



LEONARDO OLIVEIRA SILVA DA COSTA

**INFLUÊNCIA AMBIENTAL NA PRODUTIVIDADE DE
CLONES DE *Eucalyptus* E MAPEAMENTO DAS SUAS
ADAPTABILIDADES**

**LAVRAS-MG
2022**

LEONARDO OLIVEIRA SILVA DA COSTA

**INFLUÊNCIA AMBIENTAL NA PRODUTIVIDADE DE CLONES DE *Eucalyptus* E
MAPEAMENTO DAS SUAS ADAPTABILIDADES**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Genética e Melhoramento de Plantas, para a obtenção do título de Mestre.

Prof. Dr. Evandro Novaes

Orientador

Prof.^a. Dr.^a. Flavia Maria Avelar Gonçalves

Coorientadora

**LAVRAS-MG
2022**

**Ficha catalográfica elaborada pelo Setor de Repositório
Insitucional da Biblioteca Universitária da UFLA**

Costa, Leonardo Oliveira Silva da.

Influência ambiental na produtividade de clones de *Eucalyptus* e mapeamento das suas adaptabilidades / Leonardo Oliveira Silva da Costa. - 2022.

82 p. : il.

Orientador(a): Evandro Novaes.

Coorientador(a): Flavia Maria Avelar Gonçalves.

Dissertação (mestrado acadêmico) - Universidade Federal de Lavras, 2022.

Bibliografia.

1. Ambientipagem. 2. Estabilidade. 3. Interação genótipos por ambientes. 4. Predição. I. Novaes, Evandro. II. Gonçalves, Flavia Maria Avelar. III. Título.

LEONARDO OLIVEIRA SILVA DA COSTA

**INFLUÊNCIA AMBIENTAL NA PRODUTIVIDADE DE CLONES DE *Eucalyptus* E
MAPEAMENTO DAS SUAS ADAPTABILIDADES**

**ENVIRONMENTAL INFLUENCE ON THE PRODUCTIVITY OF *Eucalyptus*
CLONES AND ADAPTABILITIES MAPPING**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Genética e Melhoramento de Plantas, para a obtenção do título de Mestre.

APROVADA em 15 de fevereiro de 2022

Dr^a. Aline Cristina Miranda Fernandes – Suzano S.A.

Dr. Germano Martins Ferreira Costa Neto - Cornell University

Prof. Dr. Evandro Novaes

Orientador

Prof.^a. Dr^a. Flavia Maria Avelar Gonçalves

Coorientadora

**LAVRAS-MG
2022**

AGRADECIMENTOS

Agradeço primeiramente à minha família por todo o suporte necessário para a conclusão dessa etapa durante um período tão difícil de pandemia. Aos meus irmãos Lucas e Gustavo, meu pai Denilson Freitas e à minha vó Francisca Oliveira por todo amor, carinho e confiança. Especialmente à minha grande amiga e mãe Josinea Oliveira, a quem eu possuo grande admiração por sua força e sabedoria.

Aos meus amigos de Lavras que nos meus breves momentos na cidade constituíram uma valiosa rede de apoio: Leonardo Costa, Luiz Guilherme, Camila Teixeira, Andreza Martins, Gleice Lima, Tiago Yukio, Bruna Natália e Lorena Queiroz. Aos meus amigos de Goiânia e Brasília que participaram indiretamente desse momento: Ana Carolina Alves, Maria Paula Ataídes, Alef Solon, Thaís Pereira, Marcela Pires, Larissa Sousa, Laíssa Matos, Leonardo Carlos Corvalan, Christian Pereira, Liza Gabriela, Julia Pereira, Paulo Sérgio, Mariane Brom, Lucas Vieira, Ana Beatriz e Maria Caroline.

A Universidade Federal de Lavras e ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas. À minha coorientadora Flavia Maria Avelar Gonçalves, a meu supervisor de estágio docência José Airton Rodrigues Nunes e todo corpo docente pela oportunidade de crescimento pessoal e profissional. Em especial, ao meu orientador Evandro Novaes por todos os conselhos, apoio e compreensão e paciência durante todo o curso de mestrado e parte da graduação. Ainda, à Carolina Boaventura que, juntamente com meu orientador, me recebeu em Lavras com muita hospitalidade.

