambientes, a utilização de modelos bayesianos permite a inclusão de matrizes de covariância diversas (Montesinos-López et al., 2018). Na prática, o emprego de modelos mistos bayesianos

também pode prever as características de difícil mensuração e de alto custo (Crossa et al., 2021).

Outro fator que afeta a qualidade das predições de produtividade é o baixo número de observações. Assim, para contornar a limitação do número insuficiente de genótipos, é possível simular novos indivíduos a partir dos dados existentes. Duas abordagens comuns para essa simulação são baseadas nas classes definidas por Eberhart e Russel (1966) e nas classes proposta por Cruz, Torres e Vencovsky (1989). Após a geração desses novos indivíduos simulados, eles podem ser incorporados ao conjunto de dados para o treinamento do modelo. Essa ampliação do conjunto de dados permite aumentar a confiabilidade dos resultados dos valores preditos (Nascimento et al., 2013).

A utilização de informações provenientes de um ambiente específico para prever o desempenho e recomendar genótipos com características desejáveis em locais não testados tem sido um tema relevante nos programas de melhoramento de plantas (Resende et al., 2021). Este tópico já foi explorado por outros autores, os quais consideram o uso de covariáveis ambientais para ajustar modelos de predição. Essa abordagem tem se mostrado promissora na melhoria da precisão das predições de rendimento em culturas variadas, como evidenciado em diversos estudos (Annicchiarico et al., 2006; Jarquín et al., 2014; Pérez-Rodríguez et al., 2015; Malosetti et al., 2016; Acosta-Pech et al., 2017; Schmidt et al., 2018). A inclusão de informações ambientais em modelos de predição tem aumentado significativamente a precisão das predições da produtividade, chegando a aumentar cerca de 34% da capacidade preditiva dos modelos (Jarquín et al., 2014).

Por fim, mesmo com o crescente número de pesquisas focadas em explorar os ambientes de cultivo detalhadamente para diferentes objetivos, ainda existem lacunas a serem preenchidas para diferentes culturas. Neste estudo, utilizaremos modelos mistos bayesianos para prever a produtividade na cana-de-açúcar, onde ainda há poucos estudos, com aplicação da técnica *enviromics*. Com isso, o objetivo deste estudo foi avaliar se a inclusão de covariáveis ambientais resultaria em uma melhoria no ajuste dos modelos na predição da produtividade (TCH – Tonelada de cana por hectare) da cana-de-açúcar e observar se o aumento no número de genótipos usados na ampliação dos dados resultaria em menores erros.

## 4.2 MATERIAL E MÉTODOS

### 4.2.1 Descrição da área de estudo

Os dados experimentais foram obtidos a partir da fase experimental do programa de melhoramento da cana-de-açúcar da Universidade Federal de Viçosa (PMGCA-UFV). Foram avaliados dez clones e uma variedade (RB 966928 – G11) alocados em oito ambientes, chamadas localidades. Em cada localidade, os experimentos foram delineados em blocos ao acaso, com quatro repetições. As parcelas experimentais foram de quatro sulcos de cinco metros de comprimento, com espaçamento de 1,4 metros entre linhas.

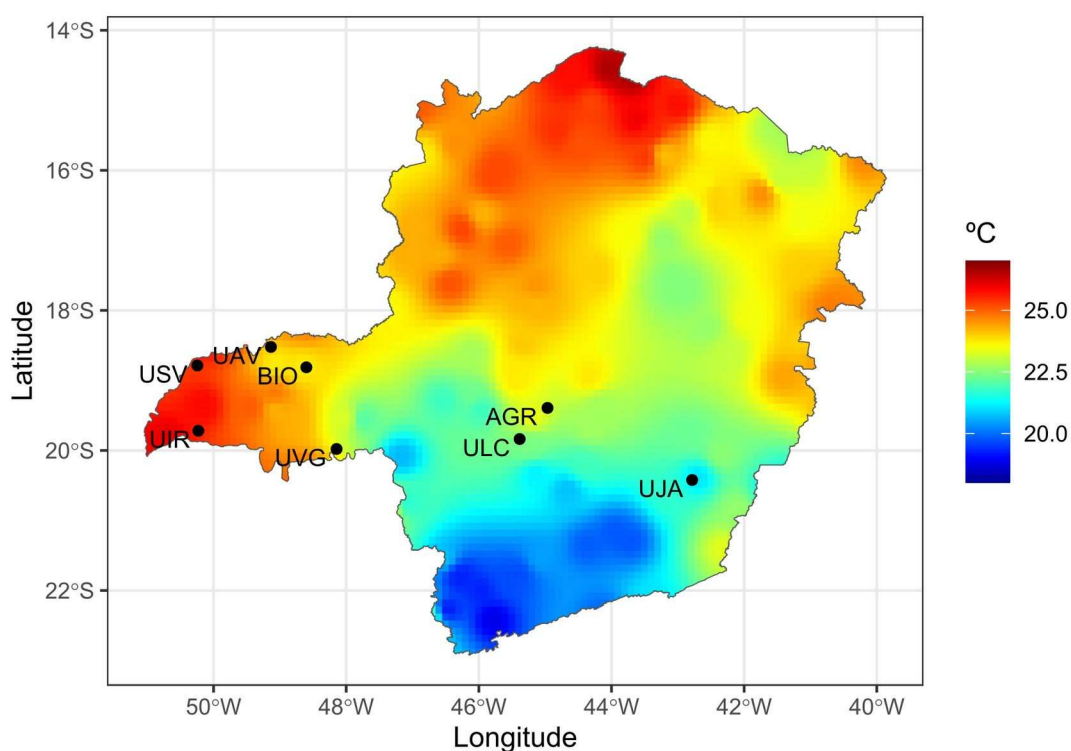


Figura 3 Mapa de temperaturas médias em Minas Gerais (julho de 2020 a agosto de 2021). Os pontos no mapa indicam a localização geográfica dos experimentos realizados nas seguintes usinas: UAV (Usina Alvorada, município de Alceu Garcia), UIR (Usina Coruripe, município de Iturama), AGR (Usina Agropéu, município de Pompéu Velho), BIO (Usina Bioaroeira, município de Uberlândia), USV (Usina Santa Vitória, município de Perdilandia), ULC (Usina Santa Lúcia, município de Moema), UJA (Usina Jatiboca, município de Oratórios) e UVG (Usina Delta Sucroenergia, município de Conceição das Alagoas).

