UFG
UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS

USO DE COVARIÁVEIS AMBIENTAIS PARA PREDIÇÃO DA INTERAÇÃO DE CLONES DE CANA-DE-AÇÚCAR COM AMBIENTES

LARYSSA BARBOSA DE SOUZA REIS

Orientadora:
Profª. Bruna Mendes de Oliveira

Coorientador:
Prof. Márcio Lisboa Guedes

Goiânia - GO
Brasil

Fevereiro – 2025

UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE AGRONOMIA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

[X] Dissertação [ ] Tese [ ] Outro*:_____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Laryssa Barbosa de Souza Reis

3. Título do trabalho

USO DE COVARIÁVEIS AMBIENTAIS PARA PREDIÇÃO DA INTERAÇÃO DE CLONES DE CANA-DE-AÇÚCAR COM AMBIENTES

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento [X] SIM [ ] NÃO[1]

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:
a) consulta ao(à) autor(a) e ao(à) orientador(a);
b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.
O documento não será disponibilizado durante o período de embargo.
Casos de embargo:
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.

Documento assinado eletronicamente por **Bruna Mendes De Oliveira**, **Professor do Magistério Superior**, em 27/03/2025, às 15:31, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020 .

Documento assinado eletronicamente por **Laryssa Barbosa De Souza Reis**, **Discente**, em 27/03/2025, às 15:35, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5269643** e o código CRC **946D4A7A**.

Referência: Processo nº 23070.011436/2025-62 SEI nº 5269643

LARYSSA BARBOSA DE SOUZA REIS

USO DE COVARIÁVEIS AMBIENTAIS PARA PREDIÇÃO DA INTERAÇÃO DE CLONES DE CANA-DE-AÇÚCAR COM AMBIENTES

Dissertação apresentada ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas, da Escola de Agronomia, da Universidade Federal de Goiás, como requisito para obtenção do título de Mestre em Genética e Melhoramento de Plantas.
Área de concentração: Genética e Melhoramento de Plantas
Linha de Pesquisa: Melhoramento de espécies cultivadas

Orientadora:
Profª. Bruna Mendes de Oliveira

Coorientador:
Prof. Márcio Lisboa Guedes

Goiânia, GO – Brasil
2025

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Reis, Laryssa Barbosa de Souza
Uso de covariáveis ambientais para predição da interação de clones de cana-de-açúcar com ambientes [manuscrito] / Laryssa Barbosa de Souza Reis. - 2025.
LX, 60 f.: il.

Orientador: Profa. Dra. Bruna Mendes de Oliveira; co-orientador Dr. Márcio Lisboa Guedes.
Dissertação (Mestrado) - Universidade Federal de Goiás, Escola de Agronomia (EA), Programa de Pós-graduação em Genética e Melhoramento de Plantas, Goiânia, 2025.
Bibliografia. Anexos. Apêndice.

1. Melhoramento genético. 2. Modelagem preditiva. 3. Ambiente. I. Oliveira, Bruna Mendes de, orient. II. Título.

CDU 633

UNIVERSIDADE FEDERAL DE GOIÁS

ESCOLA DE AGRONOMIA

ATA DE DEFESA DE DISSERTAÇÃO

Ata nº 132 da sessão de Defesa de Dissertação de Laryssa Barbosa de Souza Reis, que confere o título de Mestra em Genética e Melhoramento de Plantas, na área de concentração em Genética e Melhoramento de Plantas.

Aos vinte e sete dias do mês de fevereiro de dois mil e vinte e cinco, a partir das quatorze horas, on-line, realizou-se a sessão pública de Defesa de Dissertação intitulada "Predição da produtividade de clones de cana-de-açúcar com o uso de variáveis ambientais". Os trabalhos foram instalados pela Orientadora, Professora Doutora Bruna Mendes de Oliveira (PPGGMP/UFG), com a participação dos demais membros da Banca Examinadora: Doutor Rafael Tassinari Resende (PPGGMP/UFG), membro titular interno, Doutor Márcio Lisboa Guedes (UVA/Ceará), membro titular externo e Doutor Filipe Couto Alves (BASF/USA), membro titular externo. Durante a arguição os membros da banca fizeram sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido a candidato **Aprovada** pelos seus membros. Proclamados os resultados pela Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora.

TÍTULO SUGERIDO PELA BANCA

Uso de covariáveis ambientais para predição da interação de clones de cana-de-açúcar com ambientes

Documento assinado eletronicamente por **Filipe Couto Alves**, **Usuário Externo**, em 27/02/2025, às 17:23, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Bruna Mendes De Oliveira**, **Professor do Magistério Superior**, em 27/02/2025, às 17:24, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020 .

Documento assinado eletronicamente por **Rafael Tassinari Resende**, **Professor do Magistério Superior**, em 27/02/2025, às 17:27, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020 .

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5200538** e o código CRC **DBCB72F9**.

Referência: Processo nº 23070.011436/2025-62 SEI nº 5200538

AGRADECIMENTOS

A **Deus** por me conceder força, sabedoria e perseverança para concluir esta dissertação. Toda boa dádiva e todo dom vêm do alto.

Aos meus **pais** e **irmão**, pelo apoio incondicional.

A minha professora e orientadora, **Dra. Bruna Mendes de Oliveira**, não apenas pelas aulas, mas também, pelas lições sobre resiliência e inteligência emocional. Sua orientação se estendeu muito além da academia, moldando-me como pesquisadora e como pessoa. Agradeço também ao meu coorientador **Dr. Márcio Lisboa Guedes**, por toda ajuda e pelas incontáveis horas de aulas extras. Seu incentivo garantiu que eu permanecesse focada em minha jornada acadêmica.

Aos meus **amigos**, especialmente aqueles que conheci no Setor de Melhoramento de Plantas da Escola de Agronomia nesses últimos dois anos. Agradeço profundamente a vocês pelo apoio não apenas em questões acadêmicas, mas também na vida além da universidade.

A **sociedade brasileira**, cujo investimento em ciência e educação possibilita o funcionamento de instituições como a CAPES e CNPq, tornando viável minha trajetória acadêmica.

Ao **Programa de Pós-Graduação em Genética e Melhoramento de Plantas** da **Universidade Federal de Goiás** (PPGGMP-UFG) e a todos os seus docentes, que ministram disciplinas com tanta dedicação e profissionalismo. Em especial, às docentes mulheres, que são para mim grande fonte de inspiração.

Ao **Programa de Melhoramento Genético da Cana-de-Açúcar**, da Universidade Federal de Goiás da **Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético** (PMGCA/UFG/Ridesa) por possibilitar o desenvolvimento desse estudo e por me acolher. Em especial, ao melhorista Dr**. Renato de Carvalho Menezes**, por todo auxilio na execução desse trabalho, por me ensinar sobre cana-de-açúcar, estatística, programação e pelas vicências em campo.

Sou grata a todos que, de alguma forma, contribuíram para a realização deste trabalho.

SUMÁRIO

RESUMO 6
ABSTRACT 7
1 INTRODUÇÃO 8
2 REVISÃO BIBLIOGRÁFICA 10
2.1 CULTURA DA CANA-DE-AÇÚCAR 10
2.1.1 Relevância da cultura 10
2.1.2 Aspectos fenológicos 11
2.1.3 Melhoramento genético de cana-de-açúcar 13
2.2 AMBIENTE EM GENÉTICA E MELHORAMENTO DE PLANTAS 17
2.2.1 Interação genótipos com ambientes 17
2.2.2 Uso de covariáveis ambientais no melhoramento de plantas 19
2.2.3 Ambientômica 23
3 MATERIAL E MÉTODOS 26
3.1 ESTRUTURA EXPERIMENTAL 26
3.2 ANÁLISE CONJUNTA E CORREÇÃO DOS DADOS 27
3.3 AMBIENTIPAGEM 28
3.4 MODELOS PARA PREDIÇÃO 29
3.4.1 Modelo 1: G + E 30
3.4.2 Modelo 2: G + G × E 31
3.4.3 Modelo 3: G + E + G × E 31
3.4.4 Modelo 4: G + V 32
3.4.5 Modelo 5: G + G × V 33
3.4.6 Modelo 6: G + V + G × V 33
3.5 VALIDAÇÃO CRUZADA E HABILIDADES PREDITIVAS 34
4 RESULTADOS 36
4.1 ANÁLISES FENOTÍPICAS 36
4.2 HABILIDADES PREDITIVAS VALIDAÇÃO 5-*FOLD* 37
4.3 HABILIDADES PREDITIVAS VALIDAÇÃO *LEAVE-ONE-OUT* 40
5 DISCUSSÃO 42
6 CONCLUSÕES 46
7 REFERÊNCIAS 47
APÊNDICES 58

RESUMO

REIS, L. B. S. **Uso de covariáveis ambientais para predição da interação de clones de cana-de-açúcar com ambientes**. 2025. 51 f. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Escola de Agronomia, Universidade Federal de Goiás, Goiânia, 2025.[1]

A avaliação de genótipos em ensaios multiambientais que representem a diversidade das condições de cultivo é um dos principais desafios dos programas de melhoramento genético. Em espécies perenes ou semiperenes, como a cana-de-açúcar, esse desafio é ampliado pela experimentação prolongada ao longo de várias safras. O uso de covariáveis ambientais tem se mostrado uma abordagem promissora para aumentar a eficiência da seleção de genótipos. Este trabalho teve como objetivo avaliar diferentes modelos para a predição do comportamento de clones de cana-de-açúcar em distintos ambientes, utilizando covariáveis ambientais. Os dados fenotípicos de produtividade (TCH) foram obtidos de 53 experimentos conduzidos pelo Programa de Melhoramento Genético da Cana-de-Açúcar (PMGCA/UFG/Ridesa) entre 2012 e 2022. Dados agroclimáticos de 32 covariáveis ambientais foram coletados da plataforma POWER, da NASA. Seis modelos de regressão de cumeeira bayesiana foram desenvolvidos, integrando diferentes combinações de efeitos genotípicos (G), de experimentos (E), covariáveis ambientais (V) e suas interações. Para avaliar os modelos, dois esquemas de validação cruzada foram utilizados, 5-*fold* e *leave-one-out*, representando, respectivamente, os seguintes cenários: (1) capacidade dos modelos em predizer o comportamento de genótipos em experimentos nos quais não foram plantados; (2) capacidade dos modelos com covariáveis em predizer o comportamento de genótipos em ambientes não experimentados. A habilidade preditiva foi avaliada globalmente e por experimento. No cenário 1, as habilidades preditivas globais variaram de 0,93 a 0,13, enquanto, por experimento, variaram de 0,50 a 0,16. No cenário 2, as habilidades preditivas globais variaram de 0,47 a 0,01 e, por experimento, de 0,58 a 0,21. Os modelos sem covariáveis ambientais, G + E e G + E + G × E, mostraram-se adequados para preencher lacunas de informação em programas de melhoramento. O modelo G + V, apesar de não considerar a interação genótipos × covariáveis ambientais, destacou-se como a opção mais promissora para recomendações de cultivares em locais sem dados prévios. Os resultados evidenciam diferenças no desempenho dos modelos testados e a importância de considerar a variabilidade ambiental na interpretação das predições. Identificar experimentos com baixa capacidade preditiva é essencial para orientar estratégias de coleta de dados e aprimorar modelos capazes de lidar com a complexidade ambiental de maneira eficaz.

Palavras-chave: melhoramento genético, modelagem preditiva, ambiente.

[1] Orientadora: Prof[a]. Dr.[a] Bruna Mendes de Oliveira. EA – UFG.
Coorientador: Prof. Dr. Márcio Lisboa Guedes. UVA

ABSTRACT

REIS, L. B. S. **Environmental covariables to predict the interaction of sugarcane clones with environments**. 2025. 51 f. Dissertation (Master's in Genetics and Plant Breeding) – Escola de Agronomia, Universidade Federal de Goiás, Goiânia, 2025.[1]

The evaluation of genotypes in multi-environment trials that represent the diversity of cultivation conditions is one of the main challenges of plant breeding programs. In perennial or semi-perennial species, such as sugarcane, this challenge is amplified by prolonged experimentation over several harvest seasons. The use of environmental covariables has proven to be a promising approach to increase the efficiency of genotype selection. This study aimed to evaluate different models for predicting the performance of sugarcane clones in distinct environments using environmental covariables. Phenotypic data for yield (tons of stalks per hectare) were obtained from 53 trials conducted by the Sugarcane Breeding Program (PMGCA/UFG/Ridesa) between 2012 and 2022. Envirotyping data for 32 environmental covariates were collected from NASA's POWER platform. Six Bayesian ridge regression models were developed, integrating different combinations of genotypic effects (G), trial effects (E), environmental covariates (V), and their interactions. To evaluate the models, two cross-validation schemes were used, 5-fold and leave-one-out, representing the following scenarios, respectively: (1) the ability of the models to predict genotype performance in trials where they were not evaluated; (2) the ability of models with covariates to predict genotype performance in untested environments. Predictive ability was assessed globally and by individual trial. In scenario 1, global predictive abilities ranged from 0.93 to 0.13, while predictive ability by trial ranged from 0.50 to 0.16. In scenario 2, global predictive abilities ranged from 0.47 to 0.01, and by trial from 0.58 to 0.21. Models without environmental covariates, G + E and G + E + G × E, proved suitable for filling information gaps in breeding programs. The G + V model, although it did not consider genotype × environmental covariables interactions, stood out as the most promising option for cultivar recommendations in locations without prior data. The results highlight differences in the performance of the tested models and the importance of considering environmental variability in the interpretation of predictions. Identifying trials with low predictive ability is essential to guide data collection strategies and improve models capable of effectively handling environmental complexity.

Keywords: plant breeding, predictive modeling, environment, covariables.

[1] Advisor: Prof[a]. Dr.[a] Bruna Mendes de Oliveira. EA – UFG.
Co-Advisor: Prof. Dr. Márcio Lisboa Guedes. UVA

1 INTRODUÇÃO

A avaliação de genótipos em ensaios multiambientais que representem diversas condições de cultivo, é um dos principais desafios enfrentados pelos programas de melhoramento genético de plantas. Para recomendações de cultivares nos ambientes, deve-se captar o máximo da variabilidade ambiental possível. Contudo, ensaios multiambientais, de modo geral, são a fase mais cara e trabalhosa de um programa de melhoramento (Costa-Neto et al., 2020a). No caso do melhoramento de espécies perenes, ou semiperenes, como a cana-de-açúcar, esse desafio se intensifica. A experimentação prolongada, ao longo de várias safras, torna os ensaios ainda mais onerosos e logisticamente complexos.

O desenvolvimento de uma nova variedade de cana-de-açúcar pode estender-se de 10 a 15 anos (Gazaffi et al., 2016; Cursi et al., 2022). Ao final do melhoramento, a recomendação adequada dos genótipos é um fator decisivo para a expansão da cultura em novas áreas e para o aumento da eficiência produtiva. A depender do manejo oferecido e dos ambientes, as cultivares apresentam desempenhos diferenciais, fenômeno biológico conhecido como interação de genótipos com ambientes (Chaves, 2001). Covariáveis ambientais como temperatura e condições referentes a estresses hídricos, são relacionadas às interações de genótipos com ambientes e aos aspectos produtivos de cana-de-açúcar (Ramburan et al., 2011; Montes et al., 2021).

Aumentos contínuos de produtividade, entre outros caracteres, demonstram o sucesso dos programas de melhoramento em várias culturas (Siddiq & Vemireddy, 2021; Vogel et al., 2021; Kumar et al., 2021; Grandis et al., 2024), persistindo, contudo, a necessidade de otimizar a seleção genotípica e a reunião de múltiplas características desejáveis em uma única cultivar (Shen et al., 2021; Resende & Brondani, 2023). A predição do fenótipo se destaca como uma abordagem promissora para acelerar avaliações e ganho genético de caracteres complexos (Fradgley et al., 2023). Esses avanços têm sido impulsionados por modelos preditivos que integram covariáveis ambientais, contribuindo significativamente para pesquisas em culturas como milho (Costa-Neto et al., 2020b, 2021; Westhues et al., 2021; Resende et al., 2024b), feijão-

comum (Heinemann et al., 2022), trigo (Costa-Neto et al., 2022) e sorgo (Winans et al., 2023).

O potencial do uso desses modelos estende-se à otimização e alocação de recursos. A predição do desempenho de genótipos em ambientes ainda não testados pode reduzir o número de experimentos necessários e, consequentemente, os custos envolvidos. A capacidade de predizer o comportamento das plantas em diferentes condições ambientais é ainda uma promessa substancial para enfrentar os desafios da segurança alimentar global, particularmente diante de mudanças climáticas (Khan et al., 2022). Neste contexto, o presente estudo tem como objetivo avaliar diferentes modelos para a predição do comportamento de clones de cana-de-açúcar em diferentes ambientes utilizando covariáveis ambientais.

2 REVISÃO BIBLIOGRÁFICA

2.1 CULTURA DA CANA-DE-AÇÚCAR

2.1.1 Relevância da cultura

A cana-de-açúcar pertence ao gênero *Saccharum* L. e à família *Poaceae* (Cronquist & Takhtadzhian, 1981). A taxonomia do gênero *Saccharum* e suas espécies é complexa, já foi revisada várias vezes devido à presença de espécies sintéticas que complicam a classificação taxonômica (Amalraj & Balasundaram, 2006; Soreng et al., 2017; Mirajkar et al., 2019). Atualmente, o banco de dados *GrassBase*, mantido pelo *Kew's Royal Botanic Gardens*, reconhece 36 espécies dentro do gênero *Saccharum* (Clayton et al., 2024). O uso do termo *Saccharum* spp. ou complexo *Saccharum* é amplamente adotado para se referir ao conjunto de espécies do gênero (Amalraj & Balasundaram, 2006; Cursi, et al., 2021; Cursi et al., 2022).

Essa gramínea é cultivada principalmente em regiões tropicais e subtropicais, destacando-se como uma importante cultura energética e principal fonte de sacarose do mundo (Mirajkar et al., 2019; Babu et al., 2022). Dentre as quatro culturas que representam metade da produção global (cana-de-açúcar, milho, trigo e arroz), a cana-de-açúcar é elencada como a maior contribuinte, sendo Brasil e Índia os maiores produtores. (FAO, 2021, 2022, 2023).

O Brasil é o maior produtor de açúcar do mundo (FAO, 2021, 2022, 2023). A área colhida para a safra brasileira de 2023/24 foi estimada em 8,3 milhões de hectares com produção de 713,2 milhões de toneladas, recorde da série histórica da Companhia Nacional de Abastecimento (Conab), com aumento de 16,8%, quando comparada a safra anterior (Conab, 2024). O censo varietal IAC da safra 2023/24, englobou 271 unidades produtoras que somaram 6,8 milhões de hectares recenseados, em que a variedade RB867515 ocupou a maior área cultivada (12,5%), seguida de CTC4 (12,3%), RB966928 (11,1%), RB92579 (6,9%) e CTC9001 (5,4%) (Braga Júnior et al., 2024).

Em termos econômicos, a cana-de-açúcar é uma importante cultura para o estado de Goiás, a geração de empregos, renda e receitas oriundos do setor sucroenergético resultam em impactos na economia local e regional (Cunha & Pasqualetto, 2022). Ao longo dos anos, com incremento em área plantada, produção e produtividade média, Goiás encontra-se entre os três maiores produtores nacionais da cultura, juntamente com São Paulo e Minas Gerais (Conab, 2015, 2018, 2022, 2024).

Segundo a série histórica do sistema *Cidades@* (IBGE, 2023), de modo geral, o aumento de produção de cana-de-açúcar em Goiás, ocorreu mais de forma extensiva, ou seja, pela expansão de novas áreas de cultivo, do que de forma intensiva, pelo aumento de produtividade (t/ha). Vale ressaltar que em algumas regiões de Goiás, a disponibilidade de novas áreas de cultivo está cada vez menor (Alves & Wander, 2012; Rossi & Fernandes, 2020). A estimativa de área colhida, na safra 2023/24, foi ligeiramente menor em comparação a safra anterior. Essa diminuição pode ser atribuída à competição com lavouras de grãos como soja e milho (Conab, 2024).