Agradeço a banca composta por Aline Cristina Miranda Fernandes, Dr. José Airton Rodrigues Nunes, Dr. Germano Martins Ferreira Costa Neto e Dr. João Batista Duarte. Ao Prof. Alexandre Coelho da Universidade Federal de Goiás por gentilmente ceder os servidores de alta performance “Saccharum” e “Caryocar”, adquiridos pelo financiamento da Petrobras (P1409/ANP681C) e do consórcio MCT/CNPq/FNDCT/FAPs/MEC/CAPEs (031/2010), respectivamente. Ao Aurélio Mendes Aguiar, Aline Fernandes e Izabel Gava pela colaboração e *feedbacks* durante o trabalho e à empresa Suzano S.A. que forneceu os dados para execução desse trabalho. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

RESUMO

O caráter continental do Brasil contempla uma ampla gama de condições edafoclimáticas. Como a produção vegetal possui uma distribuição quantitativa, essa heterogeneidade ambiental representa dificuldades para o melhoramento de plantas devido à interação de genótipos com ambientes (GxA). Assim, o conhecimento detalhado da influência dos fatores ambientais sobre a expressão do fenótipo é importante para o melhoramento genético. Esse conhecimento pode indicar estratégias de incremento na capacidade preditiva dos efeitos da interação GxA e, com isso, melhorar a avaliação e eficiência da recomendação de cultivares. O objetivo do estudo foi quantificar o impacto da variabilidade ambiental na produtividade de eucalipto no Brasil e o mapeamento da adaptabilidade de clones comerciais com base em sistema de informação geográfica (SIG). Para isso, um conjunto de dados de 13.483 talhões de inventário florestal de seis clones comerciais foi fornecido pela Suzano S.A. e abrangeu as quatro principais unidades de plantio da empresa (Mato Grosso do Sul, São Paulo, Bahia e Maranhão). Dessa forma, o conjunto de dados de 13.483 talhões foi reduzido para 2.626 valores médios de produtividade. A modelagem dos efeitos das covariáveis geográficas e edafoclimáticas na produtividade dos clones foi realizada via regressão de quadrados mínimos parciais (PLS, do inglês *Partial Least Squares*). Os dados de solo foram obtidos a partir do banco de dados SoilGrids. Os dados climáticos de três bancos (INMET, NASA Power e WorldClim) foram comparados quanto às suas capacidades preditivas para cada clone pelo método *leave-one-out*. As métricas avaliadas foram: coeficiente de determinação (R^2); erro preditivo médio (RMSE, do inglês *Root Mean Square Error*); e índice de concordância. O melhor banco variou entre os clones quanto ao R^2 , porém o WorldClim possibilitou menores erros preditivos médios para todos os genótipos avaliados e, portanto, foi escolhido para as predições. A partir desse banco foram elaborados mapas de produtividade para cada clone com resolução de $\sim 5 \text{ km}^2$. A análise de similaridade ambiental, indicou que a unidade do Maranhão apresenta os ambientes mais distintos, resultando na menor proporção de predições adequadas. Nesse sentido, o clone com a maior proporção de dados de inventário na região (CLZ003) foi prejudicado e retirado da versão final do mapa do genótipo vencedor. Dentro de cada unidade florestal destacaram-se os: clone CLZ005 na Bahia (recomendado em 29,99% da área no estado) e São Paulo (57,35%), CLZ001 em Mato Grosso do Sul (50,18%) e CLZ004 no Maranhão (77,29%). As covariáveis que mais afetaram o desempenho dos clones foram a precipitação anual (BIO12), precipitação do mês mais seco (BIO14), precipitação do trimestre mais seco (BIO17), temperatura máxima do mês mais quente (BIO5) e temperatura média do trimestre mais úmido (BIO8). A metodologia permitiu a identificação dos principais fatores edafoclimáticos que reagem com a interação GxA de diferentes genótipos e representa um recurso potencial para avanço na recomendação de clones de *Eucalyptus* com o uso de SIG.