### 4.2.2 Coleta e ampliação de dados de TCH da cana-de-açúcar

Na análise, foram incluídos os registros de TCH do ano de 2021, provenientes de 11 genótipos cultivados em oito localidades distintas, correspondentes à safra de cana soca. Os dados de soca foram utilizados em vez dos de cana planta devido à grande quantidade de

dados faltantes resultante da pandemia. O plantio, realizado em 2019, deveria ser colhido em 2020, mas em muitos locais a colheita não pôde ser executada. Os cálculos seguiram as diretrizes estabelecidas no Manual de Instruções do Conselho dos Produtores de Cana-de-Açúcar, Açúcar e Álcool do Estado de São Paulo (CONSECANA, 2012).

Para a predição de genótipos em locais não testados, são necessário dois conjuntos de dados: o conjunto de treinamento e o de validação. Devido ao número insuficiente de observações para o treinamento dos modelos, foi necessário utilizar um método de ampliação dos dados existentes para expandir o conjunto de treinamento. A simulação de novos indivíduos foi realizada com base nas informações já disponíveis, para compensar o baixo número de genótipos presentes no banco de dados.

A ampliação dos valores de TCH por genótipo e ambientes foi inspirada no método desenvolvido por Barbosa et al. (2023). No entanto, os autores utilizaram o método de adaptabilidade estabilidade de Verma et al. (1978). Neste estudo, foi realizada uma modificação na metodologia, na qual houve uma mudança no método de adaptabilidade e estabilidade utilizado, optando-se pelo proposto por Cruz, Torres e Vencovsky (1989). Esta metodologia, baseada na regressão bissegmentada, compreende os parâmetros de adaptabilidade e estabilidade, que são: a média ( $\hat{\beta}_{0i}$ ) e a resposta linear aos ambientes favoráveis ( $\hat{\beta}_{1i}$ ) e desfavoráveis ( $\hat{\beta}_{1i} + \hat{\beta}_{2i}$ ). Portanto, o modelo  $Y_{ij} = \beta_{0i} + \beta_{1i} I_i + \beta_{2i} T(I_j) + \delta_{ij} + \bar{\epsilon}_{ij}$ , onde,  $\hat{\sigma}_{\delta_i}^2$  representa o desvio da regressão de cada genótipo (Cruz et al., 1999), é utilizado para gerar novos genótipos nas seguintes classes (Table 1):

Tabela 1. Classes de adaptabilidade e estabilidade genotípica, determinadas pelos valores dos parâmetros  $\beta_{1i}$ ,  $\beta_{2i}$  e  $\hat{\sigma}_{\delta_i}^2$  conforme descrito no método de Cruz, Torres and Vencovsky (1989).

Classes	Valores dos parâmetros beta		Adaptabilidade
	Ambiente		
	desfavorável	Favorável	
1	$\beta_{1i} < 1$	$\beta_{1i} + \beta_{2i} < 1$	Específico para ambientes desfavoráveis
2	$\beta_{1i} < 1$	$\beta_{1i} + \beta_{2i} = 1$	Específico para ambientes desfavoráveis
3	$\beta_{1i} < 1$	$\beta_{1i} + \beta_{2i} > 1$	Ideal
4	$\beta_{1i} = 1$	$\beta_{1i} + \beta_{2i} < 1$	Não recomendado

5	$\beta_{1i}=1$	$\beta_{1i}+\beta_{2i}=1$	Geral
6	$\beta_{1i}=1$	$\beta_{1i}+\beta_{2i}>1$	Específico para ambientes favoráveis
7	$\beta_{1i}>1$	$\beta_{1i}+\beta_{2i}<1$	Não recomendado
8	$\beta_{1i}>1$	$\beta_{1i}+\beta_{2i}=1$	Não recomendado
9	$\beta_{1i}>1$	$\beta_{1i}+\beta_{2i}>1$	Específico para ambientes favoráveis

Além disso, os genótipos simulados foram classificados de acordo com sua estabilidade. Para cada combinação de valores do parâmetro  $\beta_{1i}$  e  $\beta_{1i}+\beta_{2i}$  é atribuído um valor para  $\hat{\sigma}_\delta^2$  maior ou igual a zero, a fim de determinar se o genótipo é instável ou estável, respectivamente, resultando em 18 classes de adaptabilidade e estabilidade.

#### 4.2.3 Coleta de dados climáticos e ecofisiológicos

Os dados de clima e ecofisiológicos serão doravante referidos como dados ambientais, os quais foram obtidos ao longo dos anos de 2019, 2020 e 2021, adquiridos por meio de banco de dados de sensoriamento remoto do Nasa Power (<https://power.larc.nasa.gov>). Por meio da função `get_weather` do pacote `EnvRtype` (Costa-Neto et al., 2021) no ambiente de programação R (R Core Team, 2019) é possível fazer o download de dados climáticos de diversas localidades (Costa Neto et al., 2021). Os dados utilizados nesse estudo incluem: irradiação solar na parte superior da atmosfera, irradiação solar incidente em uma superfície horizontal, fluxo radiativo de onda longa descendente médio, medidos em megajoules por metro quadrado ( $\text{MJ m}^{-2} \text{d}^{-1}$ ), velocidade do vento a 10 metros acima da superfície da terra ( $\text{m s}^{-1}$ ), temperatura mínima do ar, temperatura máxima do ar, temperatura de ponto de orvalho, em  $^{\circ}\text{C}$ , umidade relativa do ar (%), precipitação ( $\text{mm d}^{-1}$ ) e elevação (m). Algumas das variáveis que caracterizam os processos ecofisiológicos, como evapotranspiração, déficit de pressão de vapor, inclinação da curva do déficit de pressão de vapor e impacto da temperatura na eficiência da utilização da radiação, amplitude térmica e radiação solar global, foram computadas seguindo os métodos descritos por Soltani & Sinclair (2012).

Os dados ambientais foram utilizados para criar uma matriz de covariáveis ( $W$ ), conforme proposto por Costa-Neto et al. (2021b). Posteriormente, foi aplicado o método Kernel Gaussiano na matriz  $W$  para calcular a similaridade entre os ambientes. Ambas as técnicas foram executadas com o uso do pacote `EnvRtype`, por meio das funções `W_matrix` e

env\_kernel, respectivamente. O método proposto assemelha-se àqueles utilizados na modelagem de efeitos genômicos (Eq. 1) (Costa-Neto et al., 2021a).