Para o setor sucroenergético goiano, faz-se necessário buscar prioritariamente ganhos em produtividade, para diminuir a pressão sobre o uso do solo e possibilitar maior lucro econômico de produção (Rossi & Fernandes, 2020; Cunha & Pasqualetto, 2022).

2.1.2 Aspectos fenológicos

A cana-de-açúcar é classificada como uma planta semiperene e sua fenologia pode ser dividida em cinco estágios principais: brotação, perfilhamento, crescimento, maturação e florescimento, sendo este último estágio indesejável agronomicamente. A duração de cada fase varia em cada país e região, conforme a cultivar, as condições culturais e os parâmetros geográficos (Segato et al., 2006; Som-Ard et al., 2021).

No centro-sul brasileiro, o plantio se dá em duas épocas distintas: (1) de outubro a novembro, em que nestas condições, normalmente o canavial apresenta ciclo de 14 a 21 meses, denominado como “cana de ano e meio”; (2) de fevereiro a início de abril, as vezes prolongando o plantio até maio, com ciclo médio de 12 meses, sendo o desenvolvimento limitado nos meses de abril a agosto (Segato et al., 2006). No estado de Goiás, predomina o plantio na época 2, em que o estabelecimento inicial da cultura ocorre majoritariamente em março e o ciclo da cultura é denominado como “cana de ano”.

O plantio pode ocorrer a partir de sementes ou por propagação vegetativa, contudo o plantio via semente é realizado apenas por programas de melhoramento

(Bakker, 1999; Daros et al., 2015). Em nível comercial, a cana-de-açúcar é propagada vegetativamente, principalmente utilizando o colmo cortado em pedaços com três a seis gemas, que são plantados nos sulcos. Uma outra forma de propagação é através de mudas pré-brotadas (MPB), que consistem em pequenas plantas obtidas a partir de gemas individuais cultivadas em condições controladas.

A brotação ocorre 20 a 30 dias após o plantio, em que os primórdios radiculares darão origem as raízes fasciculadas e cada gema desenvolve o perfilho primário (Segato et al., 2006). Na fase de brotação, plantas daninhas são um ponto crítico do manejo, pois competem por nutrientes com as novas raízes da cana-de-açúcar. Além disto, raízes e brotos primários são altamente vulneráveis a pragas e doenças (Zhao & Li, 2015; Usaborisut, 2018).

Após a brotação, ocorre a fase de perfilhamento, em que são emitidos perfilhos secundários, terciários e assim sucessivamente, formando touceiras e posteriormente inicia-se um período de grande crescimento desses novos colmos (Segato et al., 2006). O perfilhamento tem influência direta na produção da cana-de-açúcar, sendo o número de perfilhos por touceira um indicador importante da futura produtividade da cultura. Um maior número de perfilhos resulta em um maior número de colmos adultos a serem colhidos (Segato et al., 2006; Silva & Silva, 2012).

O potencial de perfilhamento varia conforme a cultivar (Coelho et al., 2019; Benett et al., 2011) e é influenciado por fatores ambientais, principalmente a luz. A limitação de luz nessa fase pode ocasionar à morte de brotos e colmos jovens (Segato et al., 2006). Outro fator limitante é a disponibilidade hídrica, Segundo Zhao & Li (2018), a baixa disponibilidade de água nos estágios iniciais e intermediários de crescimento reduz principalmente, o rendimento dos colmos, resultando, consequentemente, em um rendimento inferior de sacarose em termos quantitativos.

O desenvolvimento vegetativo dura de 8 a 10 meses, a depender da região e das estações chuvosas e quentes, sendo seguido pela maturação, que ocorre no inverno e dura de 4 a 5 meses (Segato et al., 2006). Segundo Bull (2000), o amadurecimento da cana-de-açúcar é um processo contínuo que se torna evidente quando as condições ambientais restringem o alongamento do caule, mas não restringem significativamente a fotossíntese, favorecendo a translocação de açúcares para os tecidos de armazenamento (parênquima).

A cana-de-açúcar é uma gramínea que armazena carboidratos nos colmos e rizomas. O colmo é composto por uma sucessão de entrenós, nos quais a sacarose se

acumula de maneira progressiva, do entrenó basal para o ápice da planta (Toppa et al., 2010; Rae et al., 2014). Cada cultivar de cana-de-açúcar possui características específicas de crescimento e tempo de maturação, podendo ser classificadas em precoces, médias e tardias. Essas variações influenciam diretamente no planejamento do plantio e colheita, proporcionando flexibilidade no manejo e ajudando na distribuição do trabalho e utilização da capacidade industrial das usinas ao longo do ano.

Nos canaviais, a inflorescência é indesejada, pois o fluxo de sacarose produzido é redirecionado para a parte reprodutiva da planta. Além disso, a inflorescência está associada a um fenômeno chamado isoporização, que se caracteriza pelo aspecto esponjoso e menos denso dos colmos devido à desidratação dos tecidos, resultando na redução da produtividade e da qualidade da matéria-prima (Tasso Júnior et al., 2009; Araldi et al., 2010; Oliveira & Braga, 2011). O florescimento, isto é, a emissão das panículas, é geralmente evitado com o plantio de variedades que não florescem ou que raramente o fazem sob condições normais de cultivo (Bakker, 1999; Segato et al., 2006).

A cana-de-açúcar é uma cultura semiperene, cuja produtividade se reduz ao longo dos cortes. No primeiro ano de cultivo, entre o plantio e o primeiro corte, as plantas são chamadas de cana-planta e o termo cana-soca é utilizado para se referir à cana que já passou por pelo menos um corte (Segato et al., 2006; Silva & Silva, 2012). Após o primeiro corte, o rebrote é cultivado e colhido por um período de quatro a seis anos, dependendo da produtividade da cana-soca. Após este período, ocorre a renovação dos canaviais (Barros & Milan, 2010; Som-Ard et al., 2021).

Devido ao longo ciclo da cultura e aos vários estágios fenológicos, o manejo adotado na lavoura e as condições ambientais podem fornecer condições favoráveis ou desfavoráveis ao potencial produtivo das cultivares. Considerando as diversas condições ambientais e de cultivo da cana-de-açúcar no Brasil, o melhoramento genético da cultura precisa atentar-se às covariáveis ambientais durante o desenvolvimento das cultivares, a fim de contribuir com a evolução do setor sucroenergético, fornecendo cultivares cada vez mais adaptadas (Morais et al., 2015; Cursi et al., 2022).

2.1.3 Melhoramento genético de cana-de-açúcar

Dentre as espécies de cana-de-açúcar, *Saccharum officinarum* e *Saccharum spontaneum* são as maiores contribuintes para o genoma dos clones cultivados. O alto teor de sacarose acumulado pelas cultivares modernas é uma característica derivada de *S.*

officinarum ou cana-nobre, enquanto características de resistência a doenças e tolerância a estresses abióticos provém de *S. spontaneum* (Amaral et al., 2015; Zhang et al., 2018). O termo "complexo de híbridos *Saccharum"* pode ser encontrado na literatura para descrever os genótipos agronomicamente importantes que são produtos de longos anos de melhoramento (Zhang et al., 2018; Sanghera & Jamwal, 2019; Xu et al., 2021; Vasquez et al., 2022).

Os primeiros híbridos da cultura foram introduzidos na Índia em 1919 (Viswanathan, 2021). No Brasil, o primeiro lançamento ocorreu na década de 1930 pelo Instituto Agronômico de Campinas (IAC). A partir dessa época, grandes investimentos foram feitos pelos setores público e privado, tanto na importação de germoplasma quanto na implantação de programas de melhoramento (Andrade, 1985; Cursi et al., 2022).

Nos programas de melhoramento genético da cana-de-açúcar, a liberação das cultivares visa três objetivos básicos: produtividade, qualidade e longevidade do canavial (Segato et al., 2006; Silva & Silva, 2012). Por se tratar de uma cultura com colheitas sucessivas ao longo de vários anos, na seleção de um clone promissor, espera-se que o seu elevado desempenho se repita nos cultivos de cana planta e cana soca. Um aspecto relevante neste sentido é a mecanização da lavoura, uma vez que a colheita e carregamento mecânico podem abalar ou até mesmo arrancar a soca. Considerando que cerca de 92% da colheita de cana-de-açúcar no Brasil é realizada mecanicamente (Conab, 2024), a resistência ao tráfego de maquinários é uma característica essencial avaliada nos programas de melhoramento.

A resistência a patógenos é um outro aspecto relevante avaliado no melhoramento de cana-de-açúcar. Doenças como ferrugem alaranjada (*Puccinia kuehnii*), ferrugem marrom (*Puccinia melanocephala*), carvão da cana-de-açúcar (*Ustilago scitaminea*), têm grande importância na agricultura canavieira brasileira por gerarem grandes perdas de produtividade (Morais et al., 2015). O controle mais recomendado para essas doenças é o plantio de variedades resistentes (Cursi et al., 2022).

De acordo com os objetivos específicos de cada programa, são aferidas características importantes na determinação de clones superiores. No desenvolvimento de cana-energia, por exemplo, visam-se as características como alto teor de fibra, baixo teor de sacarose e alta biomassa, enquanto na produção de açúcar e etanol, almejam-se variedades com alta produtividade de colmos e alto teor de sacarose (Gazaffi et al., 2016; Mirajkar et al., 2019).

Nos últimos anos, com a intensa expansão da cultura em solos de baixa fertilidade e a adoção da mecanização, novos desafios têm sido impostos aos programas de melhoramento (Cursi et al., 2022). Em países com grande extensão territorial, um dos maiores desafios enfrentados pelos melhoristas é a resposta diferencial dos genótipos em diferentes ambientes (Elias et al., 2016), fenômeno denominado interação de genótipos com ambientes ou simplesmente interação G×A (Chaves, 2001). Assim, além de bom desempenho e resistência, as variedades precisam apresentar adaptação a uma ampla gama de ambientes (Mirajkar et al., 2019).

O desenvolvimento das cultivares de cana-de-açúcar plantadas no Brasil é, principalmente, de domínio nacional (Medina, 2024). Em contraste, em outras culturas relevantes no país, como a soja, há uma forte participação de empresas estrangeiras no registro e proteção de cultivares (Degani et al., 2021). Embora o melhoramento da cana-de-açúcar seja predominantemente realizado por métodos clássicos, as abordagens biotecnológicas que utilizam os recentes avanços em recursos genômicos, já são realidade na busca pelo progresso genético em programas de melhoramento tradicionais (Cursi et al., 2022).

Os principais programas de melhoramento genético da cana-de-açúcar em atividade no Brasil são: a Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético (Ridesa), o Centro de Tecnologia Canavieira (CTC) e o Instituto Agronômico de Campinas (IAC) (Gazaffi et al., 2016). Das variedades de cana-de-açúcar cultivadas no Brasil, dois terços provêm da Ridesa (Medina, 2024).

O programa de melhoramento da Ridesa, dispõe de 101 bases de pesquisa, englobando laboratórios, estações de cruzamento, estações experimentais e bases de seleção (Oliveira et al., 2021). Dez universidades públicas fazem parte da rede, que operam com gestões independentes, mas integradas nacionalmente, a saber: Universidade Federal de Alagoas, Universidade Federal de Goiás, Universidade Federal de Mato Grosso, Universidade Federal de São Carlos, Universidade Federal de Sergipe, Universidade Federal de Viçosa, Universidade Federal do Paraná, Universidade Federal do Piauí, Universidade Federal Rural de Pernambuco e Universidade Federal Rural do Rio de Janeiro. O fluxo e logística dentro de cada etapa do programa de melhoramento, da escolha de genitores, até a obtenção de uma variedade comercial variam em função das condições e preferências de cada universidade.

O desenvolvimento de uma nova cultivar de cana-de açúcar demanda de 10 a 15 anos de trabalho (Gazaffi et al., 2016; Cursi et al., 2022). O processo de melhoramento

começa com a escolha de genitores com potencial de complementaridade. Os cruzamentos são programados e ocorrem nas estações de cruzamentos da Ridesa: Estação de Floração e Cruzamento Serra do Ouro, em Murici - AL e Estação de Floração e Cruzamento de Devaneio, em Amaraji - PE. Uma vez realizados os cruzamentos, cada estação experimental vinculada a uma universidade recebe as sementes.

No âmbito do Programa de Melhoramento Genético da Cana-de-Açúcar, da Universidade Federal de Goiás, da Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético (PMGCA/UFG/Ridesa), as sementes são semeadas por famílias, em bandejas preenchidas com substrato (Ridesa, 2024). Após a germinação das sementes, as bandejas com as plântulas são transplantadas para o campo. Esse sistema é denominado "tapetinho", devido ao aspecto visual das bandejas transplantadas nos sulcos de plantio (Figura 1).

Figura 1. Plantio da fase de "tapetinhos" no campo experimental do PMGCA/UFG/Ridesa na Escola de Agronomia da Universidade Federal de Goiás em dezembro de 2024.

Após a fase de tapetinhos, é iniciado o processo de seleção, passando pelas fases de testes denominadas como T1, T2 e T3. Ao fim da fase T3, os clones pré-selecionados são submetidos a um rigoroso ensaio final, conhecido como fase experimental (FE). Nessa etapa, o desempenho dos clones é avaliado durante quatro cortes, em diferentes épocas e em vários ambientes. Após essa avaliação os clones são

validados e multiplicados em áreas pré-comerciais, podendo ocorrer, ou não, a liberação de uma nova cultivar (Daros et al., 2015). Nas etapas de pós-melhoramento, são realizados ensaios de validação para demonstrar o desempenho produtivo e as características morfoagronômicas dos clones superiores, além das atividades relacionadas à divulgação, registro e lançamento.

2.2 AMBIENTE EM GENÉTICA E MELHORAMENTO DE PLANTAS

2.2.1 Interação genótipos com ambientes

A interação genótipos com ambientes (G × A) é definida por Chaves (2001) como o efeito diferencial dos ambientes sobre os genótipos o que resulta na resposta diferencial dos genótipos à variação ambiental. Nos ensaios finais dos programas de melhoramento, é comum ocorrer inconsistência da performance produtiva dos genótipos ao longo dos ambientes. Sendo assim, um genótipo ranqueado como o melhor em um dado ambiente, pode não performar tão bem em todos os ambientes.

A interação G × A pode ser nula, simples ou complexa. Na interação simples ocorre a mudança na magnitude de performance dos genótipos, mas a classificação destes permanece inalterada, ou seja, um genótipo superior no ambiente A permanece superior no ambiente B. Por outro lado, a interação complexa reflete o ordenamento distinto dos genótipos nos ambientes, não havendo um genótipo específico que seja consistentemente superior em todos os locais (Chaves, 2001; Squilassi, 2003). Antagonicamente, caso a interação G × A fosse nula, uma variedade se ajustaria a praticamente todos os ambientes de cultivo. Assim, um único ensaio, realizado em um só local seria o bastante para obter resultados amplos (Duarte & Vencovsky, 1999).

Segundo o conceito proposto por Mariotti (1976), a adaptabilidade refere-se à capacidade do genótipo em assimilar de forma vantajosa os estímulos provenientes do ambiente. Já o termo estabilidade diz respeito à capacidade de os genótipos exibirem resultados constantes em diferentes ambientes, apresentando menor variação diante das mudanças ambientais (Chaves, 2001; Ramalho et al., 2012). Em ensaios multiambientais, espera-se que os diferentes ambientes representem a faixa de condições agrícolas nas quais os genótipos podem apresentar adaptação específica ou ampla adaptação (Van Eeuwijk et al., 2005).

O desempenho dos genótipos em ensaios multiambientais é analisado por modelos estatísticos desenvolvidos para interpretar e descrever as interações G × A (Tabela 1). A escolha de qual método utilizar depende de alguns fatores, tais como a quantidade de ambientes e genótipos, precisão e tipo de dados e objetivos específicos do estudo (Yan et al., 2007; Elias et al., 2016; Costa-Neto et al., 2020a). Para recomendações mais precisas quanto à seleção de cultivares nos ambientes deve-se captar o máximo da variabilidade ambiental possível, contudo os ensaios multiambientais, de modo geral, são a fase mais cara e trabalhosa de um programa de melhoramento (Squilassi, 2003; Costa-Neto et al., 2020a).

Tabela 1. Exemplo de modelos para analisar estabilidade e/ou adaptabilidade de genótipos.

Método	**Descrição geral**
Wricke ou método da ecovalência[1]	As estimativas de W_i^2 partem da decomposição da soma de quadrados da interação. Indica genótipos estáveis, independente do rendimento médio. Quanto menor o valor de W_i^2 maior a estabilidade.
Eberhart e Russel[2]	Baseado em regressão linear simples. Um genótipo desejável é aquele com média alta (intercepto = β_{0i}), coeficiente de regressão maior ou pelo menos igual a 1,0 (coeficiente angular = β_{1i}) e desvio da regressão próximo ou igual 0 (δ_{ij}).
Lin e Binns[3]	Método não paramétrico e determina a estabilidade pelo quadrado médio da distância entre a média do genótipo e a resposta média máxima obtida no ambiente. Estima o índice de estabilidade P_i, que é o desvio do genótipo i em relação ao genótipo de desempenho máximo em cada ambiente, logo, quanto menor o valor de P_i, mais previsível e adaptado é o material.
Cruz, Torres e Vencovsky[4]	Baseia-se na análise de regressão bi-segmentada, considerando como parâmetros de adaptabilidade a média e a resposta dos genótipos aos ambientes favoráveis e desfavoráveis. Um genótipo é estável se o quadrado médio dos desvios de regressão que avalia a estabilidade é pequeno.
Toler[5]	Baseada em regressão não linear. Sugere dois padrões de resposta em função do parâmetro que mede o efeito ambiental: padrão convexo (duplamente desejável) e côncavo (duplamente indesejável).

Continua.

Método	Descrição geral
Annicchiarico[6]	Engloba em um único parâmetro os conceitos de adaptação, adaptabilidade e estabilidade. Para isso, o método proporciona uma medida de estabilidade denominada de índice de confiança genotípico ou *index reliability* (I_i). Quanto maior for este índice, maior será a confiança no posicionamento e recomendação da cultivar.
AMMI (*Additive Main Effects and Multiplicative Interactions*)[7]	Combina a análise de variância para estimar os efeitos principais dos genótipos e ambientes e a decomposição por valores singulares para repartir G×A em vetores ortogonais. A representação visual dos efeitos da interação é feita por meio de um gráfico biplot, que fornece informações sobre a estabilidade e adaptabilidade dos genótipos.
GGE *biplot* (*Genotype main effect plus Genotype-by-Environment interaction*)[8]	Avalia de forma multiplicativa o efeito de G + G × A. A análise gráfica do GGE é feita em um gráfico *biplot*, gerando informações sobre genótipos com ampla adaptabilidade.
Método gráfico de Nunes[9]	Método baseado em gráficos tipo radar. O formato "bola cheia" descreve um genótipo que tem comportamento acima da média ou desempenho consistente. Já, o formato "bola murcha" remete à um genótipo com desempenho abaixo da média, ou com comportamento instável.
WAASB (*Weighted Average of Absolute Scores from the singular value decomposition of the matrix of Best linear unbiased predictions for the GEI effects generated by an linear mixed-effect*)[10]	O índice WAASB é uma forma de modelo misto de AMMI, considerando genótipos como covariáveis aleatórias e levando em consideração todos os eixos de componentes principais de interação para análise de estabilidade. O genótipo com o menor valor de WAASB pode ser considerado como o mais estável. O gráfico biplot usa todos os eixos estimados para identificar a estabilidade.

[1]Wricke (1965); [2]Eberhart & Russel (1966); [3]Lin & Binns (1988); [4]Cruz et al. (1989); [5]Toler (1990); [6]Annicchiarico (1992); [7]Gauch (1992), [8]Yan (2001), [9]Nunes et al. (2005) [10]Olivoto et al. (2019).