Palavras-chave: Ambientipagem. Estabilidade. Interação genótipos por ambientes. Predição.

ABSTRACT

The continental character of Brazil contains a wide range of edaphoclimatic conditions. As plant production has a quantitative distribution, environmental heterogeneity represents difficulties for plant breeding due to the interaction of genotypes by environments (GxE). Thus, the knowledge of the environmental influences on phenotype expression is important for genetic breeding. This knowledge can indicate strategies to increase the predictive capacity of the effects of the GxA interaction and, therefore, improve the evaluation and efficiency of cultivar recommendation. The objective of the study was to quantify the impact of the environmental variability in eucalyptus production in Brazil and map the adaptability of commercial clones based on geographic information system (GIS). A dataset of 13,483 forest inventory data points of six commercial clones was provided by Suzano S.A. and encompassed the four main plantation units of the company (Mato Grosso do Sul, São Paulo, Bahia e Maranhão). The 13,483 was reduced for 2,262 mean productivity values. Partial Least Square (PLS) regressions were used to model the productivity of the six Eucalyptus clones using the environmental covariables. Soil data were obtained from the SoilGrids database. Climatic data from three databases (INMET, NASA Power, and WorldClim) were compared by their predictive capabilities for each clone using the leave-one-out method. The metrics evaluated were coefficient of determination (R^2); mean predictive error (RMSE); and concordance index. WorldClim was considered the most suitable bank for three clones considering R^2 and all of them considering RMSE and therefore selected for predictions. Productivity maps were plotted for each clone with a spatial resolution of $\sim 5 \text{ km}^2$. The similarity analyses showed that Maranhão unit presented the most distinct environments and, therefore, the lowest proportion of adequate predictions. Consequently, the clone with the highest proportion in the region (CLZ003) was harmed and removed from the last version of the winning genotype map. Within each forest unit, the most recommended genotypes were clone CLZ005 in Bahia (29.99%) and São Paulo (57.35%), CLZ001 in Mato Grosso do Sul (50.18%) and CLZ004 in Maranhão (77.29%). The covariates that most affected the performance of the clones were annual rainfall (BIO12), rainfall of the driest month (BIO14), rainfall of the driest quarter (BIO17), maximum temperature of the hottest month (BIO5), and mean temperature of the wettest quarter (BIO8). The methodology allowed the identification of important edaphoclimatic factor evolved in GxE interaction of different genotypes and represents a resource for potential advance in the recommendation of cultivars.

Keywords: Environment typing. Genotype by environment interaction. Stability. Prediction.

LISTA DE ILUSTRAÇÕES

Figura 1 – Série histórica da produção de madeira em tora dos estados alvo do estudo.....	13
Figura 2 – Normais climatológicas (1970-2000) mapeadas da precipitação pluviométrica e temperatura no Brasil.....	15
Figura 3 – Distribuição geográfica das localidades consideradas para estudo através de quatro regiões no Brasil.	22
Figura 4 – Histograma e boxplots das distâncias (km) entre as estações e as florestas inventariadas para esse estudo.....	25
Figura 5 – Esquema da extração de covariáveis hipotéticas (COV1 e COV2) por ponto a partir de arquivos raster.....	26
Figura 6 – Caracterização da produtividade de eucalipto (IMA, $m^3 \cdot ha^{-1} \cdot ano^{-1}$) de seis clones (CLZ001, CLZ002, CLZ003, CLZ004, CLZ005, CLZ006) e geral em quatro unidades florestais (MA, BA, SP, MS).....	31
Figura 7 – Caracterização ambiental.	33
Figura 8 – Comparação dos coeficientes de determinação (R^2) da capacidade preditiva utilizando modelos ajustados a partir de diferentes combinações de bancos de dados, com e sem dados edáficos do SoilGrids.....	34
Figura 9 – Valores da razão de soma de quadrados dos valores estimados (RESS) pela soma de quadrados residual dos valores preditos (PRESS) para cada variável latente (Q^2).	36
Figura 10 – Gráfico de dispersão dos valores preditos e observados para cada clone avaliado no estudo.....	37
Figura 11 – Frequência das classes de valores preditos por unidade florestal.	38
Figura 12 – Círculos de correlação entre as covariáveis ambientais e a variável resposta (IMA) para cada um dos seis clones do estudo.	39
Figura 13 – Correlações par a par das cinco covariáveis mais influentes na produtividade dos clones (BIO12, BIO14, BIO17, BIO5 e BIO8) e elevação (ELEV) por clone. Na primeira linha, os valores de correlação geral e separado por clone. Na última linha, os gráficos de dispersão (pontos coloridos por clone).....	42
Figura 14 – Correlações par a par das cinco covariáveis mais influentes na produtividade dos clones (BIO12, BIO14, BIO17, BIO5 e BIO8) e elevação (ELEV) por unidade florestal. Na primeira linha, os valores de correlação geral e separado por unidade florestal. Na última linha, os gráficos de dispersão (pontos coloridos por unidade florestal).....	43
Figura 15 – Adaptabilidade mapeada	45
Figura 16 – Mapas dos clones vencedores a partir de valores de produtividades preditas com modelos PLS utilizando covariáveis ambientais.	46
Figura 17 – Mapas dos clones vencedores a partir de valores de produtividade predita excluindo o clone CLZ003. As produtividades foram preditas com as covariáveis ambientais utilizando o modelo PLS.	48