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

Em que: O fator de largura de banda (suponha que  $h = 1$  por padrão)  $h$ , é multiplicado pela distância euclidiana  $D_{ii'}^2 = \sum_k (w_{ik} - w_{i'k})^2$  para cada par de elementos  $W = \{w_i, w_{i'}\}$ . Isso implica que a similaridade ambiental é determinada pela distância entre os ambientes, conforme calculadas pelas covariáveis ambientais. A variável escalar  $Q$  representa o quantil usado para ponderar a distância ambiental, assumindo-se  $Q=0,5$ , que equivale ao valor mediano de  $D_{ii'}^2$ .

#### 4.2.4 Descrição dos modelos mistos com o uso da abordagem bayesiana

Após a criação dos kernels de similaridade ambiental, iniciou-se o ajuste de modelos mistos bayesianos, os quais foram realizados utilizando a função kernel\_model do pacote EnvRtype. Essa função emprega um algoritmo de otimização para modelagem bayesiana hierárquica, o mesmo utilizado no pacote BGGE (*Bayesian Genomic Genotype × Environment*). Os detalhes teóricos sobre o pacote e o algoritmo bayesiano estão disponíveis em Granato et al. (2018) e Costa-Neto et al. (2021b). O modelo descrito (Eq. 2) é uma estrutura genérica que pode ser aplicada a diversas combinações de kernels, visando modelar a variação fenotípica em diferentes localidades.

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

Em que: o vetor  $y$  combina as médias de cada genótipo em cada ambiente, com o escalar  $\mu$  representando a média geral.  $X_f$  denota a matriz de incidência associada ao vetor de efeitos fixos  $\beta$ . O vetor aleatório para efeito genotípico ( $g_s$ ) varia de 1 a  $k$  e o efeito baseado no ambiente ( $w_r$ ) varia de 1 a  $l$  variáveis ambientais, e  $\varepsilon$  é o resíduo.

Neste estudo, foram ajustadas três estruturas de modelos distintas aos dados. O modelo MMD (main genomic effects (MM) + single G×E deviation), que considera exclusivamente

efeitos genotípicos, em que  $\sum_{s=1}^p g_s \neq 0$  e  $\sum_{r=1}^q w_r = 0$ , onde  $g_s$  está associado ao efeito principal do genótipo mais o efeito da interação  $G \times A$ ,  $\sum \dot{g}_s \neq 0$  (Jarquín et al., 2014). Com a inclusão de variáveis ambientais, os modelos consideram  $\sum_{s=1}^p g_s \neq 0$  e  $\sum_{r=1}^q w_r \neq 0$ , onde  $g_s$  está relacionado a  $G$  ou a  $G+G \times E$ , resultando nos modelos EMM (MM + *main environmic effects*) e EMDs (EMM + *single G × A deviation*), respectivamente. O  $w_r$  são os efeitos principais do ambiente (Souza et al., 2017).

O conjunto de treinamento e validação foi organizado em três cenários distintos. Inicialmente, oito genótipos foram retirados do conjunto de dados para serem utilizados na validação. A partir dos três restantes, foram gerados dez genótipos em cada classe, totalizando 18 classes, que foram utilizadas no treinamento (C3). Em seguida, no segundo cenário, seis genótipos foram removidos para validação, enquanto os cinco restantes foram usados na ampliação e utilizados no treinamento (C5). No terceiro cenário, quatro genótipos foram removidos para validação, e a ampliação foi realizada a partir dos genótipos restantes, os quais foram utilizados no treinamento (C7). Dessa forma, o número total de genótipos utilizados no treinamento foi de 183, 185 e 187, para os cenários C3, C5 e C7, respectivamente.

#### 4.2.5 Análise estatística

A relação entre os valores observados e estimados da produtividade da cana-de-açúcar foi verificada por meio da raiz do erro quadrático médio (RMSE), expressa em tonelada métrica por hectare:

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

Em que  $\hat{Y}_i$  e  $Y_i$  o  $i$ -ésimo valor da variável estimada e observada, respectivamente,  $n$  é o número amostral. Se os valores forem próximos ou iguais a zero, isso indica que o modelo se ajusta perfeitamente aos dados (Mangiafico, 2016).

Além disso, foi calculado um Índice de Coincidência (IC) segundo o método de Hamblin e Zimmerman (1986). O IC é calculado pela fórmula:  $IC = (A \times C) / (B \times C) \times 100$ , em que: A é o

número de plantas que coincidem em ambos os conjuntos de dados, B número de genótipos selecionados e C é o número de coincidência aleatória atribuída. Neste estudo, B foi definido como 10%. Os dados originais e preditos foram ranqueados em ordem decrescente, e os cinco e oito clones com as maiores médias foram selecionados em cada conjunto de dados. Em seguida, foi verificado se esses mesmos clones estavam presentes na seleção tanto dos dados observados quanto dos preditos.

#### 4.3 RESULTADOS

A raiz do erro quadrático médio (RMSE) foi utilizada para avaliar quais modelos se ajustaram melhor aos dados de TCH. Ao observar os valores de RMSE (Figura 3), percebe-se que, no geral, o modelo MMD (sem informação ambiental) apresentou valores mais altos em comparação com o EMM e o EMD, ambos enriquecidos com informações do ambiente.

Ao analisar os valores de RMSE no modelo MMD, um aspecto relevante é a presença de valores extremamente altos nas localidades BIO e UJA. Nessas localidades, os valores de RMSE ultrapassaram  $50 \text{ t ha}^{-1}$  em todos os cenários de validação, especialmente na localidade BIO, onde esses valores foram superiores a  $100 \text{ t ha}^{-1}$ . Além disso, outras localidades, como UVG e UAV, apresentaram valores de RMSE acima de  $34 \text{ t ha}^{-1}$ . Em contraste, a localidade AGR, que registrou o quarto menor valor de RMSE, apresentou um valor máximo de  $27 \text{ t ha}^{-1}$ , o que representa uma diferença de aproximadamente  $14 \text{ t ha}^{-1}$  entre elas. Essa diferença é significativa e indica uma variação considerável nos valores de RMSE entre as localidades. Seguindo essas, as localidades USV, UIR e ULC apresentaram-se com valores de RMSE subsequentemente menores (Figure 2).