2.2.2 Uso de covariáveis ambientais no melhoramento de plantas

Os caracteres agronômicos de importância no melhoramento de plantas são, em sua maioria, quantitativos e sofrem grande influência do ambiente (Chaves, 2001). Na agricultura de precisão, diversos componentes ambientais, como por exemplo, as condições de solo, são estudados para compreender a variabilidade espacial dentro de

uma região ou lavoura, possibilitando a aplicação de manejos específicos (Souza et al., 2022; Mazzi et al., 2023). Compreender como os fatores ambientais influenciam esses caracteres permite que práticas mais precisas e ajustadas sejam implementadas na agricultura.

Além da influência da qualidade ambiental nas características de produção, ainda existem os efeitos da interação G × A. Embora a interação possa ser considerada por melhoristas como um problema a ser resolvido, é possível tirar proveito deste fator ao praticar a seleção, visto a ocorrência de interações positivas devido a fatores edafoclimáticos previsíveis que podem oferecer ganhos na eficiência produtiva (Duarte & Vencovsky, 1999; Chaves, 2001). Por isso, o conhecimento da resposta ecofisiológica de cada genótipo em cada ambiente de produção é fundamental.

Integrar covariáveis ambientais na interpretação dos fatores causais da interação G × A e das respostas dos genótipos, pode ser um método que otimize a avaliação genotípica e facilite a recomendação de cultivares (Martins, 2004; Costa-Neto et al., 2020a; Montes et al., 2021). Por muitos anos, abordagens empíricas a partir do valor fenotípico observado *in situ* foram bastante difundidas e utilizadas (Eberhart & Russel 1966; Toler 1990; Annicchiarico, 1992; Gauch, 1992). Com a facilidade na coleta de informações ambientais, tais como covariáveis ambientais disponíveis em banco de dados globais ou coletadas em campo, essas informações adicionais estão sendo utilizadas na tomada de decisão dentro dos programas de melhoramento (Montes et al., 2021; Albuquerque et al., 2023).

Resende & Brondani (2023) dividem as tecnologias para coleta de dados ambientais em quatro categorias: (1) Sensores instalados em campo, como estações meteorológicas e amostragens de campo; (2) Tecnologias de sensoriamento remoto, que utilizam satélites, drones e aviões para capturar informações; (3) Bancos de dados globais, que reúnem e interpolam informações de fontes locais e remotas, armazenando-as digitalmente, por exemplo, NASA POWER (https://power.larc.nasa.gov/); (4) Registros manuais no campo, como aqueles relacionados às práticas de manejo.

De posse dos dados, diferentes análises quantitativas podem ser desenvolvidas em ensaios multiambientais integrando covariáveis de informação ambiental. Costa-Neto et al. (2020a) dividem essas análises em duas classes de modelos analíticos: (1) métodos de ambientipagem, com o objetivo de caracterização dos fatores ambientais chave dentro da população de ambientes-alvos definida pelos programas de melhoramento e (2) métodos exploratórios preditivos, baseados em regressões fatorial e

quadrados mínimos parciais, como o foco na interpretação ecofisiológica da adaptação fenotípica.

O avanço dessas abordagens no melhoramento de plantas ocorreu, em grande parte, devido ao avanço das tecnologias de coleta, armazenamento e processamento de dados (Costa-Neto et al., 2020b, 2021, 2022; Westhues et al., 2021; Heinemann et al., 2022; Winans et al., 2023, Resende et al., 2024b, Cruz et al., 2025). Entretanto, existem trabalhos pioneiros, como Annicchiarico (2002), que apresenta um estudo de caso com 24 cultivares de trigo cultivadas em 18 localidades da Argélia ao longo de dois anos. O objetivo principal foi otimizar as recomendações de cultivares e apoiar decisões nos programas de melhoramento. Seis covariáveis ambientais foram coletadas e utilizadas em análises de regressão fatorial e agrupamento (cluster), o que contribuiu na recomendação dos materiais e classificação dos ambientes em sub-regiões.

Para identificar os principais fatores ambientais que influenciam a interação entre genótipos e ambientes, alguns estudos aplicam análises de regressão linear via procedimento *stepwise* em culturas como soja (Martins, 2004), melão (Nunes et al., 2011) e cana-de-açúcar (Montes et al., 2021). Em síntese, o procedimento *stepwise* é utilizado para identificar um subconjunto de covariáveis explicativas que forneçam o melhor ajuste a um modelo estatístico. Através desta abordagem, pode-se estimar se os efeitos da interação são relacionados a fatores previsíveis como latitude, longitude, altitude, ou a fatores imprevisíveis como precipitação e temperatura. Esses resultados podem contribuir para a definição da rede de avaliação e direcionar esforços localizados.

No estudo de Martins (2004) e Montes et al. (2021), além da identificação de covariáveis ambientais relevantes, os autores integram a avaliação genotípica a sistemas de informação geográfica (SIG), permitindo a recomendação de cultivares de forma mais dinâmica e acessível, baseada em mapas temáticos. Montes et al. (2021) utilizaram 14 fatores ambientais que estabeleceram 113 covariáveis ambientais para caracterizar nove estações experimentais em usinas de cana-de-açúcar no estado de Goiás. No estudo, estimou-se que cerca 57% dos efeitos da interação G × A no estudo são relacionados às covariáveis longitude, temperatura média na fase de brotação da cultura e temperatura máxima no início da fase de maior crescimento.

Ramburan et al. (2011) usaram 14 covariáveis ambientais na análise de interações G × A em variedades de cana-de-açúcar, na África do Sul. Os autores utilizaram a análise de componentes principais (PCA) para determinar os padrões de separação regional e para caracterizar a influência relativa de cada covariável nos

ambientes. A interação G × A foi então modelada pelo modelo AMMI e através da análise de correlação, determinou-se a relação das covariáveis com escores de ambiente AMMI. A abordagem utilizada neste estudo identificou fatores ambientais chaves que separavam os ambientes do ensaio, embora as correlações entre as covariáveis e os escores dos eixos da análise de componentes principais da interação (IPCA) do AMMI foram geralmente fracas. Segundo as análises, a idade de colheita, a temperatura e o estresse hídrico foram os principais responsáveis pela separação dos ambientes e se relacionaram com os *biplots* de rendimento de cana e açúcar total recuperável.

A construção de índices ambientais é uma alternativa para sintetizar múltiplos fatores em um único valor contínuo. O primeiro passo dessa abordagem é obter covariáveis ambientais que representem o gradiente ambiental da área de estudo (Marcatti et al., 2017). No estudo de Marcatti et al. (2017), um índice ambiental foi desenvolvido por meio da análise de componentes principais (PCA). Os autores utilizaram esse índice para definir estratos quantitativos em uma área de produção de eucalipto, permitindo otimizar a recomendação de genótipos com base nas variações ambientais.

Um outro campo de estudo dentro do contexto do uso de covariáveis ambientais, é a predição de caracteres utilizando essas informações. Embora algumas ferramentas baseadas em predição com informações genômicas são consolidadas na literatura, como a seleção genômica ampla (*Genome Wide Selection* – GWS) (Meuwissen et al., 2001), a predição com covariáveis ambientais tem alavancado a pesquisa no melhoramento de plantas nos últimos cinco anos (Costa-Neto et al., 2020b; Costa-Neto et al., 2021; Westhues et al., 2021; Heinemann et al., 2022; Winans et al., 2024; Resende et al., 2024b).

Para predição, esses estudos incluem GBLUPs baseados em diferentes métodos de predição (Costa-Neto et al., 2021), GBLUP kernels (Winans et al., 2024; Costa-Neto et al., 2020a), regressão de cumeeira bayesiana (Winans et al., 2024), modelos aditivos generalizados (Heinemann et al., 2022), métodos baseados em aprendizado de máquina (Westhues et al., 2021; Resende et al., 2024b) e modelos com marcadores ambientômicos (Resende et al., 2024b), podendo haver a integração ou não de dados ambientais e dados genômicos.

No estudo realizado por Jarquín et al. (2014), 68 covariáveis ambientais foram usadas para avaliar o rendimento de grãos de linhagens comerciais de trigo, permitindo a predição de linhagens não observadas, em ambientes não testados, com base em dados ambientais e genômicos. A inclusão de informações ambientais nos modelos

genéticos resultou em ganhos significativos na predição, de 2 – 21% de ganho, a depender do modelo. A maior correlação preditiva foi de 0,51, no modelo mais abrangente, que continha efeitos dos ambientes, informações genômicas, covariáveis ambientais, interação genótipos × ambientes e covariáveis × genótipos.

Winans et al. 2024 também combinaram informações genômicas e ambientais para predizer caracteres de híbridos de sorgo granífero, testando sete modelos diferentes. Foram coletadas 99 covariáveis ambientais, em dez ambientes. Os autores utilizaram três esquemas de validação cruzada, que simulam cenários de predição encontrados por programas de melhoramento: (1) Predição de híbridos testados em ambientes não avaliados; (2) Predição de híbridos não testados em ambientes já avaliados; (3) Predição de híbridos testados em pelo menos um ambiente, mas não em todos. Em relação a predição da produtividade, as maiores habilidades preditivas foram obtidas nos cenários 2 e 3, com os coeficientes de correlação de Pearson entre 0,46 e 0,88, a depender do modelo.

Os resultados obtidos por Winans et al. 2024, apoiam o uso de ensaios desbalanceados quando os programas de melhoramento aplicam modelos de predição. Isso permite reduzir o número de híbridos avaliados no campo, aplicando os modelos para predizer o desempenho dos híbridos que não foram plantados. Além disto, os autores destacam que os modelos com melhor desempenho nos ambientes temperados, não são os mesmos dos ambientes subtropicais. O que indica que cada região tem características específicas que exercem influência diferencial na cultura de sorgo. Logo, a modelagem preditiva deve levar esses fenômenos em consideração.

A capacidade de predizer com precisão o desempenho das culturas em ambientes não testados é uma promessa substancial para enfrentar os desafios da segurança alimentar global, particularmente em face das mudanças climáticas (Bowerman, 2024). O uso de covariáveis ambientais no melhoramento de plantas tem mostrado grande potencial para aumentar a eficiência na seleção de genótipos e otimizar a adaptação das culturas aos ambientes de cultivo (Jarquín et al., 2014; Costa-Neto et al., 2020a, Resende et al., 2024a, Winans et al., 2024).

2.2.3 Ambientômica

A contribuição do genótipo (G), ambiente (A) e interação (G × A) na variação fenotípica (F) é descrito na genética quantitativa pela equação básica: F = G + A + G × A

(Ramalho et al., 2012). Entre esses componentes de variação, o genético foi o primeiro para o qual medições de alto rendimento se tornaram disponíveis, especialmente após a conclusão da primeira sequência do genoma vegetal, o da herbácea *Arabidopsis thaliana*, em 2000 (*The Arabidopsis Genome Initiative*, 2000). Desde então, diversos genomas de plantas foram sequenciados. De maneira análoga, com o avanço de tecnologias de caracterização, o componente ambiental também tem sido um campo de estudo com grandes avanços no melhoramento de plantas, dentre esses, estão os estudos ambientômicos (Costa-Neto et al., 2021, 2022; Resende et al., 2024a, 2024b).

O termo "ômica" refere-se ao estudo dos processos biológicos como sistemas, o que envolve a medição de um grande número de parâmetros (Maffia et al., 2017; Resende & Brondani, 2023). Dados ambientais detalhados permitem tomadas de decisões com mais embasamento, levando, em última análise, a culturas de melhor desempenho adaptadas a condições ambientais específicas (Bowerman, 2024). Nesse contexto, os conceitos relacionados a ambientômica, juntamente com fenômica e genômica, podem ser incorporados para melhor entender as contribuições dos componentes genéticos, ambientais e interação G × A no rendimento ou desempenho de um genótipo e auxiliar na tomada de decisões voltadas ao futuro.

A integração do melhoramento com grandes volumes de dados ambientais só é possível devido a tecnologias de geoprocessamento, como o SIG (Sistema de Informação Geográfica). Segundo Resende et al. (2024a), a implementação dessa tecnologia na ambientômica envolve a identificação da população de ambientes alvo (*Target Population Environment* - TPE). Os ambientes do TPE representam as condições ambientais esperadas dentro uma região produtora de determinada cultura, nos quais os genótipos de um programa de melhoramento serão cultivados (Chenu, 2015; Cooper et al., 2023; Vitale et al., 2024).

A caracterização de um TPE é normalmente conduzida para identificar classes ambientais chave que afetam o desempenho do genótipo (Chenu, 2015). Cruz et al. (2025) propõem o uso de ferramentas de SIG para definir a população de ambientes alvo, possibilitando a delimitação espacial das regiões de interesse nos estudos ambientômicos. Essa caracterização mais precisa dos ambientes pode auxiliar no delineamento de ensaios mais informativos e eficientes.

Quando a interação G × A é significativamente associada a grupos de ambientes dentro do TPE, maiores ganhos genéticos podem ser obtidos reorganizando experimentos (Callister et al., 2024; Resende et al., 2024b; Krause, et al., 2025). Callister

et al. (2024) abordam a interação G × A no melhoramento florestal de eucalipto na Austrália, enfatizando a importância de delinear zonas de melhoramento (*breeding zones*). Essas zonas, são polígonos delimitados geograficamente visando minimizar G × A (Callister et al., 2024; Resende et al., 2024a; Resende et al., 2024b).

No estudo de Resende et al. (2024b), a produtividade de grãos foi avaliada em 183 experimentos distribuídos em quatro estados brasileiros. Foram definidas quatro zonas de melhoramento, segundo suas similaridades ambientais e análise G × A. Um total de 1324 covariáveis ambientais foram utilizadas na predição de produtividade de milho em regiões experimentadas e não experimentadas dentro dos limites do TPE. Uma contribuição fundamental desta pesquisa é o desenvolvimento de Marcadores Ambientômicos Projetados (EEM – *Engineered Enviromic Markers*). Diferentes modelos foram elaborados para predizer a produtividade de grãos de milho. As habilidades preditivas calculadas através da correlação de Pearson, variaram de 0,33 a 0,58 a depender do modelo e esquema de validação cruzada.

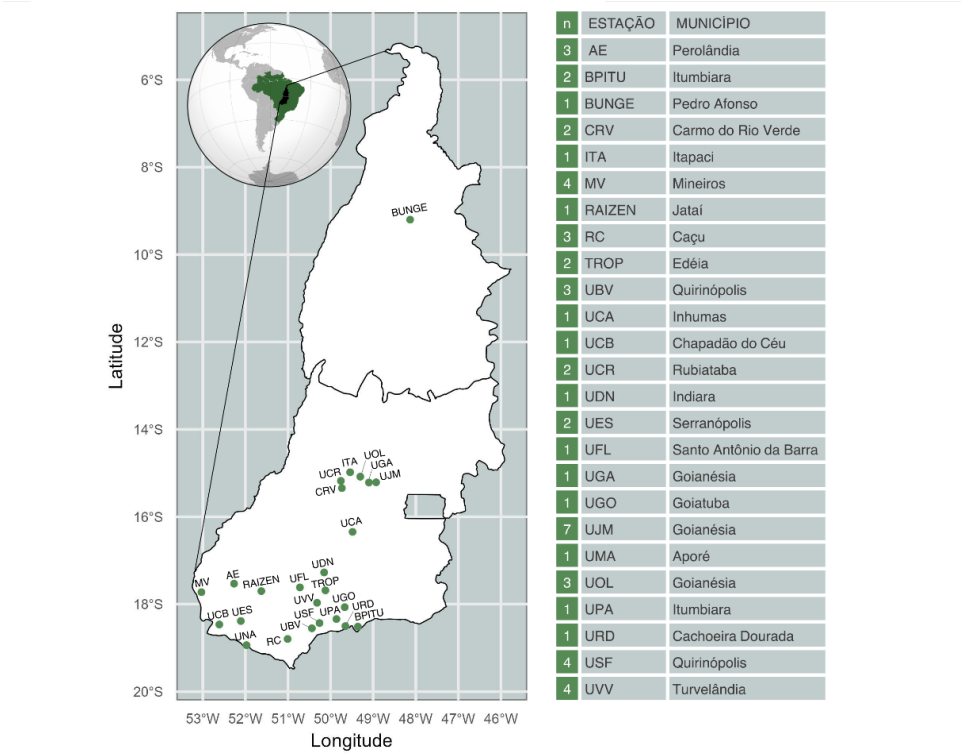
Por vezes, abordagens ambientômicas integram informações genômicas. O estudo de Costa-Neto et al. (2022) examina o impacto de dados ambientais na melhoria da predição genômica em programas de melhoramento de trigo. Um total de 108 covariáveis ambientais foram utilizadas. Os autores propuseram uma nova estrutura capaz de reciclar os dados ambientais históricos e usá-los para melhorar a precisão da predição dos métodos GBLUP convencionais.

Uma das possibilidades da ambientômica, utilizando informações multidimensionais de dados ambientais de alta precisão, é a recomendação genotípica com extrapolação preditiva para todos os pixels de uma área, possibilitando uma alocação mais eficiente dos recursos (Resende et al. 2024a). Apesar dessas possibilidades, ainda existem aspectos que precisam ser aprimorados. Entre eles, maior precisão dos dados climáticos, pixels menores para permitir análises cada vezes mais detalhadas e previsões desses dados para cenários futuros (Resende & Brondani, 2023).

3 MATERIAL E MÉTODOS

3.1 ESTRUTURA EXPERIMENTAL

Os dados utilizados são provenientes de ensaios de fases finais conduzidos no âmbito do Programa de Melhoramento Genético da Cana-de-Açúcar, da Universidade Federal de Goiás, da Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético (PMGCA/UFG/Ridesa). Ao todo, foram coletadas informações de 53 experimentos, distribuídos em 25 usinas conveniadas ao PMGCA/UFG/Ridesa, no período de 2012 a 2022 (Figura 2). Esses locais representam as áreas de produção comercial de cana-de-açúcar em Goiás, incluindo uma usina localizada no Tocantins, a única em operação no estado. Cada experimento foi gerenciado com práticas agronômicas representativas do cultivo comercial da cultura.



n	ESTAÇÃO	MUNICÍPIO
3	AE	Perolândia
2	BPITU	Itumbiara
1	BUNGE	Pedro Afonso
2	CRV	Carmo do Rio Verde
1	ITA	Itapaci
4	MV	Mineiros
1	RAIZEN	Jataí
3	RC	Caçu
2	TROP	Edéia
3	UBV	Quirinópolis
1	UCA	Inhumas
1	UCB	Chapadão do Céu
2	UCR	Rubiataba
1	UDN	Indiara
2	UES	Serranópolis
1	UFL	Santo Antônio da Barra
1	UGA	Goianésia
1	UGO	Goiatuba
7	UJM	Goianésia
1	UMA	Aporé
3	UOL	Goianésia
1	UPA	Itumbiara
1	URD	Cachoeira Dourada
4	USF	Quirinópolis
4	UVV	Turvelândia

Figura 2. Localização das estações experimentais PMGCA/UFG/Ridesa e municípios compreendidos no trabalho. No quadro, n se refere a quantidade de experimentos analisados em cada usina.

O experimento de menor tamanho avaliou três genótipos, enquanto o de maior tamanho incluiu 17 genótipos. Ao todo, foram avaliados 37 genótipos, totalizando 419 observações (Figura 3). Como covariável resposta, foi utilizado o caráter toneladas de colmos por hectare (TCH). Nos experimentos, o TCH foi obtido pela transformação do peso total das parcelas em toneladas por hectare. O delineamento experimental utilizado foi o de blocos casualizados, com quatro ou três repetições, a depender da usina. Ressalta-se que, neste estudo, os valores de TCH são referentes ao ciclo de cana planta (primeiro corte).