LISTA DE TABELAS

Tabela 1 – Número de pontos geográficos referentes aos valores médios de produtividade por clone e unidade florestal.	23
Tabela 2 – Covariáveis Bioclimáticas.	24
Tabela 3 – Métricas da capacidade preditiva dos modelos obtida com o método leave-one-out para cada conjunto de dados ambientais.....	35
Tabela 4 – Coeficientes de regressão normalizados das cinco covariáveis de maior influência na produtividade de cada um dos seis clones avaliados.	40
Tabela 5 – Clones vencedores e seus percentuais de área de adaptação produtiva superior aos demais clones nas regiões alvo.....	47
Tabela A.1 – Coeficientes de regressão via Quadrados Mínimos Parciais (PLS) para os seis clones avaliados.....	62

SUMARIO

1. INTRODUÇÃO	10
2. REVISÃO BIBLIOGRÁFICA	13
2.1. A eucaliptocultura nas localidades alvo do estudo	13
2.2. Efeitos do ambiente na produção florestal.....	14
2.2.1. Heterogeneidade ambiental no Brasil	14
2.2.2. Modelagem da produtividade utilizando covariáveis ambientais.....	16
2.2.3. Regressão via quadrados mínimos parciais (PLS) e suas aplicações	17
2.3. Interação genótipos x ambientes (GxA)	18
3. MATERIAL E MÉTODOS	22
3.1. Áreas de estudo e material vegetal	22
3.2. Bancos de dados edafoclimáticos	23
3.2.1. Dados do banco INMET	24
3.2.2. Dados do NASA Power	25
3.2.3. Dados do WorldClim	26
3.3. Avaliação da capacidade preditiva com os diferentes bancos de dados	27
3.4. Influência das covariáveis ambientais na produtividade dos clones de eucalipto	27
3.4.1. Modelagem da produtividade de madeira via regressão PLS.....	27
3.4.2. Elaboração de mapas temáticos sob valores preditos	29
3.5. Recursos computacionais.....	30
4. RESULTADOS	31
4.1. Caracterização dos dados de inventário e ambientais.....	31
4.2. Definição dos bancos de dados ambientais	33
4.3. Modelagem da produtividade de madeira via regressão PLS.....	36
4.4. Análises dos efeitos das covariáveis ambientais na produtividade dos clones.....	38
4.4.1. Mapeamento da produtividade predita dos clones.....	44
4.4.2. Mapas dos clones vencedores	46
5. DISCUSSÃO	49
6. CONCLUSÃO	54
REFERÊNCIAS	55
APÊNDICE A – Modelos Preditivos	62
APÊNDICE B – Script em linguagem R	63