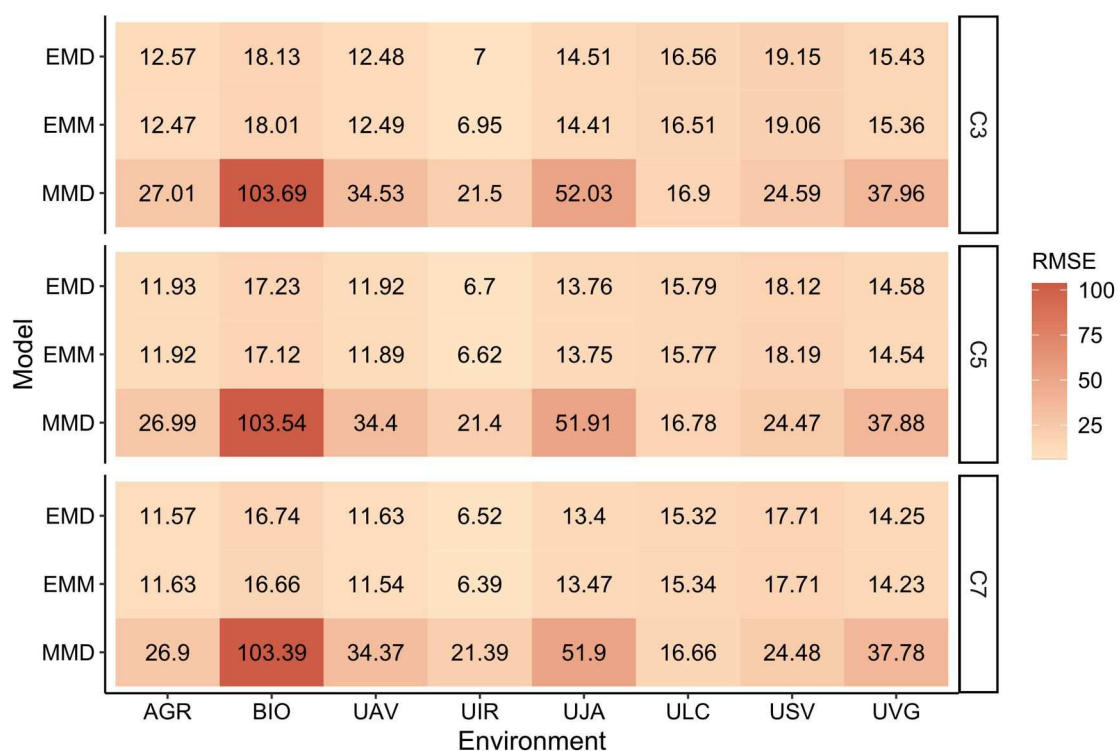


Figura 4. Valores de RMSE (Erro Quadrático Médio da Raiz) em  $t\ ha^{-1}$  para cada modelo (MMD, EMM e EMD) e cenário de validação (C3, C5 e C7) em cada localidade. A coloração varia de marrom-avermelhado mais claro a mais escuro, representando, respectivamente, os menores e maiores valores de RMSE.

Dado que os valores de RMSE do modelo MMD estavam altos, optou-se por removê-lo do próximo gráfico para melhorar a apresentação da escala. Assim, torna-se mais visível a diferença de RMSE nos três diferentes cenários de validação. Conforme aumenta o número de genótipos a serem ampliados e posteriormente utilizados no treinamento, os valores de RMSE diminuem. É importante observar também que os valores variam em cada ambiente; no entanto, eles apresentam um comportamento semelhante ao aumentar o número de genótipos na ampliação.

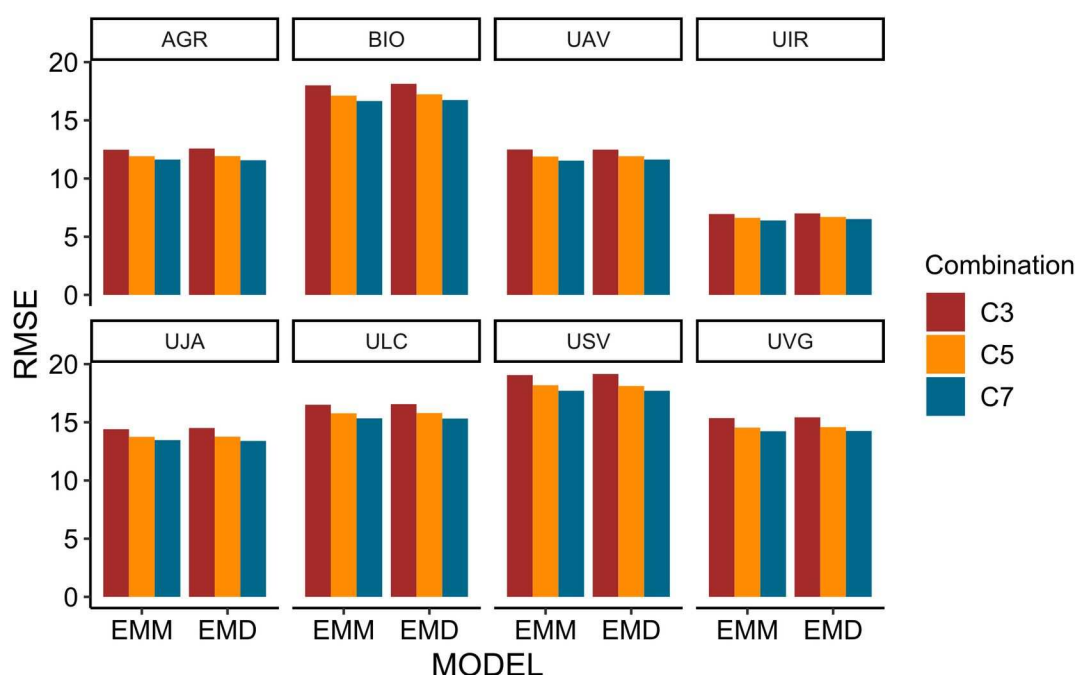


Figura 5. Gráfico de barras ilustrando os valores de RMSE para cenários de validação (C3, C5 e C7) e modelos EMM e EMD em diferentes localidades.