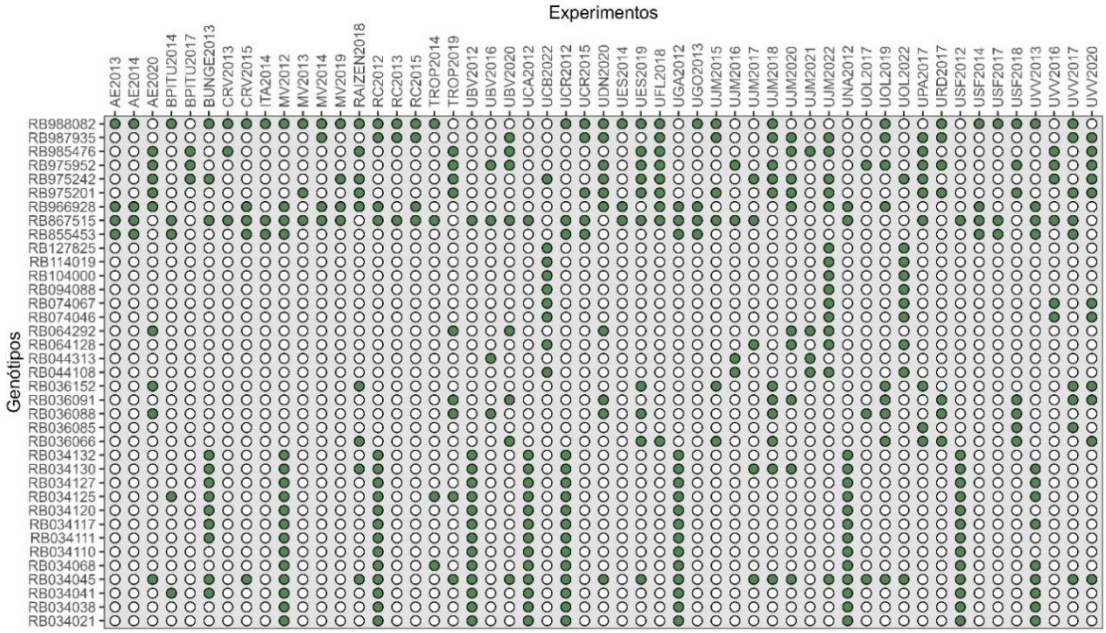


Figura 3. Estrutura experimental. Círculos em branco representam a ausência do genótipo no experimento e círculos coloridos verdes representam a presença do genótipo no experimento.

3.2 ANÁLISE CONJUNTA E CORREÇÃO DOS DADOS

Os componentes de variância associados aos efeitos do modelo de análise conjunta foram estimados utilizando o Método de Máxima Verossimilhança Residual (*Residual Maximum Likelihood* – REML), por meio o pacote lme4 do R (Bates, 2010), seguindo o modelo estatístico:

$$\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{Z_{01}r} + \boldsymbol{Z_{02}g} + \boldsymbol{Z_{03}e} + \boldsymbol{Z_{04}ge} + \boldsymbol{\varepsilon}$$

Em que: y é o vetor de observações da covariável resposta TCH; $\boldsymbol{\mu}$ é o vetor do efeito fixo de uma constante comum a todas observações; $\boldsymbol{r} \sim N(0, \mathbf{I}\sigma^2_r)$ é o vetor de efeitos aleatórios das repetições dentro de cada experimento; $\boldsymbol{g} \sim N(0, \mathbf{I}\sigma^2_g)$ é o vetor de

efeitos aleatórios dos genótipos; $\boldsymbol{e} \sim$ N (0, $\mathbf{I}\sigma^2_e$) é o vetor de efeitos aleatórios dos experimentos; $\boldsymbol{ge} \sim$ N (0, $\mathbf{I}\sigma^2_{ge}$) é o vetor de efeitos aleatórios da interação genótipos × experimentos; $\boldsymbol{\varepsilon} \sim$ N (0, $\mathbf{I}\sigma^2_{\varepsilon}$) é o vetor de resíduos. $\boldsymbol{Z_{01}}$, $\boldsymbol{Z_{02}}$, $\boldsymbol{Z_{03}}$ e $\boldsymbol{Z_{04}}$ são respectivamente as matrizes de incidências referentes a cada vetor de efeitos e **I** é matriz identidade. As significâncias dos efeitos aleatórios foram obtidas por meio do Teste de Razão de Verossimilhanças (*Likelihood Ratio Test* – LRT).

Os valores ajustados da covariável resposta TCH ($\hat{\boldsymbol{y}}$) foram obtidos por meio de BLUPs (*Best Linear Unbiased Prediction*) associados aos fatores do modelo acima descrito, conforme a equação:

$$\hat{\boldsymbol{y}} = \boldsymbol{\mu} + \hat{\boldsymbol{g}} + \hat{\boldsymbol{e}} + \widehat{\boldsymbol{ge}}$$

Em que: $\hat{\boldsymbol{y}}$ é o vetor dos valores ajustados da covariável resposta TCH; $\boldsymbol{\mu}$ é o vetor do efeito fixo de uma constante comum a todas observações; $\hat{\boldsymbol{g}}$ é o vetor de BLUPs dos genótipos; $\hat{\boldsymbol{e}}$ é o vetor de BLUPs de experimentos e $\widehat{\boldsymbol{ge}}$ é o vetor de BLUPs das interações genótipos × experimentos. Os valores ajustados ($\hat{\boldsymbol{y}}$) foram utilizados como covariável resposta para as análises subsequentes.

3.3 AMBIENTIPAGEM

As coordenadas geográficas de latitude, longitude e altitude das usinas conveniadas do PMGCA/UFG/Ridesa foram obtidas por GPS (*Global Positioning System*). Com base na localização de cada usina, coletou-se covariáveis ambientais na plataforma POWER (*Prediction Of Worldwide Energy Resources*) da NASA. As covariáveis da NASA POWER são derivadas principalmente do *Global Modeling and Assimilation Office* (GMAO) *Modern Era Retrospective-Analysis for Research and Applications* (MERRA-2) (Gelaro et al., 2017) e do GMAO Forward Processing – Instrument Teams (FP-IT) *Goddard Earth Observing System* (GEOS 5.12.4) (Molod et al., 2015). O sistema de referência de coordenadas utilizado é *datum* geodésico *World Geodetic System* 1984 (WGS84). Para os conjuntos de dados de radiação, a resolução é de 1,0° de latitude por 1,0° de longitude. Para os conjuntos de dados meteorológicos, a resolução é de ½° de latitude por ⅝° de longitude. Os dados foram baixados em formato CSV utilizando o pacote *nasapower* (Sparks, 2018) do software R (R Core Team, 2024).

Dentre as covariáveis disponíveis na plataforma, foram selecionadas aquelas definidas como agroclimáticas, devido a sua relevância no contexto agronômico. Dados

diários dessas covariáveis foram coletados a partir do plantio até a data de colheita de cada experimento. De posse dos dados, avaliou-se a qualidade das mensurações, excluindo-se covariáveis com grande quantidade de observações ausentes em um ou mais experimentos.

Os dados de precipitação e temperatura média foram divididos em três categorias, de acordo com a época de cada experimento (Figura 4). Calculou-se o acumulado da precipitação e a média da temperatura nos seguintes períodos: 1 – Fase de brotação (30 dias após plantio); 2 – Desenvolvimento vegetativo (do fim da brotação até 1 de abril do próximo ano); 3 – Fase de maturação (de 1 de abril até a colheita). Essa abordagem visa aproximar os períodos de desenvolvimento da cana-de-açúcar comercial: brotação, desenvolvimento vegetativo e maturação (Figura 4). Para as demais covariáveis agroclimáticas, foi estimada a média geral de cada covariável. No total, obteve-se 32 covariáveis (Apêndice A), que posteriormente foram padronizadas para um novo conjunto de dados com média igual a zero e desvio padrão igual a um.

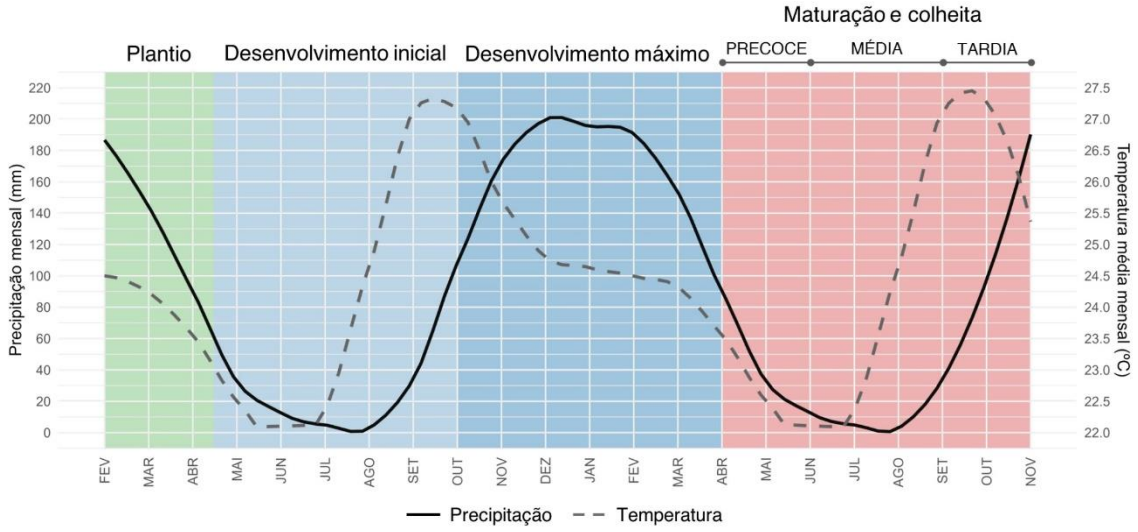


Figura 4. Médias de precipitação e temperatura ao longo do ciclo de cana-de-açúcar em 53 experimentos amostrados no Estado de Goiás e Tocantins. Dados climáticos: Plataforma NASA POWER (https://power.larc.nasa.gov/).

3.4 MODELOS PARA PREDIÇÃO

Neste estudo, foram estabelecidos seis modelos de regressão de cumeeira bayesiana (*Bayesian Ridge Regression* – BRR) para predizer o caráter TCH (Hoerl & Kennard, 1970). Todos os modelos foram implementados por meio do pacote

BGLR (Pérez & De Los Campos, 2014). O processo de amostragem de Gibbs foi realizado com 10.000 iterações, sendo *burn-in* de 1000 e *thin* de 5.

3.4.1 Modelo 1: G + E

Neste modelo, foram utilizados apenas os efeitos principais de genótipos e de experimentos para a predição de TCH, segundo a equação:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Z_1 a} + \boldsymbol{\varepsilon}$$

Em que: $\boldsymbol{y}$ é o vetor dos valores ajustados de TCH; μ é uma constante comum a todos elementos de $\boldsymbol{y}$; $\mathbf{1}$ é um vetor coluna de 1´s de mesma dimensão de $\boldsymbol{y}$; $\boldsymbol{a}$ é o vetor de efeitos aleatórios de genótipos e experimentos; $\boldsymbol{Z_1}$ é a matriz de incidência dos efeitos contidos em $\boldsymbol{a}$ e $\boldsymbol{\varepsilon}$ é o vetor de resíduos. As distribuições a *priori* assumidas para os efeitos aleatórios contidos em $\boldsymbol{a}$, para os resíduos e para seus respectivos componentes de variância foram, respectivamente: $\boldsymbol{a} \mid \sigma^2_a \sim N(0, \mathbf{I}\sigma^2_a)$; $\boldsymbol{\varepsilon} \mid \sigma^2_{\varepsilon} \sim N(0, \mathbf{I}\sigma^2_{\varepsilon})$; $\sigma^2_a \mid df_a, S_a \sim \chi_a^{-2}(df_a, S_a)$ e $\sigma^2_{\varepsilon} \mid df_{\varepsilon}, S_a \sim \chi_{\varepsilon}^{-2}(df_{\varepsilon}, S_{\varepsilon})$. As mesmas distribuições a *priori* foram assumidas para os efeitos aleatórios e resíduos dos demais modelos e no ambiente computacional R, foram adicionados valores ausentes (NAs) no vetor $\boldsymbol{y}$.

Neste modelo, $\boldsymbol{y}$ foi ordenado por experimentos e a matriz $\boldsymbol{Z_1}$ foi construída pela concatenação das matrizes de incidência de genótipos ($\boldsymbol{G}$, de dimensão 1961 × 37) e de experimentos ($\boldsymbol{E}$, de dimensão 1961 × 53), resultando em uma matriz de incidência global ($\boldsymbol{Z_1}$, dimensão 1961 × 90). Conforme a seguinte estrutura:

$$\boldsymbol{G} = \begin{bmatrix} I_{37} \\ I_{37} \\ I_{37} \\ \vdots \\ I_{37} \\ I_{37} \end{bmatrix} \quad \boldsymbol{E} = \begin{bmatrix} 1\,0\,0 \cdots 0 \\ 1\,0\,0 \cdots 0 \\ \vdots\ \vdots\ \vdots\ \ddots\ \vdots \\ 1\,0\,0 \cdots 0 \\ 0\,1\,0 \cdots 0 \\ 0\,1\,0 \cdots 0 \\ \vdots\ \vdots\ \vdots\ \ddots\ \vdots \\ 0\,1\,0 \cdots 0 \\ 0\,0\,1 \cdots 0 \\ 0\,0\,1 \cdots 0 \\ \vdots\ \vdots\ \vdots\ \ddots\ \vdots \\ 0\,0\,1 \cdots 0 \\ 0\,0\,0 \cdots 1 \\ \vdots\ \vdots\ \vdots\ \ddots\ \vdots \\ 0\,0\,0 \cdots 1 \end{bmatrix} \quad \boldsymbol{Z_1} = [\,\boldsymbol{G} \mid \boldsymbol{E}\,] \qquad (1)$$

3.4.2 Modelo 2: G + G × E

Neste modelo, o vetor de efeitos aleatórios contém efeitos principais de genótipos e os efeitos da interação genótipos × experimentos, segundo a equação:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Z_2 b} + \boldsymbol{\varepsilon}$$

Em que: $\boldsymbol{y}$ é o vetor dos valores ajustados de TCH; μ é uma constante comum a todos elementos de $\boldsymbol{y}$; $\mathbf{1}$ é um vetor coluna de 1´s de mesma dimensão de $\boldsymbol{y}$; $\boldsymbol{b}$ é o vetor de efeitos aleatórios de genótipos e da interação genótipos × experimentos; $\boldsymbol{Z_2}$ é a matriz de incidência dos efeitos contidos em $\boldsymbol{b}$ e $\boldsymbol{\varepsilon}$ é o vetor de resíduos. Neste modelo, $\boldsymbol{y}$ foi ordenado por genótipos e a matriz $\boldsymbol{Z_2}$ (dimensão 1961 × 1998) é uma matriz bloco diagonal em que cada bloco (B, dimensão 53 × 54) representa a estrutura de *design* de um genótipo, em que a primeira coluna é o intercepto e as demais colunas representam cada experimento, conforme a seguinte estrutura:

$$B = \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \dots & 1 \end{bmatrix} \quad \boldsymbol{Z_2} = \begin{bmatrix} B & 0 & 0 & \dots & 0 \\ 0 & B & 0 & \dots & 0 \\ 0 & 0 & B & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & B \end{bmatrix} \qquad (2)$$

3.4.3 Modelo 3: G + E + G × E

Esse modelo é uma expansão do modelo anterior, acrescido do efeito principal de experimento, segundo a equação:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Z_3 c} + \boldsymbol{\varepsilon}$$

Em que: $\boldsymbol{y}$ é o vetor dos valores ajustados de TCH; μ é uma constante comum a todos elementos de $\boldsymbol{y}$; $\mathbf{1}$ é um vetor coluna de 1´s de mesma dimensão de $\boldsymbol{y}$; $\mathbf{c}$ é o vetor de efeitos aleatórios de genótipos, experimentos e da interação genótipos × experimentos; $\boldsymbol{Z_3}$ é a matriz de incidência dos efeitos contidos em $\boldsymbol{c}$ e $\boldsymbol{\varepsilon}$ é o vetor de resíduos. O vetor $\boldsymbol{y}$ foi ordenado por experimentos e a matriz $\boldsymbol{Z_3}$ foi construída pela concatenação das matrizes de incidência de genótipos ($\boldsymbol{G}$, de dimensão 1961 × 37), de experimentos ($\boldsymbol{E}$, de dimensão 1961 × 53) e uma matriz identidade ($\boldsymbol{GE}$, dimensão 1961 × 1961), resultando em uma matriz de incidência global ($\boldsymbol{Z_3}$, dimensão 1961 × 2051). Dessa forma, a estrutura de $\boldsymbol{G}$, $\boldsymbol{E}$, $\boldsymbol{GE}$ e $\boldsymbol{Z_3}$ podem ser representadas como:

$$\boldsymbol{G} = \begin{bmatrix} I_{37} \\ I_{37} \\ I_{37} \\ \vdots \\ I_{37} \\ I_{37} \end{bmatrix} \quad \boldsymbol{E} = \begin{bmatrix} 1\,0\,0\cdots 0 \\ 1\,0\,0\cdots 0 \\ \vdots\;\vdots\;\vdots\;\ddots\;\vdots \\ 1\,0\,0\cdots 0 \\ 0\,1\,0\cdots 0 \\ 0\,1\,0\cdots 0 \\ \vdots\;\vdots\;\vdots\;\ddots\;\vdots \\ 0\,1\,0\cdots 0 \\ 0\,0\,1\cdots 0 \\ 0\,0\,1\cdots 0 \\ \vdots\;\vdots\;\vdots\;\ddots\;\vdots \\ 0\,0\,1\cdots 0 \\ 0\,0\,0\cdots 1 \\ \vdots\;\vdots\;\vdots\;\ddots\;\vdots \\ 0\,0\,0\cdots 1 \end{bmatrix} \quad \boldsymbol{GE} = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix} \quad \boldsymbol{Z_3} = [\boldsymbol{G} \mid \boldsymbol{E} \mid \boldsymbol{GE}] \qquad (3)$$

3.4.4 Modelo 4: G + V

Foram utilizados neste modelo, os efeitos principais de genótipos e das 32 covariáveis ambientais para a predição de TCH, segundo a equação:

$$\boldsymbol{y} = \boldsymbol{1}\mu + \boldsymbol{Z_4 d} + \boldsymbol{\varepsilon}$$

Em que: $\boldsymbol{y}$ é o vetor dos valores ajustados de TCH; μ é uma constante comum a todos elementos de $\boldsymbol{y}$; $\boldsymbol{1}$ é um vetor coluna de 1´s de mesma dimensão de $\boldsymbol{y}$; $\boldsymbol{d}$ é o vetor de efeitos aleatórios de genótipos das covariáveis ambientais; $\boldsymbol{Z_4}$ é a matriz de incidência dos efeitos contidos em $\boldsymbol{d}$ e $\boldsymbol{\varepsilon}$ é o vetor de resíduos. O vetor $\boldsymbol{y}$ foi ordenado por experimentos e a matriz $\boldsymbol{Z_4}$ foi construída pela concatenação das matrizes de incidência de genótipos ($\boldsymbol{G}$, de dimensão 1961 × 37) e de covariáveis ambientais ($\boldsymbol{V}$, de dimensão 1961 × 32) em *design* por experimento, resultando em uma matriz de efeito global ($\boldsymbol{Z_4}$, dimensão 1961 x 69) com a seguinte estrutura:

$$\boldsymbol{G} = \begin{bmatrix} I_{37} \\ I_{37} \\ I_{37} \\ \vdots \\ I_{37} \\ I_{37} \end{bmatrix} \quad \boldsymbol{V} = \begin{bmatrix} V_{11} & V_{12} & V_{13} & \ldots & V_{1,32} \\ V_{11} & V_{12} & V_{13} & \ldots & V_{1,32} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ V_{11} & V_{12} & V_{13} & \ldots & V_{1,32} \\ V_{21} & V_{22} & V_{23} & \ldots & V_{2,32} \\ V_{21} & V_{22} & V_{23} & \ldots & V_{2,32} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ V_{21} & V_{22} & V_{23} & \ldots & V_{2,32} \\ V_{31} & V_{32} & V_{33} & \ldots & V_{3,32} \\ V_{31} & V_{32} & V_{33} & \ldots & V_{3,32} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ V_{53,1} & V_{53,2} & V_{53,3} & \ldots & V_{53,32} \end{bmatrix} \quad \boldsymbol{Z_4} = [\,\boldsymbol{G} \mid \boldsymbol{V}\,] \qquad (4)$$