1. INTRODUÇÃO

O setor florestal brasileiro é constituído por uma cadeia produtiva resiliente, que contribui de maneira crescente para o PIB brasileiro (IBÁ, 2022). Tal relevância e crescimento só foi possível devido às constantes melhorias tecnológicas relacionadas ao setor no país. Na eucaliptocultura, por exemplo, a produtividade média de madeira saltou de 10 para 36 $\text{m}^3.\text{ha}^{-1}.\text{ano}^{-1}$ no período de 50 anos (IBÁ, 2016, 2020), com algumas regiões atingindo os 60 $\text{m}^3.\text{ha}^{-1}.\text{ano}^{-1}$. Essa produtividade de 60 $\text{m}^3.\text{ha}^{-1}.\text{ano}^{-1}$ é praticamente o potencial produtivo dos eucaliptos (STAPE et al., 2010).

O desenvolvimento da eucaliptocultura brasileira está fortemente relacionado ao melhoramento genético (FERREIRA, 2016). A cultura tem um robusto histórico de melhoramento desde a sua introdução no país, destacando-se as décadas de 1960 a 1980 (CASTRO et al., 2016). Foi nessa época que a necessidade de combate ao cancro (causado pelo fungo *Chrysosporthe cubensis*) impulsionaram importantes avanços tecnológicos, como a hibridação interespecífica e a clonagem via estaquia. Essa estratégia de clonagem de híbridos se tornou a base do melhoramento genético e da maior parte da produção de eucalipto no país (ASSIS; MAFIA, 2007; CASTRO et al., 2016). Além disso, houve um interesse crescente no estudo da interação genótipos por ambientes (GxA), com o uso de ferramentas biométricas para o entendimento do fenômeno (MORI; LELLO; KAGEYAMA, 1986; MORI; KAGEYAMA; FERREIRA, 1988).

A compreensão da magnitude da influência do ambiente sobre a produção florestal aliada a estratégias para lidar com os efeitos de interação GxA é essencial no melhoramento de plantas. Dentre as estratégias mais empregadas estão a seleção de genótipos mais estáveis em uma ampla gama de ambientes ou a seleção de genótipos mais adaptados a ambientes específicos. Além disso, a estratificação ambiental por meio da definição de grupos ambientais homogêneos, i.e. com menor interação GxA, também é uma estratégia importante para a recomendação de cultivares (CRUZ; REGAZZI; CARNEIRO, 2012; SANTOS et al., 2015; CASTRO et al., 2018; PUPIN et al., 2018; DA SILVA et al., 2019). Em geral, essas estratégias dependem de um amplo acervo de ferramentas e análises biométricas.

Apesar dos avanços obtidos com os estudos da interação GxA, incluindo análises de estabilidade e adaptabilidade, a predição da produtividade/adaptabilidade de cultivares em diferentes locais ainda é um enorme desafio. Em geral, os clones devem ser testados em um maior número de locais representativos das regiões alvo do cultivo. Entretanto, mesmo com extensas redes de ensaios, permanecem as dificuldades na predição da produtividade futura

das cultivares, bem como na mitigação dos riscos eventuais de grandes quedas na produtividade, em face de estresses (bióticos e abióticos) e mudanças climáticas. Nesse sentido, é fundamental que pesquisas continuem sendo desenvolvidas para melhorar a capacidade preditiva dos modelos de interação GxA e, com isso, a recomendação de cultivares.

A modelagem da interação GxA por meio de informações ambientais usadas como covariáveis (geográficas, edáficas e climáticas) vem trazendo novas possibilidades de avanços (MALOSETTI et al., 2016; MARCATTI et al., 2017; COSTA-NETO et al., 2020). O uso de covariáveis ambientais pode iluminar os mecanismos responsáveis pela interação GxA e melhorar a capacidade preditiva de modelos biométricos, especialmente em novos ambientes alvo (PIEPHO; DENIS; VAN EEUWIJK, 1998; MALOSETTI et al., 2016). Como nesses novos ambientes ainda não houveram plantios ou experimentos, o rendimento de genótipos só pode ser estimado por meio de suas características edafoclimáticas com a utilização das covariáveis ambientais (PIEPHO; DENIS; VAN EEUWIJK, 1998).