Com a inclusão de covariáveis ambientais nos modelos de predição os valores de RMSE nas localidades BIO e UJA foram drasticamente reduzidos (Figura 3). Assim, ao comparar os modelos EMM e EMD os erros se tornam mais semelhantes entre as localidades, e a variação entre modelos EMM e EMD não é tão evidente.

Ao avaliar os modelos EMM e EMD, é possível observar uma variação nos valores de RMSE entre os ambientes. A localidade ULC apresentou valores consistentes de RMSE, independente do modelo, com valores variando de  $15,32 \text{ t ha}^{-1}$  a  $16,9 \text{ t ha}^{-1}$ , que corresponde a aproximadamente 15% da média de TCH, a depender do cenário de validação. No entanto, as localidades UIR, UAV e AGR apresentaram os menores erros com a inclusão de covariáveis ambientais (Figura 3 e 4).

Ao avaliar o Índice de Coincidência (IC) com cinco genótipos selecionados, o IC variou de 40% a 80% dependendo da localidade. Na maioria dos locais, o IC foi de 60%, e na localidade UVG chegou a 80% (Tabela 2). Especificamente, nos locais AGR, BIO, UAV, UIR e ULC, três dos cinco genótipos de maior média coincidiram entre os conjuntos de dados observados e preditos, indicando um IC de 60%. Nos locais UJA e USV, apenas dois genótipos coincidiram, resultando em um IC de 40%. Por outro lado, na localidade UVG, quatro genótipos coincidiram, alcançando um IC de 80%.

Tabela 2 Ranqueamento dos cinco clones de cana-de-açúcar com maior média em cada localidade, para os valores observados e preditos.

AGR		BIO		UAV		UIR		
Rank	Observado	Predito	Observado	Predito	Observado	Predito	Observado	Predito
1	<u>G7</u>	<b>G11</b>	<u>G4</u>	<u>G11</u>	<b>G11</b>	<u>G10</u>	<b>G11</b>	<b>G11</b>
2	<u>G10</u>	<b>G5</b>	<b>G5</b>	<b>G6</b>	<u>G7</u>	<b>G11</b>	<u>G4</u>	<u>G6</u>
3	<b>G11</b>	<u>G8</u>	<b>G10</b>	<u>G7</u>	<b>G5</b>	<b>G6</b>	<b>G10</b>	<b>G5</b>
4	<b>G6</b>	<b>G6</b>	<u>G1</u>	<b>G10</b>	<b>G6</b>	<b>G5</b>	<u>G8</u>	<b>G10</b>
5	<b>G5</b>	<u>G4</u>	<b>G6</b>	<b>G5</b>	<u>G8</u>	<u>G4</u>	<b>G5</b>	<u>G7</u>
IC (%)	60		60		60		60	
UJA		ULC		USV		UVG		
Rank	Observado	Predito	Observado	Predito	Observado	Predito	Observado	Predito
1	<u>G5</u>	<b>G11</b>	<u>G9</u>	<b>G11</b>	<u>G9</u>	<u>G6</u>	<b>G11</b>	<u>G5</u>
2	<b>G6</b>	<u>G10</u>	<b>G6</b>	<u>G10</u>	<u>G4</u>	<b>G11</b>	<u>G8</u>	<b>G11</b>
3	<b>G11</b>	<u>G7</u>	<b>G5</b>	<b>G5</b>	<b>G10</b>	<u>G5</u>	<b>G6</b>	<b>G6</b>
4	<u>G8</u>	<b>G6</b>	<b>G11</b>	<u>G8</u>	<b>G11</b>	<u>G7</u>	<b>G11</b>	<b>G10</b>
5	<u>G9</u>	<u>G4</u>	<u>G7</u>	<b>G6</b>	<u>G8</u>	<b>G10</b>	<b>G10</b>	<u>G9</u>
IC (%)	40		60		40		80	

#### 4.4 DISCUSSÃO

A predição eficaz da produtividade de clones em diferentes ambientes é um desafio crucial no melhoramento genético da cana-de-açúcar (Resende et al., 2021). Modelos mistos bayesianos oferecem uma abordagem robusta para lidar com a complexidade das interações G×A e a variabilidade ambiental. Este estudo compara três estruturas de modelos mistos bayesianos diferentes em cada localidade utilizada na fase experimental do programa de melhoramento genético. No modelo MMD, foram consideradas apenas as informações

fenotípicas, ou seja, os dados de produtividade (TCH) da cana-de-açúcar, com a  $G \times A$ . Nos modelos EMM e EMD, foi inserida a matriz de covariância ambiental, sem considerar a  $G \times A$  e incluindo-a, respectivamente.

Ao avaliar os valores de RMSE nas localidades BIO e UJA sem incluir as covariáveis ambientais, observa-se, em comparação com as demais localidades, um valor muito discrepante. Uma das razões para que esses valores altos, particularmente na localidade UJA, é o ambiente de cultivo. Nesse local, as temperaturas baixas e a baixa radiação, além de grandes volumes de chuvas em períodos pontuais durante o estágio de crescimento das plantas, representam um cenário oposto ao observado nas outras localidades, onde as temperaturas e outras variáveis correlacionadas foram mais altas, e a precipitação foi mais bem distribuída ao longo do ciclo vegetativo da cana-de-açúcar. Além disso, a localidade UJA apresenta médias de TCH significativamente inferiores às dos demais ambientes, o que pode ter comprometido a precisão da predição do modelo. Essas baixas médias podem estar relacionadas a fatores limitantes da produção, como as características do solo, conforme conhecimento prático do local. (Supplementary figure 1).

Este resultado destaca a importância da similaridade entre ambientes de cultivo para a predição de produtividade em locais não testados, neste estudo, para condição climática e as variáveis ecofisiológicas (Jarquin et al., 2021; Neyhart et al., 2022). No entanto, é crucial ressaltar que a similaridade entre os locais também depende de outros fatores além do clima, como as características do solo, que devem ser explorados em estudos futuros.