3.4.5 Modelo 5: G + G × V

Neste modelo, o vetor de efeitos aleatórios contém efeitos principais de genótipos e os efeitos da interação genótipos × covariáveis, segundo a equação:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Z_5 e} + \boldsymbol{\varepsilon}$$

Em que: $\boldsymbol{y}$ é o vetor dos valores ajustados de TCH; μ é uma constante comum a todos elementos de $\boldsymbol{y}$; $\mathbf{1}$ é um vetor coluna de 1´s de mesma dimensão de $\boldsymbol{y}$; $\boldsymbol{e}$ é o vetor de efeitos aleatórios de genótipos e da interação genótipos × covariáveis; $\boldsymbol{Z_5}$ é a matriz de incidência dos efeitos contidos em $\boldsymbol{e}$ e $\boldsymbol{\varepsilon}$ é o vetor de resíduos. O vetor $\boldsymbol{y}$ foi ordenado por genótipos e a matriz $\boldsymbol{Z_5}$ (dimensão 1961 × 1221) é uma matriz bloco diagonal em que cada bloco (B, dimensão 53 × 33) representa a estrutura de *design* de um genótipo, em que a primeira coluna é o intercepto e as demais colunas representam as covariáveis ambientais, conforme a seguinte estrutura:

$$B = \begin{bmatrix} 1 & V_{11} & V_{12} & V_{13} & \dots & V_{1,32} \\ 1 & V_{21} & V_{22} & V_{23} & \dots & V_{2,32} \\ 1 & V_{31} & V_{32} & V_{33} & \dots & V_{3,32} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & V_{53,1} & V_{53,2} & V_{53,3} & \dots & V_{53,32} \end{bmatrix} \quad \boldsymbol{Z_5} = \begin{bmatrix} B & 0 & 0 & \dots & 0 \\ 0 & B & 0 & \dots & 0 \\ 0 & 0 & B & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & B \end{bmatrix} \quad (5)$$

3.4.6 Modelo 6: G + V + G × V

Esse modelo é uma expansão do modelo anterior, adiciona-se o efeito principal das covariáveis ambientais, segundo a equação:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Z_6 f} + \boldsymbol{\varepsilon}$$

Em que: $\boldsymbol{y}$ é o vetor dos valores ajustados de TCH; μ é uma constante comum a todos elementos de $\boldsymbol{y}$; $\mathbf{1}$ é um vetor coluna de 1´s de mesma dimensão de $\boldsymbol{y}$; $\boldsymbol{f}$ é o vetor de efeitos aleatórios de genótipos, covariáveis ambientais e da interação genótipos × covariáveis; $\boldsymbol{Z_6}$ é a matriz de incidência dos efeitos contidos em $\boldsymbol{f}$ e $\boldsymbol{\varepsilon}$ é o vetor de resíduos. O vetor $\boldsymbol{y}$ foi ordenado por genótipos e a matriz $\boldsymbol{Z_6}$ (dimensão 1961 × 1253) foi construída pela concatenação das matrizes de incidência de genótipos ($\boldsymbol{G}$, de dimensão 1961 × 37), de covariáveis ambientais ($\boldsymbol{V}$, de dimensão 1961 × 32) e uma matriz de bloco diagonal, denominada $\boldsymbol{GV}$ (dimensão 1961 × 1184), em que cada bloco (B, dimensão 53 × 32) contém as covariáveis ambientais em *design* por genótipo, conforme a seguinte estrutura:

$$\boldsymbol{G} = \begin{bmatrix} 1\,0\,0\cdots 0 \\ 1\,0\,0\cdots 0 \\ \vdots\;\vdots\;\vdots\;\ddots\;\vdots \\ 1\,0\,0\cdots 0 \\ 0\,1\,0\cdots 0 \\ 0\,1\,0\cdots 0 \\ \vdots\;\vdots\;\vdots\;\ddots\;\vdots \\ 0\,1\,0\cdots 0 \\ 0\,0\,1\cdots 0 \\ 0\,0\,1\cdots 0 \\ \vdots\;\vdots\;\vdots\;\ddots\;\vdots \\ 0\,0\,1\cdots 0 \\ 0\,0\,0\cdots 1 \\ \vdots\;\vdots\;\vdots\;\ddots\;\vdots \\ 0\,0\,0\cdots 1 \end{bmatrix} \quad \boldsymbol{V} = \begin{bmatrix} V_{11} & V_{12} & V_{13} & \cdots & V_{1,32} \\ V_{21} & V_{22} & V_{23} & \cdots & V_{2,32} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ V_{53,1} & V_{53,2} & V_{53,3} & \cdots & V_{53,32} \\ V_{11} & 1_{22} & V_{13} & \cdots & V_{1,32} \\ V_{21} & V_{22} & V_{23} & \cdots & V_{2,32} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ V_{53,1} & V_{53,2} & V_{53,3} & \cdots & V_{53,32} \\ V_{11} & V_{12} & V_{13} & \cdots & V_{1,32} \\ V_{21} & V_{22} & V_{23} & \cdots & V_{2,32} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ V_{53,1} & V_{53,2} & V_{53,3} & \cdots & V_{53,32} \end{bmatrix} \tag{6}$$

$$B = \begin{bmatrix} V_{11} & V_{12} & V_{14} & \cdots & V_{1,32} \\ V_{21} & V_{22} & V_{24} & \cdots & V_{2,32} \\ V_{31} & V_{32} & V_{34} & \cdots & V_{3,32} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ V_{53,1} & V_{53,3} & V_{53,4} & \cdots & V_{53,32} \end{bmatrix} \quad \boldsymbol{GV} = \begin{bmatrix} B & 0 & 0 & \ldots & 0 \\ 0 & B & 0 & \ldots & 0 \\ 0 & 0 & B & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & B \end{bmatrix}$$

$$\boldsymbol{Z_6} = [\,\boldsymbol{G} \mid \boldsymbol{V} \mid \boldsymbol{GV}\,]$$

3.5 VALIDAÇÃO CRUZADA E HABILIDADES PREDITIVAS

Dois esquemas de validação cruzada foram utilizados para avaliar as habilidades preditivas dos modelos descritos anteriormente. Cada esquema representou um desafio preditivo diferente dentro dos programas de melhoramento. O primeiro esquema avaliou a capacidade dos modelos predizerem o desempenho de genótipos em experimentos nos quais esses não foram plantados, de modo a preencher lacunas no banco de dados e proporcionar uma avaliação mais abrangente do potencial genotípico. Para isso, utilizou-se a validação cruzada 5-*fold* (Apêndice B). O conjunto de valores ajustados de TCH foi dividido aleatoriamente em cinco subconjuntos (5-*folds*). Os modelos foram treinados utilizando quatro dos subconjuntos, enquanto o subconjunto restante foi reservado para a validação. Ao final dos cinco *folds* todas as observações foram preditas. Esse processo foi repetido 30 vezes para estimar a variação dos resultados.

O segundo esquema avaliou as habilidades dos modelos com covariáveis ambientais (4, 5 e 6) de predizerem a performance de genótipos em ambientes desconhecidos. Para isto, utilizou-se a validação cruzada *leave-one-out* (Apêndice C), em que o conjunto de dados foi dividido em 53 subgrupos (53-*fold*), no qual cada *fold*

corresponde a um experimento que foi excluído da população de treinamento. Esse cenário reflete uma situação na qual deseja-se identificar os melhores genótipos em um ambiente que não existem dados experimentais prévios. Esse processo também foi repetido 30 vezes para estimativas de variação dos resultados.

Para ambos os esquemas, a habilidade preditiva foi avaliada de duas maneiras: (1) Habilidade preditiva global: medida pelo coeficiente de correlação de Pearson entre todos os valores observados e preditos. Esse valor se refere a média deste coeficiente de correlação das 30 rodadas de validação cruzada; (2) Habilidade preditiva por experimento: medida pelo coeficiente de correlação de Spearman médio entre os valores observados e preditos das 30 rodadas de validação cruzada, considerando cada experimento separadamente. A utilização do coeficiente de Spearman teve o propósito de avaliar a habilidade de predizer o ranqueamento dos genótipos em cada experimento e não a produtividade em si. Para garantir maior confiabilidade, somente experimentos com cinco ou mais genótipos foram utilizados para calcular o coeficiente de correlação de Spearman por experimento, totalizando 34 experimentos (Figura 3). As habilidades preditivas de cada modelo foram agrupadas pelo teste de Scott–Knott a 5% de significância.

4 RESULTADOS

4.1 ANÁLISES FENOTÍPICAS

Na análise de deviance conjunta, o Teste de Razão de Verossimilhança (LRT), indicou que a variância foi significativa para os fatores experimentos, genótipos e interação genótipos × experimentos (Tabela 2). Destaca-se o efeito dos experimentos, responsável pela maioria da variação total, cerca de dez vezes maior que a variância de genótipo e cinco vezes maior que interação G × E (Tabela 2).

Tabela 2. Análise de deviance dos dados fenotípicos segundo o modelo estatístico conjunto: $\boldsymbol{y = \mu + Z_{01}r + Z_{02}g + Z_{03}e + Z_{04}ge + \varepsilon}$

Fonte de variação	Variância	Fração da variância	P(χ^2)
Experimentos (E)	1026,11	0,575	<0,0001
Genótipos (G)	101,41	0,057	<0,0001
Interação G × E	208,82	0,117	<0,0001
Blocos	44,98	0,025	<0,0001
Resíduos	403,43	0,226	

P(χ^2): p-valor obtido usando a distribuição qui-quadrado.

Com base nos BLUPs de experimentos obtidos na análise conjunta, os experimentos foram ordenados em um gradiente ambiental, o experimento considerado de pior qualidade foi o UOL2017, enquanto o de melhor qualidade foi TROP2019 (Figura 5). Observa-se que os experimentos não se agrupam por unidades, indicando a influência do fator ano no desempenho dos genótipos (Figura 5). Destacam-se os genótipos RB988082, RB975242, RB034045 por apresentarem os maiores BLUPs genotípicos.

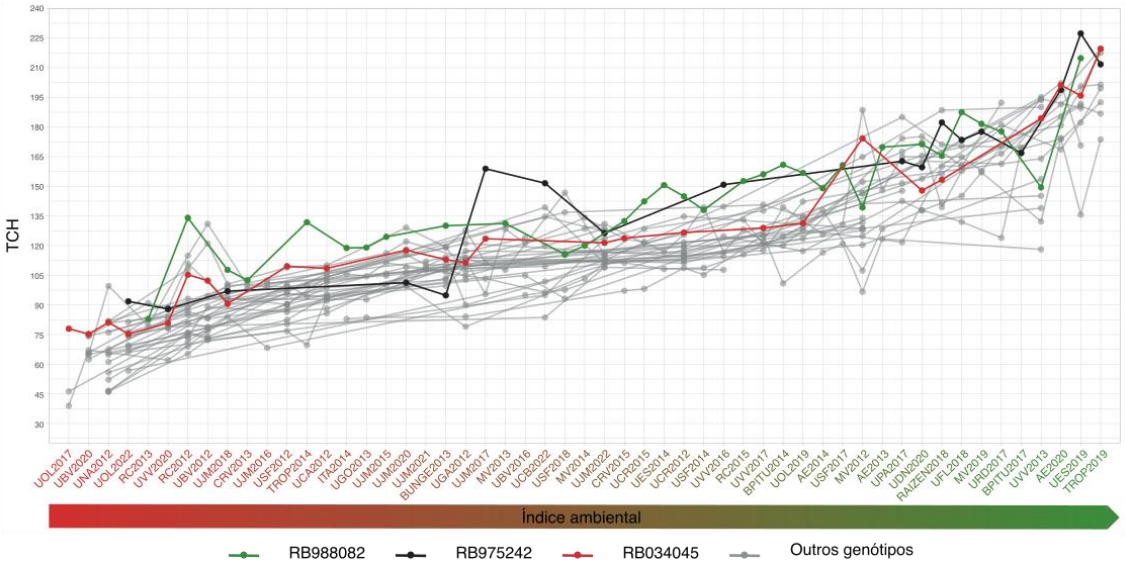


Figura 5. Visualização gráfica das interações genótipos com experimentos. No eixo X, experimentos foram ordenados de piores a melhores, segundo os seus respectivos BLUPs em relação a covariável resposta TCH.

4.2 HABILIDADES PREDITIVAS VALIDAÇÃO 5-*FOLD*

Os modelos 1, 2 e 3 apresentaram uma habilidade preditiva global de 0,93, 0,32 e 0,92, respectivamente (Figura 6; Tabela 3). A adição dos efeitos da interação G × E no modelo 3 (G + E + G × E) não melhorou a predição em relação ao modelo 1 (G + E), o mais básico, contendo apenas os efeitos principais de genótipos e de experimentos para a predição de TCH. Por outro lado, ao remover o efeito principal do fator experimento, no modelo 2 (G + G × E), houve queda na predição de 0,93 para 0,32. Em relação a performance dos modelos com covariáveis ambientais, nenhum deles superou a capacidade de predição global quando comparados aos modelos 1 e 3, sem covariáveis. Os modelos 4, 5 e 6, apresentaram habilidade preditiva global de 0,79, 0,13 e 0,54, respectivamente, destacando-se, dentre esses três, o primeiro, G +V (Figura 6; Tabela 3).

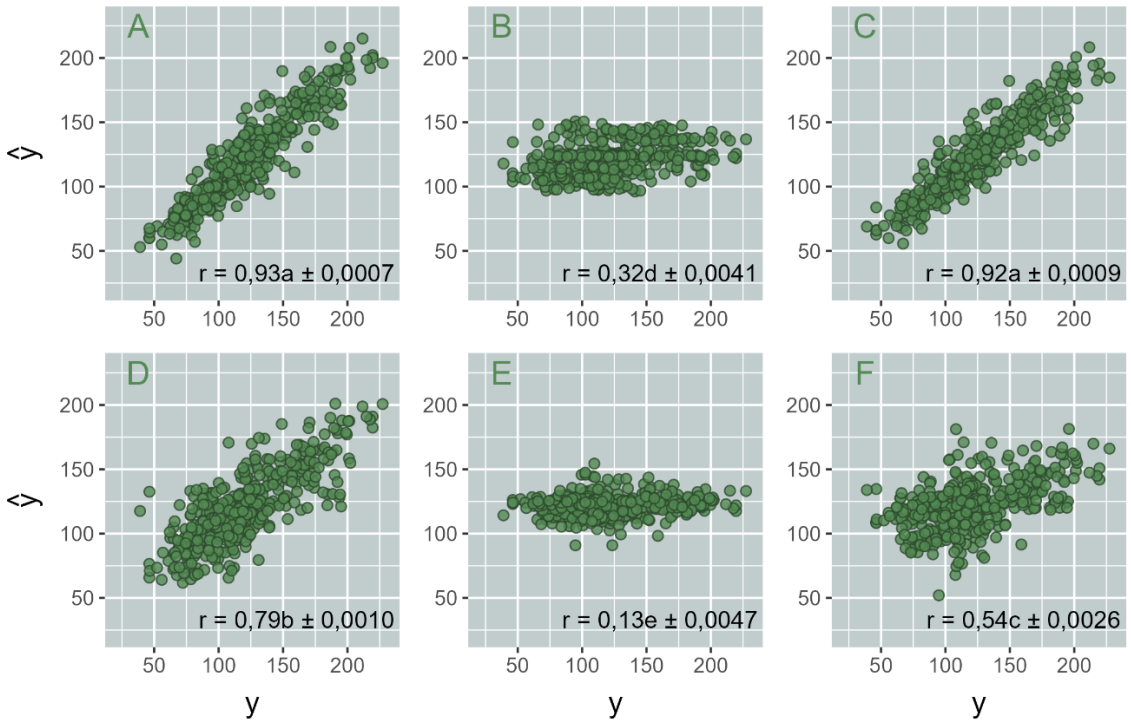


Figura 6. Gráficos de dispersão TCH observado (y) e TCH predito (ŷ). A: Modelo 1 (G + E); B: Modelo 2 (G + G × E); C: Modelo 3 (G + E + G × E); D: Modelo 4 (G + V); E: Modelo 5 (G + G × V); F: Modelo 6 (G + V + G × V). Valor r refere-se ao coeficiente de correlação de Pearson: r(y, ŷ). Letras minúsculas após valor r indicam o agrupamento das habilidades preditivas pelo teste de Scott–Knott a 5% de significância. Valores após o símbolo "±" referem-se ao erro padrão de r das 30 rodadas da validação 5-*fold*.

As habilidades preditivas analisadas por experimento foram agrupadas pelo teste de Scott-Knott em dois grupos (Figura 7; Tabela 3). Os modelos 5 e 6, respectivamente G + G × V ($\bar{r}$ = 0,17) e G + V + G × V ($\bar{r}$ = 0,16), foram alocados no grupo inferior. Para os demais modelos, estimativas de predição foram torno de 0,50 (Figura 7; Tabela 3). Ao substituir o efeito de experimentos no modelo G + E pelo efeito de covariáveis ambientais em G + V, a habilidade preditiva por experimento se manteve. No entanto, essa substituição em modelos que incluem interação levou a habilidades preditivas mais baixas. Além disso, observou-se uma inferioridade geral na predição em todos os modelos para os experimentos BUNGE2013 e URD2017, cujas habilidades preditivas foram inferiores a 0,20 (Figura 7).

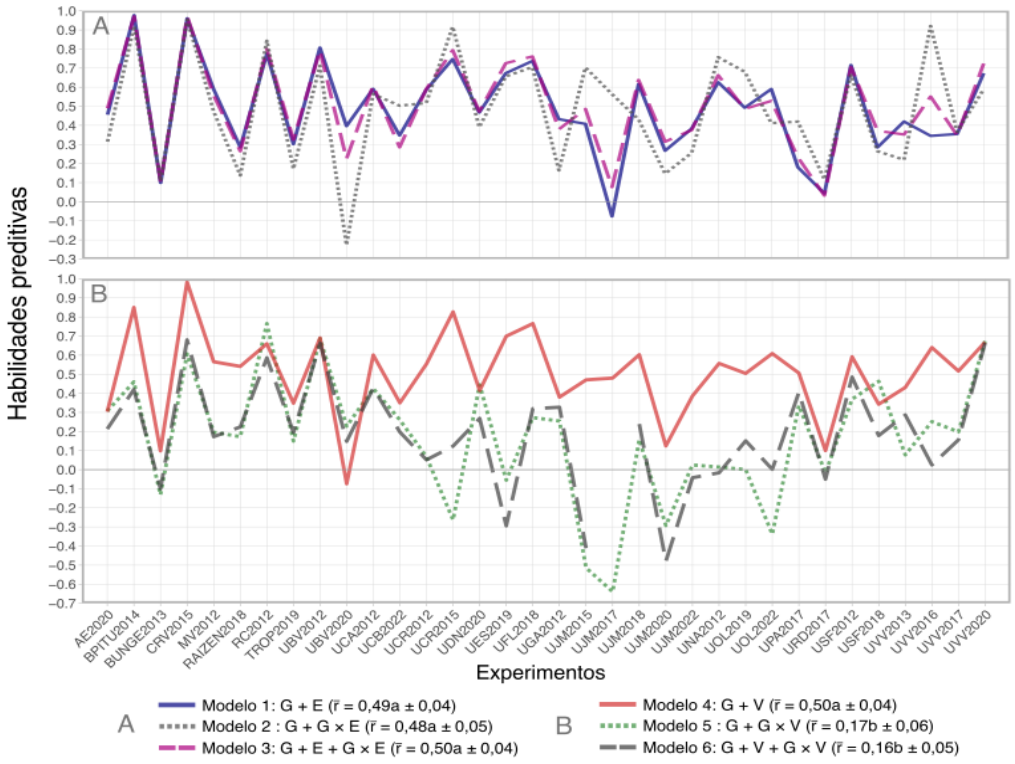


Figura 7. Habilidades preditivas por experimento. A: Modelos sem covariáveis ambientais; B: Modelos com covariáveis. Valores de habilidade preditiva referem-se ao coeficiente de correlação de Spearman médio de 30 rodadas de validação cruzada 5-*fold*, considerando cada experimento separadamente. Letras minúsculas após os valores de habilidades médias (r̄) indicam o agrupamento das habilidades preditivas pelo teste de Scott–Knott ao nível de 5% de significância. Valores após o símbolo "±" referem-se ao erro padrão da habilidade preditiva entre estes experimentos.