Nesse sentido, uma abordagem promissora consiste na combinação das ferramentas biométricas tradicionais e da caracterização ambiental combinados com sistema de informações geográficas (SIG) (XU, 2016). O ajuste de modelos em que a produtividade de um ou mais genótipos foi avaliada em vários locais possibilita a predição fenotípica em áreas em que um material genético nunca foi testado (COOPER et al., 2014). Com isso, é possível realizar o mapeamento da adaptabilidade desse genótipo pixel a pixel, na mesma resolução em que as covariáveis ambientais foram avaliadas ou estimadas.

Essa metodologia vem sendo aplicada em culturas como cana-de-açúcar (MONTES; BRESEGHELLO; DUARTE, 2021), trigo (ANNICCHIARICO; BELLAH; CHIARI, 2006), arroz (COSTA-NETO et al., 2020) e eucalipto (MARCATTI et al., 2017) para a determinação da adaptabilidade e estabilidade de genótipos em diferentes regiões alvo. Marcatti et al. (2017) utilizaram dados de produtividade, em florestas plantadas de eucalipto, para o ajuste de um modelo misto (REML/BLUP) associado a um modelo logístico em função de um índice ambiental. De forma análoga aos modelos lineares, o coeficiente de regressão associado ao índice ambiental foi utilizado para inferir a sensibilidade do genótipo ao ambiente. Entretanto, apesar de promissora, não há muitos relatos do uso de técnicas baseadas em SIG para a cultura do eucalipto e quanto da variação fenotípica pode ser dissecada por informações ambientais. Assim, a utilização de covariáveis ambientais para modelagem da interação GxA e mapeamento da adaptabilidade fenotípica via SIG representa uma abordagem analítica interessante, que pode gerar ganhos de eficiência no melhoramento genético para

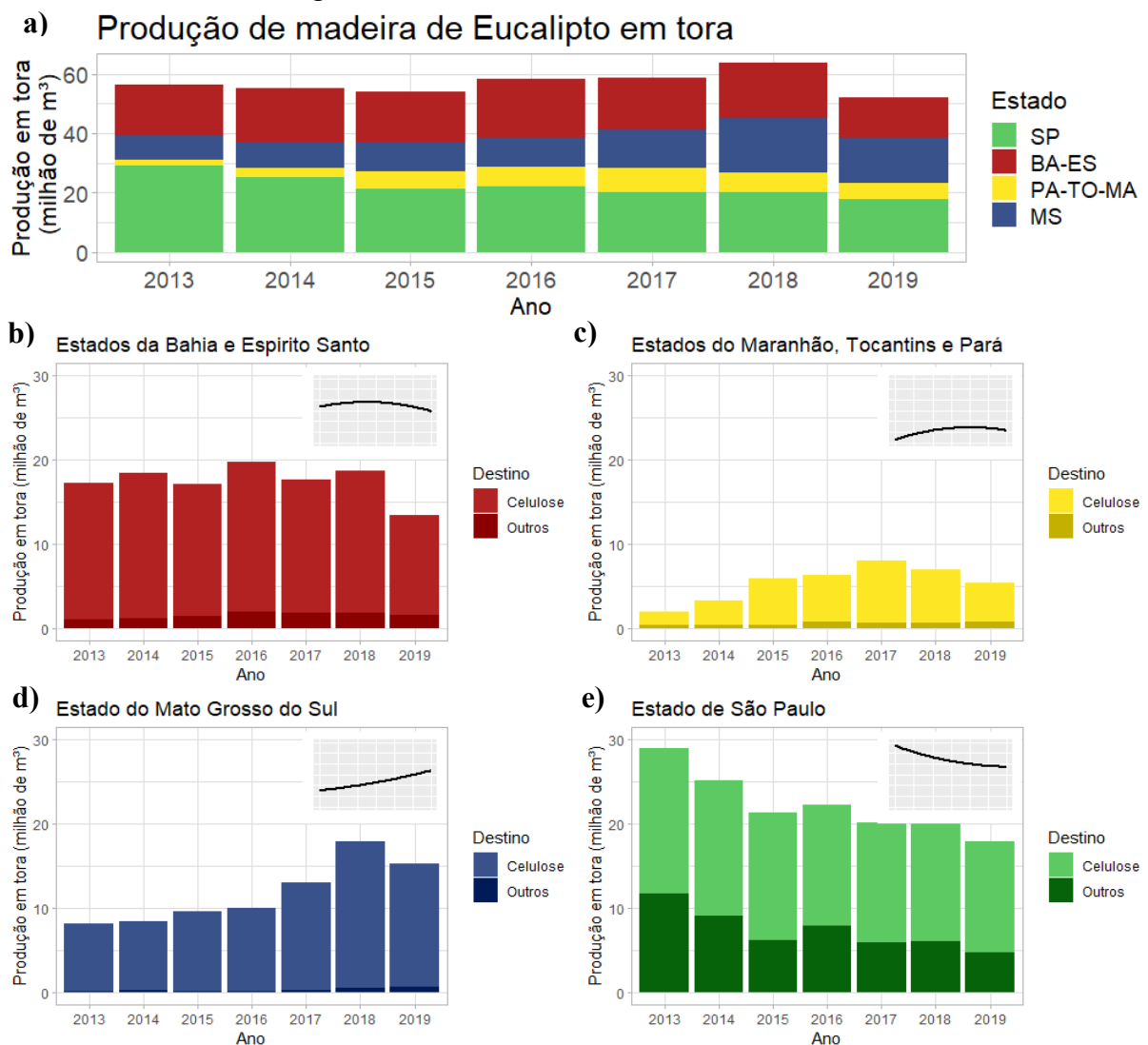
recomendação de cultivares. Dessa forma, o objetivo desse estudo foi quantificar o impacto da variabilidade ambiental na produtividade de eucalipto no Brasil e o mapeamento da adaptabilidade de clones comerciais com base em sistema de informação geográfica (SIG).