Na localidade BIO, mesmo com condições climáticas semelhantes às dos outros ambientes, os erros de predição foram muito altos devido aos valores de TCH serem significativamente discrepantes em relação aos demais locais. Isso pode ser atribuído à maior fertilidade dos solos nessa localidade e à elevadas altitudes (900 metros), conforme conhecimentos práticos sobre a área. Altitudes elevadas podem contribuir para um aumento na radiação solar (Blumthaler et al., 1997), potencialmente aumentando a eficiência fotossintética da planta. Essa condição é vantajosa durante o estágio de crescimento da cana-de-açúcar (Taiz e Zeiger, 2007).

A inclusão de covariáveis ambientais nos modelos de predição de produtividade mostrou-se promissora. O ajuste do modelo melhorou significativamente para todas as localidades; algumas apresentaram mudanças mais drásticas, o que sugere que as covariáveis ambientais

contribuíram para melhorar a predição na maioria dos ambientes, exceto na localidade ULC, onde as variações foram bem reduzidas. Alguns autores, ao utilizar poucas covariáveis ambientais (duas, três), conseguiram obter predições precisas (Li et al., 2018; Millet et al., 2019). Isso indica que, em ambientes heterogêneos onde a interação  $G \times A$  é alta e deve ser considerada nos modelos preditivos, um maior número de covariáveis ambientais é necessário. Em contraste, em ambientes mais homogêneos, onde  $G \times A$  é mínima e negligenciável, variáveis ambientais adicionais podem não ser necessárias para modelar o desempenho diferencial dos genótipos (Neyhart et al., 2022).

É fundamental conhecer as variáveis ambientais que afetam o desenvolvimento do genótipo em cada ambiente de teste. As localidades onde são cultivadas a cana-de-açúcar são classificadas segundo ao seu potencial produtivo, esse sistema leva em consideração as características físico-químicas dos solos e as épocas de corte, especialmente o tempo de maturação da cana-de-açúcar (Joaquim, 1997). Dessa forma, sugere-se que, em análises subsequentes, seja desenvolvido um modelo de predição específico para cada ambiente de produção, com inclusão das variáveis climáticas para aprimorar a classificação desses ambientes.

Ao considerar as particularidades de cada localidade em um programa de melhoramento, onde o objetivo é selecionar clones promissores, é preciso analisar o ranqueamento dos genótipos em cada ambiente. Para avaliar a consistência desse ranqueamento no conjunto de dados preditos, foi analisado o IC para cada localidade. Observa-se uma variação do IC entre as localidades, sem relação com os valores de RMSE; ou seja, locais com maiores valores de RMSE não apresentaram necessariamente maior IC. Essa variação no IC pode ser atribuída a fatores ambientais e genéticos que afetam o desempenho dos genótipos em diferentes locais. A alta coincidência observada na localidade UVG (80%) sugere uma maior estabilidade e previsibilidade dos genótipos neste ambiente específico. Por outro lado, a menor coincidência em UJA e USV pode indicar maior variabilidade nas condições locais ou na resposta dos genótipos.

De acordo com Jarquín et al. (2013), o uso da predição em ambientes não testados pode ser mais eficaz em muitos ambientes onde condições ambientais são mais similares. No entanto, este estudo foi conduzido apenas em oito localidades e com onze genótipos. Como alternativa para aumentar a confiabilidade dos dados dos indivíduos, foi feita a ampliação dos dados originais. Conforme o número de genótipos usados na ampliação aumentava, observou-

se uma melhoria consistente no ajuste. A simulação dos dados mostrou-se eficiente para esse estudo. Com a inclusão de novos indivíduos, foi possível obter uma consistência nos dados de produtividade, mesmo com uma redução pequena do erro, de aproximadamente  $0,10 \text{ t ha}^{-1}$ , à medida que o número de genótipos na ampliação aumentava.

Outra questão sugerida por alguns autores é que os dados climáticos, ecofisiológicos e de solo são mais eficazes quando resumidos em estádios fenológicos das plantas, especialmente para variáveis que quantificam a sacarose, já que seu acúmulo é mais sensível às variações ambientais. Esse enfoque permite identificar os fatores ambientais que influenciam cada fase de desenvolvimento, gerando resultados mais precisos (Heslot et al., 2014; Monteverde et al., 2019; Rincent et al., 2019). No entanto, nesse estudo, não foi realizado essa abordagem, pois foram utilizados dados da cana soca, que apresenta diferentes épocas de cortes da cana planta (junho e julho). Utilizar uma única data poderia comprometer a confiabilidade dos resultados.

#### 4.5 CONCLUSÃO

A inclusão de covariáveis ambientais demonstrou ser promissora para aprimorar a precisão dos modelos na previsão da produtividade da cana-de-açúcar (TCH). Além disso, o aumento no número de genótipos empregados na ampliação dos dados resultou em uma redução significativa nos erros, conforme evidenciado pela diminuição do RMSE.



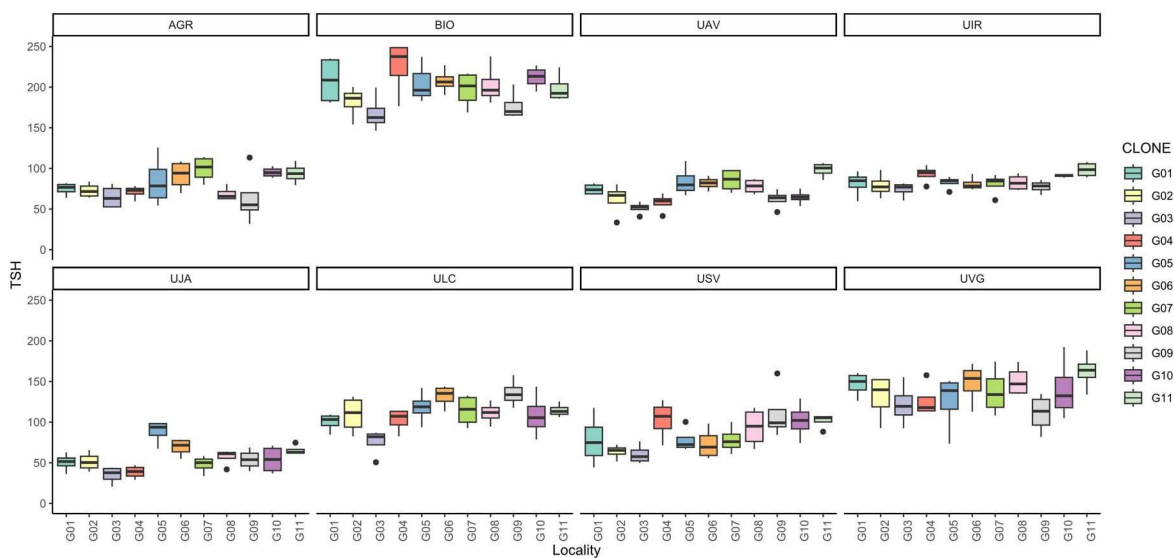


Figura Suplementar 2. Box plot da variável TCH (t ha<sup>-1</sup>) nos experimentos da fase experimental do programa de melhoramento, conduzidos em diferentes locais. Cada caixa no gráfico mostra o limite inferior (1º quartil, 25%), a mediana (50%, representada por uma linha horizontal em negrito) e o limite superior (3º quartil, 75%), enquanto os pontos indicam outliers.