Tabela 3. Habilidades preditivas na validação 5-*fold* dos modelos de regressão de cumeeira bayesiana (*Bayesian Ridge Regression* – BRR) estabelecidos.

Modelos	Efeitos aleatórios	Número de efeitos estimados	Habilidades preditivas globais	Habilidades preditivas por experimento
1	G + E	90	0,93a	0,49a
2	G + G × E	1998	0,32d	0,48a
3	G + E + G × E	2051	0,92a	0,50a
4	G + V	69	0,79b	0,50a
5	G + G × V	1221	0,13e	0,17b
6	G + V + G × V	1253	0,54c	0,16b

G representa efeito de genótipo; E de experimento e V de covariáveis ambientais. Letras minúsculas após valores de habilidades preditivas indicam o agrupamento das habilidades preditivas pelo teste de Scott–Knott a 5% de significância.

4.3 HABILIDADES PREDITIVAS VALIDAÇÃO *LEAVE-ONE-OUT*

As habilidades preditivas globais dos modelos 4, 5 e 6 foram de 0,47, 0,01 e 0,24, respectivamente (Figura 8; Tabela 4). O teste de Scott-Knott agrupou essas estimativas nos grupos a, c e b, respectivamente, assim, se destaca o modelo 4 (G + V), o mais básico dentre os modelos analisados na validação *leave-one-out*. Entretanto, para esse modelo, alguns dos valores preditos para o experimento BUNGE2013, foram não interpretáveis em termos fenotípicos (valores negativos de TCH) (Figura 8-A).

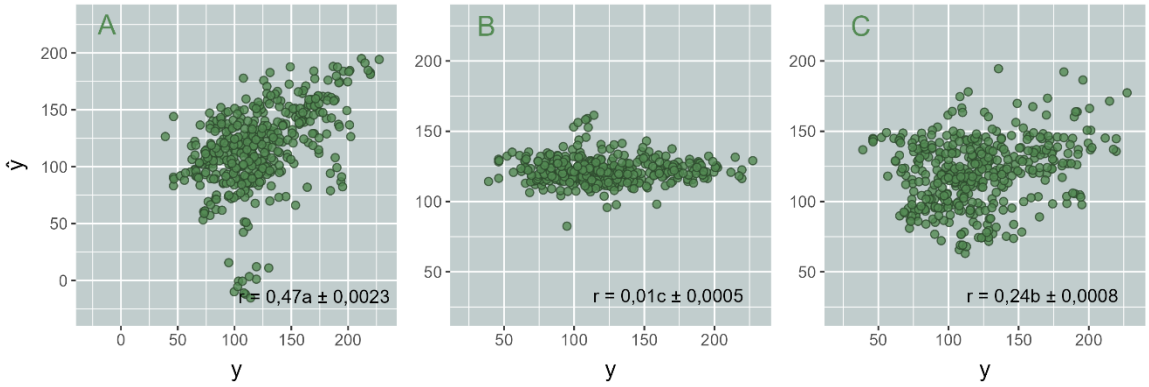


Figura 8. Gráficos de dispersão TCH observado (y) e TCH predito (ŷ). A: Modelo 4 (G + V); B: Modelo 5 (G + G × V); C: Modelo 6 (G + V + G × V). Valor r refere-se ao coeficiente de correlação de Pearson: r(y, ŷ). Letras minúsculas após valor r indicam o agrupamento das habilidades preditivas pelo teste de Scott–Knott a 5% de significância. Valores após o símbolo "±" referem-se ao erro padrão de r das 30 rodadas de validação *leave-one-out*.

Em relação as habilidades preditivas analisadas por experimento (Figura 9; Tabela 4), o modelo 4 obteve média de 0,58, o modelo 5, 0,21 e o modelo 6, 0,25. O teste de Scott-Knott agrupou essas estimativas nos grupos a, b e b, respectivamente. Similar ao observado na validação 5-*fold*, foi possível perceber a inferioridade da predição em alguns experimentos. Os modelos 5 e 6 (G + G × V e G + V + G × V), que apresentaram comportamento semelhante em relações as correlações por experimento, tiveram predições negativas em 10 experimentos, enquanto que as habilidades preditivas do modelo 4 (G + V) variaram de 0,1 a 1, sem valores negativos (Figura 9). Dos 34 experimentos aptos para estas estimativas, 19 tiveram predições com r maior ou igual a 0,60 pelo modelo 4 (G + V). Destaca-se o experimento CRV2015, com cinco genótipos,

em que o modelo 4 acertou 100% do ranqueamento em todas as rodadas de *leave-one-out* (Figura 9).

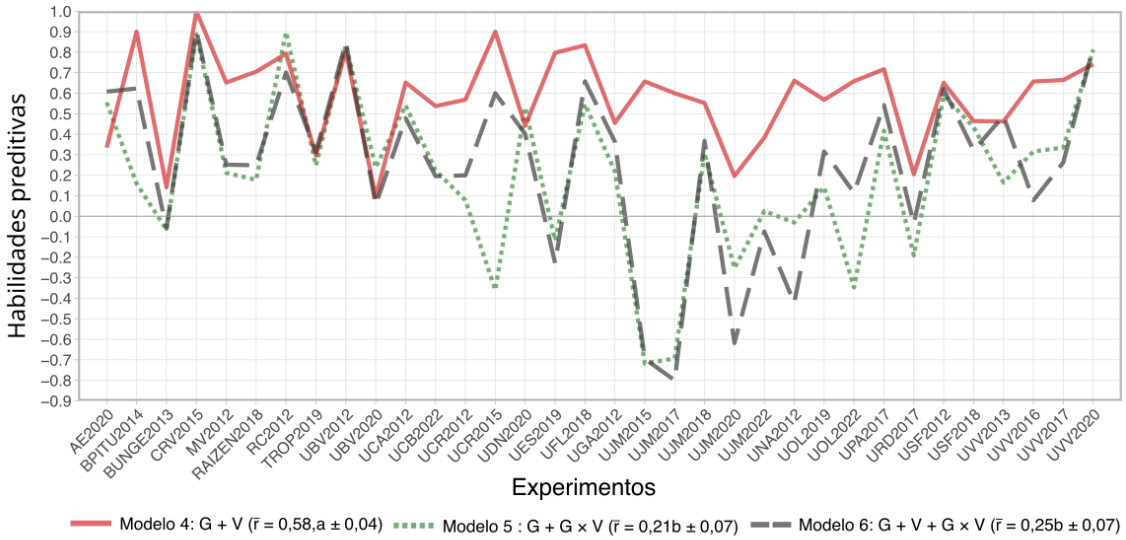


Figura 9. Habilidades preditivas por experimento, considerando experimentos com cinco ou mais genótipos. Valores de habilidade preditiva referem-se ao coeficiente de correlação de Spearman médio de 30 rodadas de validação cruzada *leave-one-out*, considerando cada experimento separadamente. Letras minúsculas após os valores de habilidades médias (r̄) indicam o agrupamento das habilidades preditivas pelo teste de Scott–Knott ao nível de 5% de significância. Valores após o símbolo "±" referem-se ao erro padrão da habilidade preditiva entre estes experimentos.

Tabela 4. Habilidades preditivas na validação *leave-one-out* dos modelos de regressão de cumeeira bayesiana (*Bayesian Ridge Regression* – BRR) estabelecidos.

Modelos	**Efeitos aleatórios**	**Número de efeitos estimados**	**Habilidades preditivas globais**	**Habilidades preditivas por experimento**
4	G + V	69	0,47a	0,58a
5	G + G × V	1221	0,01c	0,21b
6	G + V + G × V	1253	0,24b	0,25b

G representa efeito de genótipo e V de covariáveis ambientais. Letras minúsculas após valores de habilidades preditivas indicam o agrupamento das habilidades preditivas pelo teste de Scott–Knott a 5% de significância.

5 DISCUSSÃO

Os resultados deste trabalho destacam a influência do ambiente na produtividade da cana-de-açúcar em termos de toneladas de colmos por hectare (TCH). Em ensaios multiambientais de cana-de-açúcar, a maior contribuição do fator ambientes (neste trabalho experimentos) em relação a outras fontes de variação também foi encontrada por Otieno & Owuor (2019), Montes et al. (2021) e Mehareb et al. (2022). Essa predominância pode ser interpretada como uma evidência da complexidade do cultivo da cana-de-açúcar em diferentes regiões. O ambiente influencia diretamente a produtividade, refletindo variações em clima, solo e manejo. Em ensaios multiambientais, essa variação é esperada, pois o sucesso de uma cultivar depende não apenas do seu potencial genético, mas também de quão bem ela se adapta a condições ambientais específicas.

Devido a essa variabilidade, é essencial que os programas de melhoramento incluam testes em uma ampla gama de ambientes para identificar cultivares estáveis e adaptáveis para as diferentes áreas de cultivos. Resende & Bondrani (2024) destacam que trabalhar com ensaios desbalanceados é uma abordagem eficaz para captar mais informações de diferentes ambientes. Além disso, Resende et al. (2021) discutem que conjuntos de dados menores, mesmo que balanceados, não trazem vantagens em termos de predição. Ambos os estudos destacam a importância de ter o máximo possível de informação ao nível "ômico-ambiental".

No caso dos modelos preditivos 1 (G + E) e 3 (G + E + G × E), a utilização do efeito principal dos experimentos através de uma matriz de incidência foi fundamental para obter boas habilidades preditivas globais. Em todos os modelos com efeito de genótipos e de experimentos, quando determinada combinação de genótipo *i* com experimento *j* sai do conjunto de treinamento, essa interação fica impossibilitada de ser estimada e, portanto, não contribui na habilidade de predição. Por essa razão, os modelos G + E e G + E + G × E tem habilidades preditivas muito similares. Por outro lado, a inclusão de covariáveis ambientais representa uma grande vantagem. Quando se utiliza as covariáveis ambientais, a conexão entre ambientes e genótipos é mantida, pois passa-

se a ter informação daquelas covariáveis ambientais mesmo que a combinação de genótipo *i* com covariável *j* tenha ido para o conjunto de validação. Trabalhos como o de Costa-Neto et al. (2022) e Resende et al. (2024b) têm demonstrado o potencial da incorporação de covariáveis ambientais em modelos preditivos para lidar com cenários não experimentados.

Entre os modelos com covariáveis ambientais deste estudo, G + V se destaca, mesmo sem considerar interações genótipos × covariáveis. Assim, o valor do coeficiente de correlação entre os preditos e observados por experimento depende exclusivamente do valor observado, pois o ranqueamento predito se mantêm devido à ausência de interação. Experimentos com boas habilidades preditivas no modelo G + V são experimentos representativos da população de ambientes alvo. Esses ambientes refletem de forma precisa o ranqueamento médio dos genótipos na população, sendo, portanto, ambientes-chave para a avaliação de novos genótipos em programas de melhoramento. Cruz et al. (2025) propõem uma abordagem baseada em ferramentas de SIG (Sistema de Informação Geográfica) para definir a população de ambientes alvo, permitindo delimitar espacialmente as regiões de interesse para os estudos ambientômicos. Essa definição mais precisa dos ambientes pode contribuir para o delineamento de ensaios mais informativos e otimizados.

A habilidade preditiva observada em torno de 0,60 no modelo G + V também se deve, em parte, à estabilidade dos genótipos avaliados, que já estão em fases finais do programa de melhoramento. Nessa etapa do PMGCA/UFG/Ridesa, boa parte dos genótipos selecionados tendem a apresentar desempenho consistente em diferentes ambientes. Nos modelos que consideram a interação genótipos × covariáveis ambientais, a inclusão de interação parece introduzir ruído ao invés de melhorar a predição. Isso é evidenciado pelo fato de que, mesmo em experimentos onde a predição com o modelo simples G + V é insatisfatória, a inclusão das interações genótipos × covariáveis ambientais não resulta em melhorias significativas na habilidade preditiva.

Embora modelos mais complexos possam capturar interações mais detalhadas, na prática, esses modelos podem sofrer problemas como a captura de variabilidade irrelevante para a predição geral (Brighton & Gigerenzer, 2015). Brighton & Gigerenzer (2015) também discutem como a estrutura dos dados pode explicar por que, em certas situações, modelos mais simples superam os mais complexos. Segundo os autores, conforme a base de dados aumenta, modelos mais complexos podem ser

ajustados de maneira mais eficaz, mas, no atual cenário e com a quantidade de observações fenotípicas deste trabalho, a simplicidade foi vantajosa.

Segundo Brighton & Gigerenzer (2015), os principais contribuintes para a complexidade de um modelo são o número de parâmetros e a sua forma funcional, ou seja, como as covariáveis independentes estão relacionadas à covariável de resposta. Green & Armstrong (2015), comparando modelos preditivos em várias áreas da ciência, evidenciaram que modelos considerados mais simples frequentemente apresentam maior ou similar acurácia em relação aos modelos mais complexos. Para os autores, a complexidade de um modelo nem sempre é em função do número de covariáveis, ou do esforço requerido para desenvolvê-lo.

Os fatores ambientais que determinam as características de um ambiente são numerosos, complexos e frequentemente apresentam alta colinearidade entre si (Costa-Neto et al., 2020a). Essa complexidade dificulta a substituição do efeito global de um experimento por um conjunto limitado de covariáveis ambientais. Ainda que tenham sido selecionadas covariáveis relacionadas à produtividade da cana-de-açúcar, tais como temperatura em diferentes fases fenológicas, precipitação, altitude, longitude (Montes et al., 2021; Ramburan et al., 2011), essas covariáveis podem não capturar efetivamente os efeitos combinados e dinâmicos que definem as condições ambientais de um experimento. Saber como lidar com grandes volumes de dados ambientais para caracterização dentro de um ensaio ou entre ensaios, assim como a caracterização da população de ambientes alvo é um atributo fundamental na predição de ambientes representativos que podem nunca ter sido testados experimentalmente (Resende et al., 2022).

Uma realidade nos programas de melhoramento é a dificuldade na execução de ensaios multiambientais amplos e balanceados, devido ao alto custo de grandes experimentos e às perdas de parcelas causadas por condições não controláveis. Os modelos 1 e 3, aqui apresentados, possuem potencial para completar as lacunas de informações nos bancos de dados dos programas de melhoramento. Isso possibilita a expansão da rede de ensaios multiambientais, garantindo maior representatividade dos ambientes de produção, sem a necessidade de plantar todos os genótipos em todos os experimentos, sendo feita a predição dos genótipos não plantados. Dessa forma, é possível realizar predições para os genótipos não plantados, reduzindo o custo de experimentação e acelerando o processo de tomada de decisões em programas de melhoramento.

Embora o modelo 4 tenha apresentado habilidades preditivas superiores no cenário de *leave-one-out*, o experimento BUNGE2013, o único localizado no estado de Tocantins, destacou-se como uma exceção. Diferentemente dos demais experimentos, BUNGE 2013 apresentou valores preditos de TCH negativos. Essa limitação pode ser atribuída a condições ambientais locais, como diferenças na temperatura, distribuição de chuvas e características do solo, que não foram devidamente capturadas pelas covariáveis ambientais utilizadas no modelo. Além disso, a menor representatividade dessa região no conjunto de dados de treinamento pode ter contribuído para a ineficiência do modelo em ambientes menos conhecidos. Neste experimento, a habilidade foi 0,14 pelo modelo G + V.

Os resultados obtidos no experimento BUNGE2013 ressaltam a importância de incluir maior diversidade ambiental nos dados. Nos modelos que consideraram interações genótipos × covariáveis ambientais (G + G × V e G + G + G × V), os valores preditos de TCH não foram negativos, embora esses modelos não tenham apresentado habilidades preditivas superiores ao G + V. Isso sugere que a inclusão das interações, mesmo não resultando em melhores habilidades preditivas por experimento, contribuiu para gerar valores preditos mais coerentes com a realidade, evitando extrapolações extremas ou inconsistentes.

Futuras investigações devem focar nas causas subjacentes que dificultam a predição em experimentos específicos, o que pode envolver fatores como variações sazonais, interações ambientais e fatores não controláveis de manejo que os modelos não conseguem capturar adequadamente. A identificação dessas falhas é essencial para planejar o uso estratégico das predições em ambientes onde os modelos demonstram maior eficiência, além de orientar no planejamento de novos experimentos em ambientes onde a predição se mostrou ineficiente.

6 CONCLUSÕES

Os resultados deste trabalho evidenciam o desempenho diferencial dos modelos testados para predição em ensaios multiambientais de cana-de-açúcar. As habilidades preditivas obtidas pela análise global foram superiores às estimadas por experimento, evidenciando que a predição direcionada para um ambiente é desafiadora, pois envolve condições únicas e complexas. A capacidade dos modelos predizerem o desempenho de genótipos varia substancialmente entre os experimentos. Assim, identificar os experimentos nos quais a predição é ineficaz é crucial para o planejamento de novas estratégias de coleta de dados e para o aprimoramento de modelos que sejam capazes de lidar com essa complexidade ambiental de forma eficaz.

Os modelos 1 e 3 se mostraram adequados para preencher lacunas de informações nos bancos de dados de programas de melhoramento genético. Esses modelos podem ser úteis para expandir redes de ensaios multiambientais, possibilitando a predição de genótipos não testados em determinados ambientes, o que reduz custos e acelera decisões no processo de seleção. O modelo 4 (G + V), que utiliza apenas o efeito principal das covariáveis ambientais, destacou-se como a opção mais promissora para recomendações de cultivares em locais sem informações sobre seu desempenho.

Embora cada experimento avaliado apresente características únicas, os resultados do modelo G + V indicam que o desempenho geral dos genótipos avaliados tende a ser relativamente estável, independentemente das variações ambientais aqui analisadas. Isso sugere que é possível fazer predições razoáveis sem considerar interações. Além disso, os resultados destacam a importância de escolher o modelo preditivo com base no contexto de uso e nas informações disponíveis. Modelos simples, como o 1 e 4, oferecem soluções viáveis para predições do desempenho de genótipos nos ambientes do PMGCA/UFG/Ridesa.

7 REFERÊNCIAS

ALBUQUERQUE, J. R. T. D.; LINS, H. A.; SANTOS, M. G. D.; FREITAS, M. A. M. D.; SILVEIRA, L. M. D.; NUNES, G. H. S. et al. Environmental variables in the G x E interaction in soybean in the semiarid. **Ciência Agronômica**, v. 54, n. 1, p. e11-e23, 2023.

ALVES, N. C. G. F; WANDER, A. E. Competividade da produção de cana-de-açúcar no Estado de Goiás. *In*: TAVARES NETO, J. Q. (coord). **O estado e desenvolvimento regional:** realidade e perspectivas do Centro-Oeste brasileiro. Franca: Lemos & Cruz, 2012. cap. 5, p. 105-128.

AMALRAJ, V. A.; BALASUNDARAM, N. On the taxonomy of the members of 'Saccharum complex'. **Genetic Resources and Crop Evolution**, v. 53, n. 1, p. 35-41, 2006.

AMARAL, A. L.; SANTOS, J. M.; BARBOSA, G. V. S. **Denominação de cultivares de cana-de-açúcar como suporte à documentação e utilização do germoplasma conservado**. Aracajú: Embrapa Tabuleiros Costeiros, 2015. 31 p.

ANDRADE, J. C. **Esforço histórico das antigas variedades de cana-de-açúcar**. Alagoas: Asplana, 1985.

ANNICCHIARICO, P. Case study: durum wheat in Algeria. *In*: ANNICCHIARICO, P. **Genotype x environment interactions**: challenges and opportunities for plant breeding and cultivar recommendations. Rome: FAO, 2002. cap. 8, p 89-96.

ANNICCHIARICO, P. Cultivar adaptation and recommendation from alfalfa trials in Northern Italy. **Journal of Genetics and Breeding**, v. 46, p. 269-269, 1992.

ARALDI, R.; SILVA, F. M. L.; ONO, E. O.; RODRIGUES, J. D. Florescimento em cana-de-açúcar. **Ciência Rural**, v. 40, n. 3, p. 694-702, 2010.

BABU, K. S. D.; JANAKIRAMAN, V.; PALANISWAMY, H.; KASIRAJAN, L.; GOMATHI, R.; RAMKUMAR, T. R. A short review on sugarcane: its domestication, molecular manipulations and future perspectives. **Genetic Resources and Crop Evolution**, v. 69, n. 8, p. 2623-2643, 2022.

BAKKER, H. **Sugar cane cultivation and management**. 1. ed. Horsham: Springer, 1999. 679 p.

BARROS, F. F.; MILAN, M. Qualidade operacional do plantio de cana-de-açúcar. **Bragantia**, v. 69, n. 1, p 221-229, 2010.

BATES, D. M. lme4: **Mixed-effects modeling with R**. Madison: Springer, 2010. 131 p.

BENETT, C. G. S.; BUZETTI, S.; SILVA, K. S.; TEIXEIRA FILHO, M. C. M.; GARCIA, C. M. DE P.; MAESTRELO, P. R. Produtividade e desenvolvimento da cana-planta e soca em função de doses e fontes de manganês. **Revista Brasileira de Ciência do Solo**, v. 35, n. 5, p. 1661-1667, 2011.

BOWERMAN, A. F. A demonstration of the enviromics approach to integrating environmental 'big data'problems. **New Phytologist**, 2024.

BRAGA JUNIOR; R. L. C.; LANDELL, M. G. A.; XAVIER, M. A.; KANTHACK, R. A. D.; SILVA, D. N.; BIDÓIA, M. A. P. et al. **Censo varietal IAC de cana-de-açúcar no Brasil** - safra 2023/24. Campinas: Instituto Agronômico de Campinas, 2024. 64 p.

BRIGHTON, H.; GIGERENZER, G. The bias bias. **Journal of Business Research**, v. 68, n. 8, p. 1772-1784, 2015.

BULL, T. The sugarcane plant. *In*: In: Hogarth, M.; Allsopp, P. (ed.). **Manual of Cane Growing**. Brisbane: Bureau of Sugar Experimental Stations, 2000. P. 71-88.

CALLISTER, A. N.; COSTA-NETO, G.; BRADSHAW, B. P.; ELMS, S.; CROSSA, J.; BRAWNER, J. T. Enviromic prediction enables the characterization and mapping of Eucalyptus globulus Labill breeding zones. **Tree Genetics & Genomes**, v. 20, n. 1, p. 3, 2024.

CHAVES, L. J. Interação de genótipos com ambientes. *In*: NASS, C. C.; VALOIS, A. C. C.; MELO, I. S.; VALADARES-INGLIS, M. C. (ed). **Recursos genéticos e melhoramento - Plantas**. Rondonópolis: Fundação MT, 2001. cap. 22, p. 673-714.

CHENU, K. Characterizing the crop environment–nature, significance and applications. **Crop physiology**, p. 321-348, 2015.

CLAYTON, W. D.; VORONTSOVA, M. S.; HARMAN, K. T.; WILLIAMSON, H. **GrassBase – The Online World Grass Flora**. 2023. Disponível em: http://www.kew.org/data/grasses-db/sppindex.htm#S. Acesso em: 27 abr. 2024.

COELHO, A. P.; DALRI, A. B.; FARIA, R. T. de; LANDELL, E. P. A.; PALARETTI, L. F. Perfilhamento da cana-de-açúcar irrigada e plantada por mudas pré-brotadas: um novo conceito. **Acta Iguazu**, v. 7, n. 4, p. 71–84, 2019.

CONAB - Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira de cana-de-açúcar**. Brasília: Conab, 2015. 29p. Disponível em: https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar/item/download/1199_6d17e4217fb4a7eb35f603c18b0b3e06. Acesso em: 30 abr. 2023.

CONAB - Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira de cana-de-açúcar**. Brasília: Conab, 2018. 72p. Disponível em:

https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar/item/download/22956_506e8f00170422c62a452d3e319a6d6f. Acesso em: 30 abr. 2023.

CONAB - Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira de cana-de-açúcar**. Brasília: Conab, 2019. 58p. Disponível em: https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar/item/download/45225_ae2bfd50a119f157a4253d90d8364666. Acesso em: 30 abr. 2023.

CONAB - Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira de cana-de-açúcar**. Brasília: Conab, 2022. 59p. Disponível em: https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar/item/download/45236_22907ba0da29e3364b66c913e9557ec9. Acesso em: 30 abr. 2023.

CONAB - Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira de cana-de-açúcar**. Brasília: Conab, 2024. 52p. Disponível em: https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar/item/download/52732_1ba60aad418c90b0a86daf409a3703a5. Acesso em: 09 jul. 2024.

COOPER, M.; POWELL, O.; GHO, C.; TANG, T.; MESSINA, C. Extending the breeder's equation to take aim at the target population of environments. **Frontiers in Plant Science**, v. 14, p. 1129591, 2023.

COSTA-NETO, CRESPO-HERRERA, L.; FRADGLEY, N.; GARDNER, K.; BENTLEY, A. R.; DREISIGACKER, S. et al. Envirome-wide associations enhance multi-year genome-based prediction of historical wheat breeding data. **G3**, v. 13, n. 2, p. 1-22, 2022.

COSTA-NETO, G. M. F.; DUARTE, J. B.; CASTRO, A. P.; HEINEMANN, A. B. **Uso de informações ambientais na modelagem e interpretação da interação genótipo x ambiente: revisão bibliográfica**. Santo Antônio de Goiás: Embrapa Arroz e Feijão, 2020a. 46 p.

COSTA-NETO, G.; CROSSA, J.; FRITSCHE-NETO, R. Enviromic assembly increases accuracy and reduces costs of the genomic prediction for yield plasticity in maize. **Frontiers in Plant Science**, v. 12, p. 717552, 2021.

COSTA-NETO, G.; FRITSCHE-NETO, R.; CROSSA, J. Nonlinear kernels, dominance, and envirotyping data increase the accuracy of genome-based prediction in multi-environment trials. **Heredity**, v. 126, n. 1, p. 92-106, 2020b.

CRONQUIST, A.; TAKHTADZHIAN, A. L. **An integrated system of classification of flowering plants**. Nova Iorque: Columbia University Press, 1981. 1262 p.

CRUZ, C. D.; TORRES, R. A. A.; VENCOVSKY, R. An alternative approach to the stability analysis proposed by Silva and Barreto. **Revista Brasileira de Genética**, v. 12, n.3, p. 567-580, 1989.

CRUZ, D. D. M.; HEINEMANN, A. B.; MARCATTI, G. E.; RESENDE, R. T. Defining Target Population of Environments to Enviromics Studies Using R-based GIS Tools. **bioRxiv**, p. e50822519, 2025.

CULLIS, B. R.; SMITH, A. B.; COOMBES, N. E. On the design of early generation variety trials with correlated data. **Journal of agricultural, biological, and environmental statistics**, v. 11, p. 381-393, 2006.

CUNHA, N. G.; PASQUALETTO, A. Análise da expansão da cana-de-açúcar em Goiás. **Informe Gepec**, v. 26, n. 3, p. 120-153, 2022.

CURSI, D. E.; CASTILLO, R. O.; TARUMOTO, Y.; UMEDA, M.; TIPPAYAWAT, A.; PONRAGDEE, W. et al. Origin, Genetic Diversity, Conservation, and Traditional and Molecular Breeding Approaches in Sugarcane. *In*: PRIYADARSHAN, P.; JAIN, S. M. (ed.) **Cash Crops**. Cham: Springer, 2021. cap. 4. p. 83-116.

CURSI, D. E.; HOFFMANN; H. P.; BARBOSA, G. V. S.; BRESSIANI, J. A.; GAZAFFI, R.; CHAPOLA, R. G. et al. History and current status of sugarcane breeding, germplasm development and molecular genetics in Brazil. **Sugar Tech**, v. 24, n. 1, p. 112-133, 2022.

DAROS, E.; OLIVEIRA; R. A.; BARBOSA; G. V. S. (org.). **45 anos de variedades RB de cana-de-açúcar**: 25 anos de Ridesa. 1. ed. Curitiba: Graciosa, 2015, 156 p.

DEGANI, P. C.; LEITNER C. P. S.; BAGGENTOSS, S.; TOKOMIAN A. L. V; ALVES FILHO, A. G. Inovação em cultivares: panorama das proteções e registros da soja no Brasil. **Revista de Gestão e Tecnologia**, n. 11, p. 1-16, 2021.

DONZELLI, J. L.; BERTOLANI, F. C.; TROMBETA, N. C. Sugarcane Cultivation: Soil Mapping, Environmental Effects, and New Sugarcane Varieties. In: CHANDEL, A. K.; SILVEIRA, M. H. L. (ed.). **Advances in Sugacane Biorefinery**. Amsterdam: Elsevier, 2015. Cap. 1, p. 1-15.

DUARTE, J. B.; VENCOVSKY, R. **Interação genótipos x ambientes: uma introdução à análise AMMI.** Ribeirão Preto: Sociedade Brasileira de Genética, 1999. 60 p.

EBERHART, S. A.; RUSSEL, W. A. Stability parameters for comparing varieties. **Crop Science**, v. 6, n. 1, p. 36-40, 1966.

ELIAS, A. A.; ROBBINS, K. R.; DOERGE, R.W.; TUINSTRA, M. R. Half a century of studying genotype×environment interactions in plant breeding experiments. **Crop Science**, v. 56, n. 5, p. 2090-2105, 2016.
FAO. **World Food and Agriculture – Statistical Yearbook 2021**. Rome: FAO, 2021. 366p.

FAO. **World Food and Agriculture – Statistical Yearbook 2022**. Rome: FAO, 2022. 378p.

FAO. **World Food and Agriculture – Statistical Yearbook 2022**. Rome: FAO, 2023. 367p.

FRADGLEY, N.; GARDNER, K. A.; BENTLEY, A. R.; HOWELL, P.; MACKAY, I. J.; SCOTT, M. F.; MOTT, R.; COCKRAM, J. Multi-trait ensemble genomic prediction and simulations of recurrent selection highlight importance of complex trait genetic architecture for long-term genetic gains in wheat. in silico **Plants**, v. 5, n. 1, p. 1-20, 2023.

GAUCH, H. G. **Statistical analysis of regional yield trials**: AMMI analysis of factorial designs. Amsterdam: Elsevier, 1992. 278 p.

GAZAFFI, R.; CURSI, D. E.; CHAPOLA, R. G.; SANTOS, J. M.; FERNANDES JÚNIOR, A. R.; CARNEIRO, S. M. et al. RB varieties: A major contribution to the sugarcane industry in Brazil. *In:* Proceedings of the international society of sugar cane technologists, 29., 2016, Chiang Mai. **Anais eletrônicos** [...]. Chiang Mai: ISSCT, 2016. Disponível em: https://issct.org/resources/proceedings/. Acesso em: 02 maio 2023.

GELARO, R.; MCCARTY, W.; SUÁREZ, M. J.; TODLING, R.; MOLOD, A.; TAKACS, L.et al. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). **Journal of climate**, v. 30, n. 14, p. 5419-5454, 2017.

GRANDIS, A.; FORTIRER, J. D. S.; PAGLIUSO, D.; BUCKERIDGE, M. S. Scientific research on bioethanol in Brazil: history and prospects for sustainable biofuel. **Sustainability**, v. 16, n. 10, p. 4167, 2024.

GREEN, K. C.; ARMSTRONG, J. S. Simple versus complex forecasting: The evidence. **Journal of Business Research**, v. 68, n. 8, p. 1678-1685, 2015.

HEINEMANN, A. B.; COSTA-NETO, G.; FRITSCHE-NETO, R.; C. D.; DA MATTA H. D.; FERNANDES, I. K. Enviromic prediction is useful to define the limits of climate adaptation: a case study of common bean in Brazil. **Field Crops Research**, v. 286, p. 108628, 2022.

HOERL, A. E.; KENNARD, R. W. Ridge regression: Biased estimation for nonorthogonal problems. **Technometrics**, v. 12, n. 1, p. 55-67, 1970.

IAC. INSTITUTO AGRONÔMICO DE CAMPINAS. **IAC desenvolve régua que reúne informações para orientar o plantio de cana-de-açúcar**. Instituto Agronômico de Campinas, 2016. Disponível em: <https://www.iac.sp.gov.br/publicacoes/publicacoes/infocana_12.pdf>. Acesso em: 09 set. 2024.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Cidades@: Pesquisas Produção agrícola**. 2023. Disponível em: https://cidades.ibge.gov.br/brasil/go/pesquisa/14/10193?tipo=grafico&indicador=10248. Acesso em: 10 set. 2023.

JARQUÍN, D.; CROSSA, J.; LACAZE, X.; CHEYRON, P. D.; DAUCOURT, J.; LORGEOU, J. et al. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. **Theoretical and applied genetics**, v. 127, p. 595-607, 2014.

KHAN, M. H. U.; WANG, S.; WANG, J.; AHMAR, S.; SAEED, S.; KHAN, S. U. et al. Applications of artificial intelligence in climate-resilient smart-crop breeding. **International Journal of Molecular Sciences**, v. 23, n. 19, p. 11156, 2022.

KRAUSE, M. D.; DIAS, K. O. G.; SINGH, A. K.; BEAVIS, W. D. Using soybean historical field trial data to study genotype by environment variation and identify mega-environments with the integration of genetic and non-genetic factors. **Agronomy Journal**, v. 117, n. 1, p. e70023, 2025.

KUMAR, A.; BOHRA, A.; MIR, R. R.; SHARMA, R.; TIWARI, A., KHAN, M. A.; VARSHNEY, R. K. Next generation breeding in pulses: Present status and future directions. **Crop Breeding and Applied Biotechnology**, v. 21, p. e394221S13, 2021.

LIN, C. S.; BINNS, M. R. A method of analyzing cultivar x location x year experiments: a new stability parameter. **Theoretical and applied genetics**, v. 76, p. 425-430, 1988.

MAFFIA, M.; CONTE, L.; GRECO, M.; MAFFIA, M. C.; SIMEONE, P.; VERGARA, D. Omics Sciences: toward omics personalized medicine. **Journal of inter Disciplinary REsearch Applied to Medicine**, v. 1, n. 1, p. 7-20, 2017.

MARCATTI, G. E.; RESENDE, R. T.; RESENDE, M. D. V.; RIBEIRO, C. A. A.; DOS SANTOS, A. R.; DA CRUZ, J. P.; LEITE, H. G. GIS-based approach applied to optimizing recommendations of Eucalyptus genotypes. **Forest ecology and management**, v. 392, p. 144-153, 2017.

MARIOTTI, J. A.; OYARZABAL, E. S.; OSA, J. M.; BULACIO, A. N. R.; ALMADA, G. H. Análisis de estabilidad y adaptabilidad de genotipos de cana de azúcar, interacciones dentro de un localidad experimental. **Revista Agronómica del Noroeste Argentino**, San Miguel de Tucumán, v. 13, n. 14, p. 105-127, 1976.

MARTINS, A. S. **Aplicação de sistema de informações geográficas no estudo da interação de genótipos com ambientes**. 2004. 156 f. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Escola de Agronomia, Universidade Federal de Goiás, Goiânia, 2004.

MATSUOKA, S; STOLF, R. Sugarcane tillering and ratooning: key factors for a profitable cropping. Sugarcane: Production, cultivation and uses. *In*: GONÇALVES, J. F.; CORREIRA, K. D. (ed.) **Sugarcane**: production, cultivation and uses. New York: Nova Science Publishers, 2012. cap. 5, p. 137-157.

MAZZI, J. V. C.; DOS SANTOS, J. A.; MONTEBELLO, A. E. S. Agricultura de precisão no Setor Sucroalcooleiro do estado de São Paulo: análise produtiva, econômica e ambiental. **Revista de Gestão e Secretariado**, v. 14, n. 5, p. 8556-8571, 2023.

MEDINA, G. Desenvolvimento agroindustrial no Brasil. *In*: PEDROSO, M. T. M.; BRISOLA, M. V.; NAVARRO, Z (org.). **O Brasil rural**: novas interpretações. São Paulo: Editora BARAÚNA, 2024. cap. 6, p. 194-221.

MEHAREB, E. M.; OSMAN, M. A. M.; ATTIA, A. E.; BEKHEET, M. A. ABO ELENEN, F. M. Stability assessment for selection of elite sugarcane clones across multi-environment based on AMMI and GGE-biplot models. **Euphytica**, v. 218, n. 7, p. 95, 2022.

MEUWISSEN, T. H.E; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, n. 4, p. 1819-1829, 2001.

MIRAJKAR, S. J.; DEVARUMATH, R. M.; NIKAM, A. A.; SUSHIR, K.V.; BABU, H.; SUPRASANNA, P. Sugarcane (*Saccharum* spp.): Breeding and Genomics. *In*: AL-KHAYRI, J.; JAIN, S.; JOHNSON, D. (ed.). **Advances in Plant Breeding Strategies**: Industrial and Food Crops. Cham: Springer, 2019. cap. 11, p. 363-406.

MOLOD, A.; TAKACS, L.; SUAREZ, M.; BACMEISTER, J. Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2. **Geoscientific Model Development**, v. 8, n. 5, p. 1339-1356, 2015.

MONTES, R. F.; BRESEGHELLO, F.; DUARTE, J. B. Modelling of genotype by environment interaction to improve the recommendation of sugarcane cultivars for the state of Goiás, Brazil. **Pesquisa Agropecuária Brasileira**, v. 56, p. 1-10, 2021.

MORAIS, L. K.; CURSI D. E.; SANTOS, J. M.; SAMPAIO, M.; CÂMARA, T. M. M.; SILVA, P. A. **Melhoramento Genético da Cana-de-açúcar**. Aracajú: Embrapa Tabuleiros Costeiros, 2015. 40 p.

NUNES, G. H. S.; ANDRADE NETO, R. C.; COSTA FILHO, J. H. MELO, S. B. Influência de covariáveis ambientais sobre a interação genótipos x ambientes em meloeiro. **Revista Brasileira de Fruticultura**, v. 33, n.4, p. 1194-1199, 2011.

NUNES, J. A. R.; RAMALHO, M. A. P.; ABREU, A. F. Graphical method in studies of adaptability and stability of cultivars. **Annual report-bean improvement cooperative**, v. 48, p. 182-183, 2005.

OLIVEIRA, A. R.; BRAGA, M. B. **Florescimento e acamamento de cultivares de cana-de-açúcar submetidas a diferentes lâminas de irrigação**. Petrolina: Embrapa Semiárido, 2011. 23 p.

OLIVEIRA, R. A.; HOFFMANN, H. P.; BARBOSA, G. V. S. (org.) **Liberação nacional de variedades RB de cana-de-açúcar**. Curitiba: UFPR, 2021. 79 p.

OLIVOTO, T.; LÚCIO, A. D. C. SILVA, J. A. G.; MARCHIORO, V. S.; SOUZA, V. Q.; JOST, E. Mean performance and stability in multi-environment trials I: combining features of AMMI and BLUP techniques. **Agronomy Journal**, v. 111, n. 6, p. 2949-2960, 2019.

OTIENO, O. V.; OWUOR, O. N. Multivariate genotype and genotype by environment interaction biplot analysis of sugarcane breeding data using R. **International Journal of Statistical Distributions and Applications**, v. 5, p. 22, 2019.

PÉREZ, P.; DE LOS CAMPOS, G. BGLR: a statistical package for whole genome regression and prediction. **Genetics**, v. 198, n. 2, p. 483-495, 2014.

RAE, A. L.; MARTINELLI, A. P.; DORNELAS, M. C. Sugarcane: Physiology, Biochemistry, and Functional Biology. In: MOORE, P. H.; BOTHA, FREDERIK C. (ed.). **Sugarcane: physiology, biochemistry and functional biology**. Nova Jersey: John Wiley & Sons, 2014. cap. 2, p. 19-34.