#### 4.7 REFERÊNCIA BIBLIOGRÁFICA

- Acosta-Pech, R., Crossa, J., de Los Campos, G., Teyssèdre, S., Claustres, B., Pérez-Elizalde, S., & Pérez-Rodríguez, P. (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., Bellah, F., & Chiari, T. (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., & da Silveira, L. C. I. (2015). Breeding program and cultivar recommendations. In *Sugarcane* (pp. 241-255). Academic Press.
- Barbosa, Wagner Faria et al. Artificial neural networks based on segmented model for adaptability and stability evaluation of soybean genotypes. *Australian Journal of Crop Science*, v. 17, n. 9, p. 735-740, 2023.
- Blumthaler, M., Ambach, W., & Ellinger, R. (1997). Increase in solar UV radiation with altitude. *Journal of photochemistry and Photobiology B: Biology*, 39(2), 130-134.
- Caetano, J. M., & Casaroli, D. (2017). Sugarcane yield estimation for climatic conditions in the state of Goiás. *Revista Ceres*, 64, 298-306.
- CONAB - COMPANHIA NACIONAL DE ABASTCIMENTO. Acompanhamento da safra brasileira de cana-de-açúcar safra 2022/23. 4º levantamento, Brasília, DF, v. 10, n. 4 abril 2023.
- CONAB - COMPANHIA NACIONAL DE ABASTCIMENTO. Acompanhamento da safra brasileira de cana-de-açúcar safra 2023/24. 3º levantamento, Brasília, DF, v. 11, n. 3 novembro 2023.
- 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., Galli, G., Carvalho, H. F., Crossa, J., & Fritsche-Neto, R. (2021a). EnvRtype: a software to interplay enviromics and quantitative genomics in agriculture. *G3*, 11(4), jkab040.

- Costa-Neto, G., Fritsche-Neto, R., & Crossa, J. (2021b). Nonlinear kernels, dominance, and envirotyping data increase the accuracy of genome-based prediction in multi-environment trials. *Heredity*, 126(1), 92-106.
- Crossa, J., Montesinos-Lopez, O. A., Pérez-Rodríguez, P., Costa-Neto, G., Fritsche-Neto, R., Ortiz, R., & Rincent, R. (2022). Genome and environment based prediction models and methods of complex traits incorporating genotype $\times$  environment interaction. *Genomic Prediction of Complex Traits: Methods and Protocols*, 245-283.
- Crossa, J., Fritsche-Neto, R., Montesinos-Lopez, O. A., Costa-Neto, G., Dreisigacker, S., Montesinos-Lopez, A., & Bentley, A. R. (2021). The modern plant breeding triangle: optimizing the use of genomics, phenomics, and enviromics data. *Frontiers in plant science*, 12, 651480.
- De Los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., ... & Cotes, J. M. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics*, 182(1), 375-385.
- Cruz, C. D., Torres, R. D. A., & Vencovsky, R. (1989). An alternative approach to the stability analysis proposed by Silva and Barreto.
- Cruz, C. D., Regazzi, A. J., Carneiro, P. C. S. (2004). *Modelos biomáticos aplicados ao melhoramento genético*, Ed. 3. Viçosa, 480.
- de Los Campos, G., Pérez-Rodríguez, P., Bogard, M., Gouache, D., & Crossa, J. (2020). A data-driven simulation platform to predict cultivars' performances under uncertain weather conditions. *Nature communications*, 11(1), 4876.
- des Marais, D. L., Hernandez, K. M., & Juenger, T. E. (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., & Zhou, M. (2024). Predicting Ratooning Ability of Sugarcane Varieties in Selection Trials. *Sugar Tech*, 26(1), 52-62.
- Eberhart, S. T., & Russell, W. (1966). Stability parameters for comparing varieties 1. *Crop science*, 6(1), 36-40.
- Gauch, H. J. (1992). Statistical analysis of regional yield trials: AMMI analysis of factorial designs (pp. 278).

- Hayes, B. J., Wei, X., Joyce, P., Atkin, F., Deomano, E., Yue, J., ... & Voss-Fels, K. P. (2021). Accuracy of genomic prediction of complex traits in sugarcane. *Theoretical and Applied Genetics*, 134, 1455-1462.
- Heslot, N., Yang, H. P., Sorrells, M. E., & Jannink, J. L. (2012). Genomic selection in plant breeding: a comparison of models. *Crop science*, 52(1), 146-160.
- Heslot, N., Akdemir, D., Sorrells, M. E., & Jannink, J. L. (2014). Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theoretical and applied genetics*, 127, 463-480.
- Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., ... & de los Campos, G. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and applied genetics*, 127, 595-607.
- Joaquim, A. C., Bellinaso, I. F., Donzelli, J. L., Quadros, A. D., & Barata, M. Q. S. (1994). Potencial e manejo de solos cultivados com cana-de-açúcar. *Seminário Copersucar de Tecnologia Agronômica*, 6, 1-10.
- Li, X., Guo, T., Mu, Q., Li, X., & Yu, J. (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., Appunu, C., Aitken, K., Suresha, G. S., Vignesh, P., Mahadeva Swamy, H. K., ... & Ram, B. (2021). Genomic selection in sugarcane: Current status and future prospects. *Frontiers in Plant Science*, 12, 708233.
- Manhães, C. M. C., Garcia, R. F., Francelino, F. M. A., de Oliveira Francelino, H., & Coelho, F. C. (2015). Fatores que afetam a brotação e o perfilhamento da cana-de-açúcar. *Revista Vértices*, 17(1), 163-181.
- Mangiafico, S.S. Summary and Analysis of Extension Program Evaluation in R, Version 1.18.1. 2016. Available online: <https://rcompanion.org/handbook/> (accessed on 10 January 2021).