RAMALHO, M. A. P.; ABREU, A. F. B.; Santos, J. B.; NUNES, J. A. R. **Aplicações da Genética Quantitativa no Melhoramento de Plantas Autógamas**. Lavras: Universidade Federal de Lavras, 2012. 522 p.

RAMBURAN, S.; ZHOU, M.; LABUSCHAGNE, M. Interpretation of genotype× environment interactions of sugarcane: Identifying significant environmental factors. **Field Crops Research**, v. 124, n. 3, p. 392-399, 2011.

RESENDE, R. T.; BRONDANI, C. **Melhoramento de precisão**. Brasília: Embrapa, 2023. 326 p.

RESENDE, R. T.; CHENU, K.; RASMUSSEN, S. K.; HEINEMANN, A. B.; FRITSCHE-NETO, R. Enviromics in plant breeding. **Frontiers in Plant Science**, v. 13, p. 935380, 2022.

RESENDE, R. T.; HICKEY, L.; AMARAL, C. H.; PEIXOTO, L. L.; MARCATTI, G. E.; XU, Y. Satellite-enabled enviromics to enhance crop improvement. **Molecular Plant**, v. 17, n. 6, p. 848-866, 2024a.

RESENDE, R. T.; PIEPHO, H. P.; ROSA, G. J.; SILVA-JUNIOR, O. B.; SILVA, F. F.; DE RESENDE, M. D. V.; GRATTAPAGLIA, D. Enviromics in breeding: applications and perspectives on envirotypic-assisted selection. **Theoretical and Applied Genetics**, v. 134, p. 95-112, 2021.

RESENDE, R. T.; XAVIER, A.; SILVA, P. I. T.; RESENDE, M. P. M.; JARQUIN, D.; MARCATTI, G. E. GIS-based G× E modeling of maize hybrids through enviromic markers engineering. **New Phytologist**, 2024b.

RIDESA. REDE INTERUNIVERSITÁRIA PARA O DESENVOLVIMENTO DO SETOR SUCROENERGÉTICO. **Melhoramento Genético da Cana-de-açúcar**. Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético. Disponível em: < https://www.ridesa.com.br/melhoramento>. Acesso em: 15 ago. 2024.

ROSSI, R. M. FERNANDES, F. B. Análise estratégica da evolução dos custos de produção da cultura da cana-de-açúcar em Goiás. **Custos e @gronegócio on-line**, v. 16, n. 3, p. 256-289, 2020.

SANGHERA, G. S.; JAMWAL, N. S., Perspective for Genetic Amelioration of Sugarcane towards Water Logged Conditions, **International Journal of Pure & Apply Bioscience**. v. 7, n. 3, p. 484-502, 2019.

SEGATO, S. V.; MATTIUZ, C. F. M.; MOZAMBANI, A. E. (org.). Aspectos fenológicos da cana-de-açúcar. *In:* SEGATO, S. V.; PINTO, A. de S.; JENDIROBA, E.; NÓBREGA, J. C. M. de. **Atualização em produção de cana-de-açúcar**. 1.ed. Piracicaba: Livroceres, 2008. p. 19-36.

SHEN, Y.; ZHOU, G.; LIANG, C.; TIAN, Z. Omics-based interdisciplinarity is accelerating plant breeding. **Current Opinion in Plant Biology**, v. 66, p. 102167, 2022.

SIDDIQ, E. A.; VEMIREDDY, L. R. Advances in Genetics and Breeding of rice: an overview. *In*: ALI, J; WANI, S. H. (ed.). **Rice Improvement**. 1. ed. Cham: Springer, 2021. p. 1-30.

SILVA, J. P. N.; SILVA, M. R. N. **Noções da cultura da cana-de-açúcar**. Inhumas: Universidade Federal de Santa Maria, 2012. 150 p.

SOM-ARD, J.; ATZBERGER, C.; IZQUIERDO-VERDIGUIER, E.; VUOLO, F.; IMMITZER, M. Remote sensing applications in sugarcane cultivation: A review. **Remote sensing**, v. 13, n. 20, p. 40-86, 2021.

SORENG, R. J.; PETERSON, P. M.; ROMASCHENKO, K.; DAVIDSE, G.; TEISHER, J. K.; CLARK, L. G. ET AL. A worldwide phylogenetic classification of the Poaceae (Gramineae) II: An update and a comparison of two 2015 classifications. **Journal of Systematics and Evolution**, v. 55, n. 4, p. 259 -290, 2017.

SOUZA, C. G. P.; RODRIGUES, M. S.; DA SILVA, K. A.; LIRA, A. L. F.; MOREIRA, F. B. R.; LIMA, A. M. N. Precision agriculture tools for liming management in mango orchards in the Brazilian semiarid region. **Comunicata Scientiae**, v. 13, p. e3630-e3630, 2022.

SOUZA, G. M.; BERGES H.; BOCS, S.; CASU, R.; D'HONT, A.; FERREIRA, J. E. The Sugarcane Genome Challenge: Strategies for Sequencing a Highly Complex Genome. **Tropical Plant Biology**. v. 4, n. 1, p. 145-156, 2011.

SPARKS, A. H. nasapower: a NASA POWER global meteorology, surface solar energy and climatology data client for R. **Journal of Open Source Software**, v. 3, n. 30, p. 1035, 2018.

SQUILASSI, M. G. **Interação de genótipos com ambientes**. Aracajú: Embrapa Tabuleiros Costeiros, 2003. 47 p.

TASSO JÚNIOR, L. C.; MARQUES, M. O.; SILVA NETO, H. F.; CAMILOTTI, F.; BERNARDI, J. H.; NOGUEIRA, T. A. R. Variação genotípica no florescimento, isoporização e características tecnológicas em seis cultivares de cana-de-açúcar. **Revista de Biologia e Ciências da Terra**, v. 9, n. 1, p. 12-18, 2009.

TOLER, J. E. **Patterns of genotypic performance over environmental arrays**. 1990. 153 f. Tese (Doctor of philosophy) Clemson University, Clemson, 1990.

TOPPA, E. V. B.; JADOSKI, C. J.; JULIANETTI, A.; HULSHOF, T.; ONO, E. O.; RODRIGUES, J. D. **Pesquisa Aplicada & Agrotecnologia**, v. 3, n. 3, p. 215-221, 2011.

USABORISUT, P. Progress in mechanization of sugarcane farms in Thailand. **Sugar Tech**. v. 20, n. 2, p. 116-121, 2018.

VAN EEUWIJK, F. A.; MALOSETTI, M.; YIN, X.; STRUIK, P. C.; STAM, P. Statistical models for genotype by environment data: from conventional ANOVA models to eco-physiological QTL models. **Australian Journal of Agricultural Research**, v. 56, n. 9, p. 883-894, 2005.

VASQUEZ, S. I.; WELKER, C. A. D.; NEFA, V. G. S.; PEICHOTO, M. C. Morphometric analysis of the *Saccharum* complex (Poaceae, Andropogoneae). **Plant Systematics and Evolution,** v. 308, n. 10, 2022.

VISWANATHAN, R. Sustainable sugarcane cultivation in India through threats of red rot by varietal management. **Sugar Tech**, v. 23, n. 2, p. 239-253, 2021.

VITALE, P.; CROSSA, J.; VACCINO, P.; DE VITA, P. Defining the target population of environments for wheat (Triticum aestivum L.) breeding in Italy based on historical data. **Plant Breeding**, v. 143, n. 4, p. 518-533, 2024.

VOGEL, J. T.; LIU, W.; OLHOFT, P., CRAFTS-BRANDNER; S. J., PENNYCOOKE; J. C.; CHRISTIANSEN, N. Soybean yield formation physiology–a foundation for precision breeding based improvement. **Frontiers in plant science**, v. 12, p. 719706, 2021.

WESTHUES, C. C.; MAHONE, G. S.; DA SILVA, S.; THORWARTH, P.; SCHMIDT, M.; RICHTER, J. C. et al. Prediction of maize phenotypic traits with genomic and environmental predictors using gradient boosting frameworks. **Frontiers in plant science**, v. 12, p. 1-22, 2021.

WINANS, N. D.; FONSECA, J. M.; PERUMAL, R.; KLEIN, P. E.; KLEIN, R. R.; ROONEY, W. L. Envirotyping can increase genomic prediction accuracy of new environments in grain sorghum trials depending on mega-environment. **Crop Science**, 2024.

WRICKE, G. Die erfassung der wechselwirking zwischen genotyp und umwelt bei quantitativen eigenschaften. **Zeitschrift fur pflanzenzuchtung**, v. 53, n. 4, p. 266-343, 1965.

XU, F.; WANG Z.; LU, G.; ZENG, R.; QUE, Y. Sugarcane ratooning ability: Research status, shortcomings, and prospects. **Biology**, v. 10, n. 10, p. 1052-1065, 2021.

YAN, W. GGE biplot- A Windows application for graphical analysis of multienvironment trial data and other types of two-way data. **Agronomy Journal**, v. 93, n. 5, p. 1111-1118, 2001.

YAN, W.; KANG, M. S.; MA, B.; WOODS, S.; CORNELIUS, P. L. GGE biplot vs. AMMI analysis of genotype-by-environment data. **Crop science**, v. 47, n. 2, p. 643-653, 2007.

ZHANG, J.; ZHANG, X.; TANG, H.; ZHANG, Q.; HUA, X.; MA, X. et al. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. **Nature genetics**, v. 50, n. 11, p. 1565-1573, 2018.

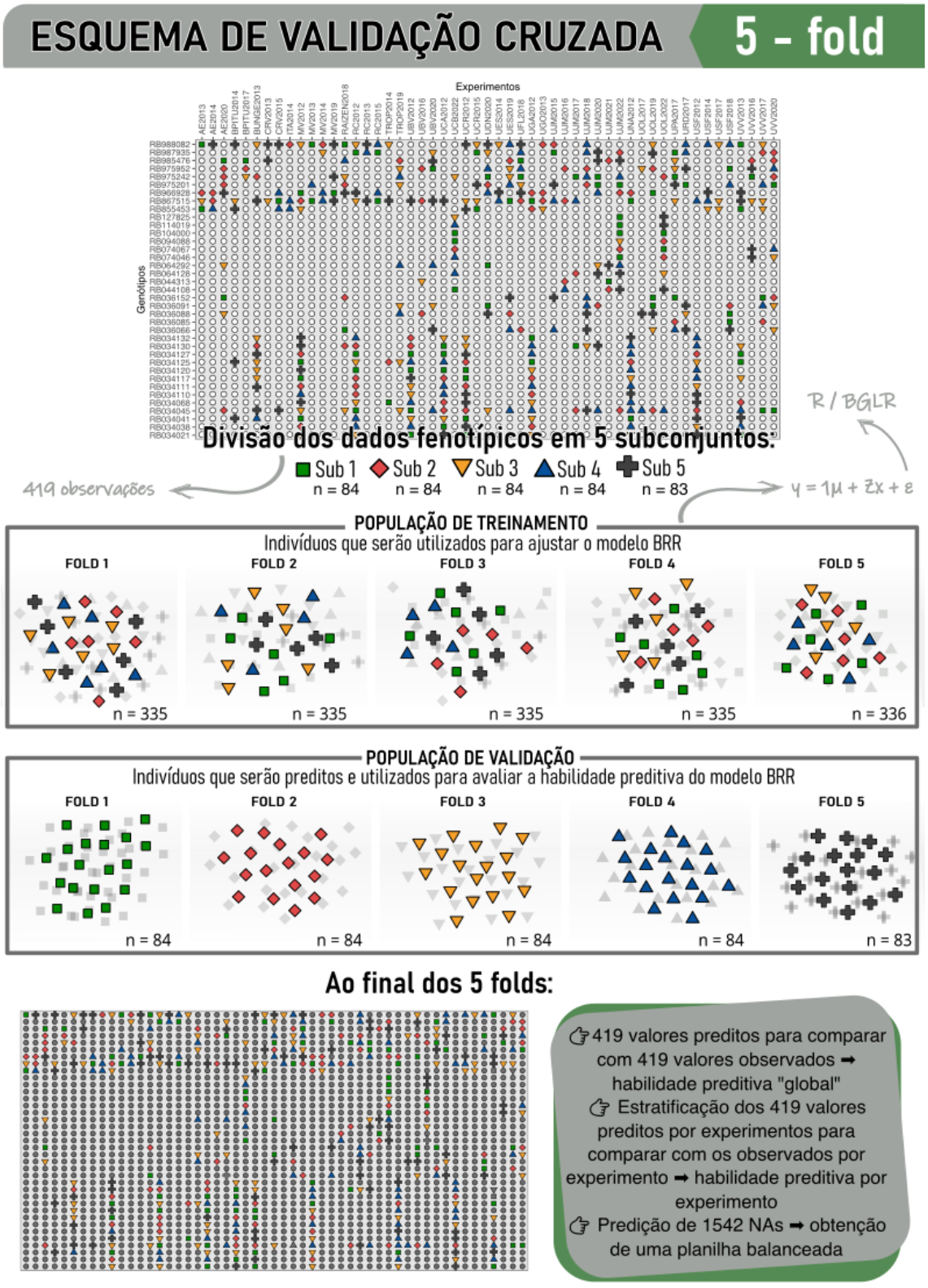
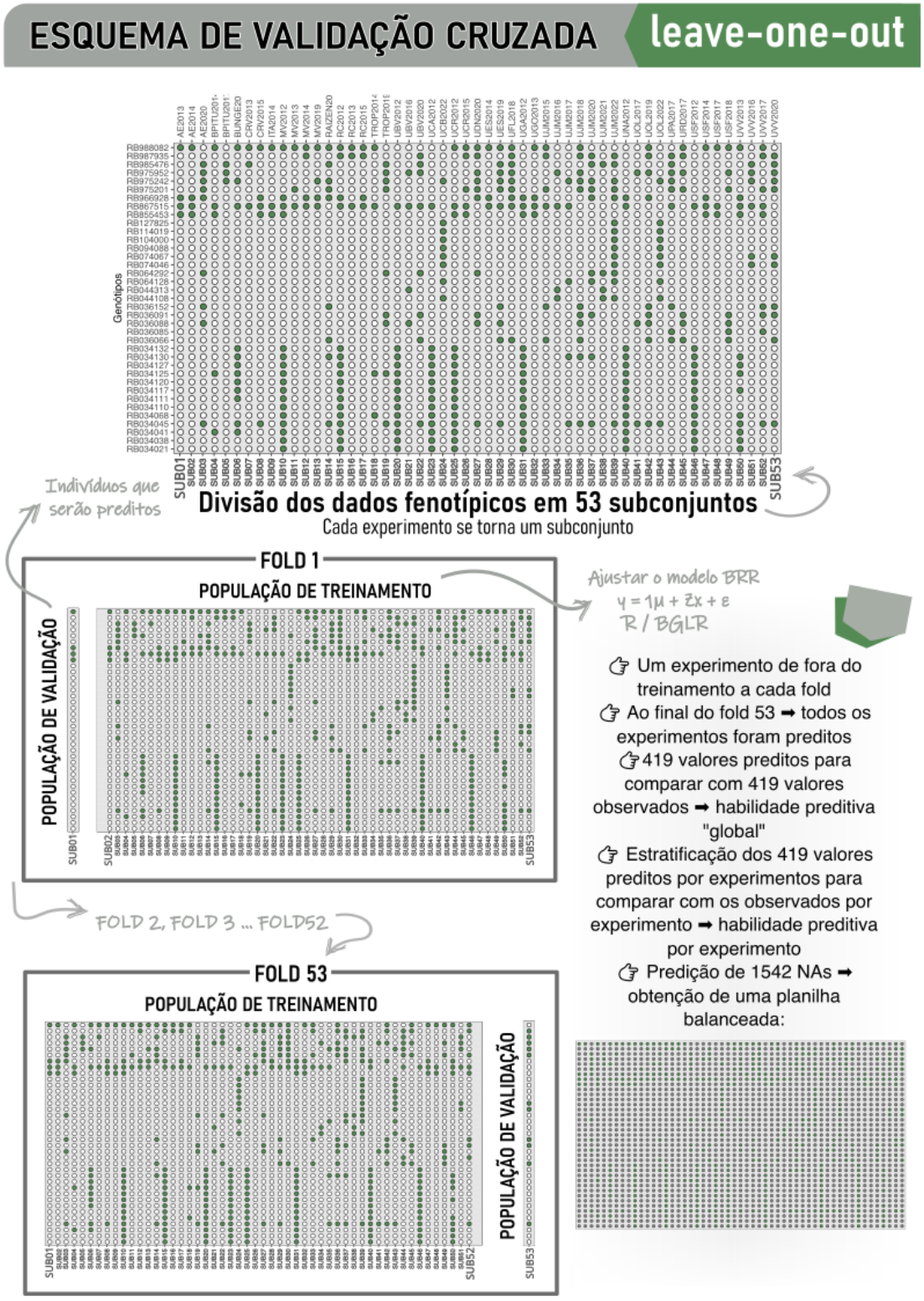
ZHAO, D.; LI, Y. Climate change and sugarcane production: potential impact and mitigation strategies. **International Journal of Agronomy**, v. 2015, n. 1, p. 1-10, 2015.

APÊNDICES

Apêndice A. Covariáveis geográficas e covariáveis ambientais obtidas através da plataforma NASAPOWER.

Sigla	Covariável (unidade de medida)
ALLSKY_SFC_LW_DWN	Irradiância de onda longa na superfície sob todas as condições de céu (W/m²)
ALLSKY_SFC_PAR_TOT	A radiação fotossinteticamente ativa total incidente na superfície sob todas as condições de céu (W/m²)
ALLSKY_SFC_SW_DWN	Irradiância descendente de onda curta na superfície sob todas as condições de céu (MJ/m²/dia)
ALLSKY_SFC_UV_INDEX	Índice UV na superfície sob todas as condições de céu (adimensional)
ALLSKY_SFC_UVA	Irradiância UVA na superfície sob todas as condições de céu (W/m²)
ALLSKY_SFC_UVB	Irradiância de UVB na superfície sob todas as condições de céu (W/m²)
ALT	Altitude (m)
CLRSKY_SFC_PAR_TOT	Radiação fotossinteticamente ativa total incidente na superfície sob condições de céu limpo (W/m²)
CLRSKY_SFC_SW_DWN	Irradiância descendente de onda curta na superfície sob condições de céu limpo (MJ/m²/dia)
GWETPROF	Umidade do perfil do solo (%)
GWETROOT	Umidade da zona radicular do solo, de 0 a 100 cm (%)
GWETTOP	Umidade superficial do solo, de 0 a 5 cm (%)
LAT	Latitude (Graus)
LONG	Longitude (Graus)
MM_BT	Precipitação acumulada na fase de brotação (mm)
MM_CRES	Precipitação acumulada no período de crescimento (mm)
MM_MAT	Precipitação acumulada na fase de maturação (mm)
PRECTOTCORR	Precipitação corrigida (mm/dia)
PS	Pressão na superfície (kPa)
QV2M	Umidade específica a 2 metros (g/kg)
RH2M	Umidade relativa a 2 metros (%)
T2M	Temperatura a 2 metros (°C)
T2M_MAX	Temperatura máxima a 2 metros (°C)
T2M_MIN	Temperatura mínima a 2 metros (°C)
T2M_RANGE	Intervalo de temperatura a 2 metros (°C)
T2MDEW	Ponto de orvalho/gelo a 2 metros (°C)
T2MWET	Temperatura do bulbo úmido a 2 metros (°C)
TOA_SW_DWN	Irradiância descendente de onda curta no topo da atmosfera (MJ/m²/dia)
TP_BT	Temperatura média na fase de brotação (°C)
TP_CRES	Temperatura média no período de crescimento (°C)
TP_MAT	Temperatura média na fase de maturação (°C)
TS	Temperatura da superfície da terra (°C)

Apêndice B. Esquema validação cruzada 5-*fold* deste trabalho



Apêndice C. Esquema validação cruzada *leave-one-out* deste trabalho