- Medar, R. A., Rajpurohit, V. S., & Ambekar, A. M. (2019). Sugarcane crop yield forecasting model using supervised machine learning. *International Journal of Intelligent Systems and Applications*, 11(8), 11.
- Millet, E. J., Kruijer, W., Coupel-Ledru, A., Alvarez Prado, S., Cabrera-Bosquet, L., Lacube, S., ... & Tardieu, F. (2019). Genomic prediction of maize yield across European environmental conditions. *Nature genetics*, 51(6), 952-956.
- Miphokasap, P., & Wannasiri, W. (2018). Estimations of nitrogen concentration in sugarcane using hyperspectral imagery. *Sustainability*, 10(4), 1266.
- Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Gianola, D., Hernández-Suárez, C. M., & Martín-Vallejo, J. (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., Gutierrez, L., Blanco, P., Pérez de Vida, F., Rosas, J. E., Bonnacarrère, V., ... & McCouch, S. (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.
- Neyhart, J. L., Silverstein, K. A., & Smith, K. P. (2022). Accurate predictions of barley phenotypes using genomewide markers and environmental covariates. *Crop Science*, 62(5), 1821-1833.
- Pagani, V., Stella, T., Guarneri, T., Finotto, G., Van den Berg, M., Marin, F. R., ... & Confalonieri, R. (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., Crossa, J., Bondalapati, K., De Meyer, G., Pita, F., & Campos, G. D. L. (2015). A pedigree-based reaction norm model for prediction of cotton yield in multi-environment trials. *Crop Science*, 55(3), 1143-1151.
- Poudyal, C., Costa, L. F., Sandhu, H., Ampatzidis, Y., Odero, D. C., Arbelo, O. C., & Cherry, R. H. (2022). Sugarcane yield prediction and genotype selection using unmanned aerial vehicle-based hyperspectral imaging and machine learning. *Agronomy Journal*, 114(4), 2320-2333.

- Resende, R. T., Piepho, H. P., Rosa, G. J., Silva-Junior, O. B., e Silva, F. F., de Resende, M. D. V., & Grattapaglia, D. (2021). Enviromics in breeding: applications and perspectives on envirotypic-assisted selection. *Theoretical and Applied Genetics*, 134, 95-112.
- Rincent, R., Malosetti, M., Ababaei, B., Touzy, G., Mini, A., Bogard, M., ... & Van Eeuwijk, F. (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., Cezar, E., dos Santos, G. L. A. A., Reis, A. S., Furlanetto, R. H., de Oliveira, R. B., & Nanni, M. R. (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*, 419-449.
- Rudorff, BFT, Aguiar, DA, Silva, WF, Sugawara, LM, Adami, M., & Moreira, MA (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., Jones, M., Marin, F., Ruane, A., & Thorburn, P. (2014). Predicting climate change impacts on sugarcane production at sites in Australia, Brazil and South Africa using the Cane-gro model. *Sugar Tech*, 16, 347-355.
- Scarpore, F. V., van Lier, O. D. J., Corrêa, S. T. R., Hugo, A., Barros, C., Marin, F. R., & Nassif, D. S. P. (2012). MODELOS DE CRESCIMENTO DA CANA-DE-AÇÚCAR E SUA PARAMETRIZAÇÃO. *Revista de Agricultura*, 87(1), 66-80.
- Scarpari, M. S., & Beauclair, E. G. F. D. (2004). Sugarcane maturity estimation through edaphic-climatic parameters. *Scientia Agricola*, 61, 486-491.
- Schmidt, P., Möhring, J., Koch, R. J., & Piepho, H. P. (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., Sharma, L., Awasthi, S. K., & Pathak, A. D. (2017). Sugarcane in India. Package of practices for different agro-climatic zones, All Indian Coordinated Research Project on Sugarcane, IISR Lucknow, Uttar Pradesh, 1-64.

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

Taiz, L., & Zeiger, E. (2007). *Fisiologia vegetal* (Vol. 10). Universitat Jaume I.

Tena, E., Goshu, F., Mohamad, H., Tesfa, M., Tesfaye, D., & Seife, A. (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., Jadoski, C. J., Julianetti, A., Hulshof, T., & Ono, E. O. (2010). Physiology development in the vegetative stage of sugarcane. *Applied Research & Agrotechnology*, 3(2), 169-186. Jadoski et al., 2010;

Verma MM, Chahal GS, Murty BR (1978) Limitations of conventional regression analysis a proposed modification. *Theor Appl Genet.* 53: 89–91.

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

Viana, J. L., de Souza, J. L. M., Hoshide, A. K., de Oliveira, R. A., de Abreu, D. C., & da Silva, W. M. (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.

## CAPÍTULO 5 CONCLUSÕES GERAIS

- A remoção de genótipos na análise de biplots GGE afeta a formação de mega-ambientes. Genótipos instáveis, sensíveis a variações ambientais, apresentam desempenho variável, perdendo-se informações importantes sobre a interação  $G \times A$ . Por outro lado, genótipos estáveis têm um efeito menor na formação de mega-ambientes, mostrando consistência em diferentes condições. A definição de mega-ambientes baseada apenas no desempenho das variedades pode ser inadequada para recomendações futuras em programas de melhoramento.
- Foi identificado padrões de comportamento entre variáveis climáticas em diferentes localidades, delineando mega-ambientes com características climáticas semelhantes. A formação dos mega-ambientes mostrou consistência ao longo do estágio fenológico das plantas. As variáveis climáticas influenciam de forma diversa o desempenho das variedades de cana-de-açúcar, com temperatura e umidade relativa do ar sendo os principais fatores contribuintes para o ajuste do modelo em todas as variedades.
- A inclusão de covariáveis ambientais demonstrou ser promissor para aprimorar a precisão dos modelos na previsão da produtividade da cana-de-açúcar (TCH). Além disso, o aumento no número de genótipos empregados na ampliação dos dados resultou em uma redução significativa nos erros, conforme evidenciado pela diminuição do RMSE.