2. REVISÃO BIBLIOGRÁFICA

2.1. A eucaliptocultura nas localidades alvo do estudo

Mesmo devido às potenciais dificuldades geradas pelo gradiente ambiental, a eucaliptocultura está presente em todas regiões do Brasil (IBÁ, 2022). Os estados da Bahia, Espírito Santo, Mato Grosso do Sul, Minas Gerais, São Paulo, Maranhão, Tocantins e Pará, alvos desse estudo, estão entres as unidades da federação com maior produção e área plantada. A Figura 1 ilustra a série história da produção de madeira de eucalipto em toras (m³) obtida a partir do banco de dados SIDRA/IBGE (IBGE, 2021).

Figura 1 – Série histórica da produção de madeira em tora dos estados alvo do estudo. **a)** Série histórica da produção total, **b-e)** Série histórica de destinação da produção por estado. No canto superior direito de cada gráfico, as linhas de tendência do volume extraído ao longo da serie historica.



Fonte: IBGE (2021)

Pode-se notar que a produção entre os sete anos avaliados não apresentou grandes oscilações e que o estado de São Paulo se configura como o maior produtor da série histórica, enquanto o Maranhão apresenta os menores volumes de madeira extraídos (Figura 1.a). No setor florestal, esses estados são caracterizados pela forte influência de indústrias de celulose. De fato, o principal destino da madeira em tora de cada estado é para a produção de celulose (Figura 1.b-e).

A partir da comparação da produção por estado, é possível notar que há uma tendência de aumento na produção no Mato Grosso do Sul. Esse aumento possibilitou que o estado apresentasse maior produção de toras para celulose que o estado de São Paulo, desde o ano de 2018. Os estados do Maranhão, Tocantins e Pará apresentam as menores extrações de madeira. Esses estados são recentes quanto à expansão do setor florestal para produção de celulose e, além da ausência de progênies/clones adaptados, possuem condições edafoclimáticas relacionadas aos estresses hídrico e térmico. Além disso, em conjunto com os estados do Espírito Santo e Bahia, a ocorrência do distúrbio fisiológico do eucalipto é um dos grandes desafios para a região (CÂMARA et al., 2018).

2.2. Efeitos do ambiente na produção florestal

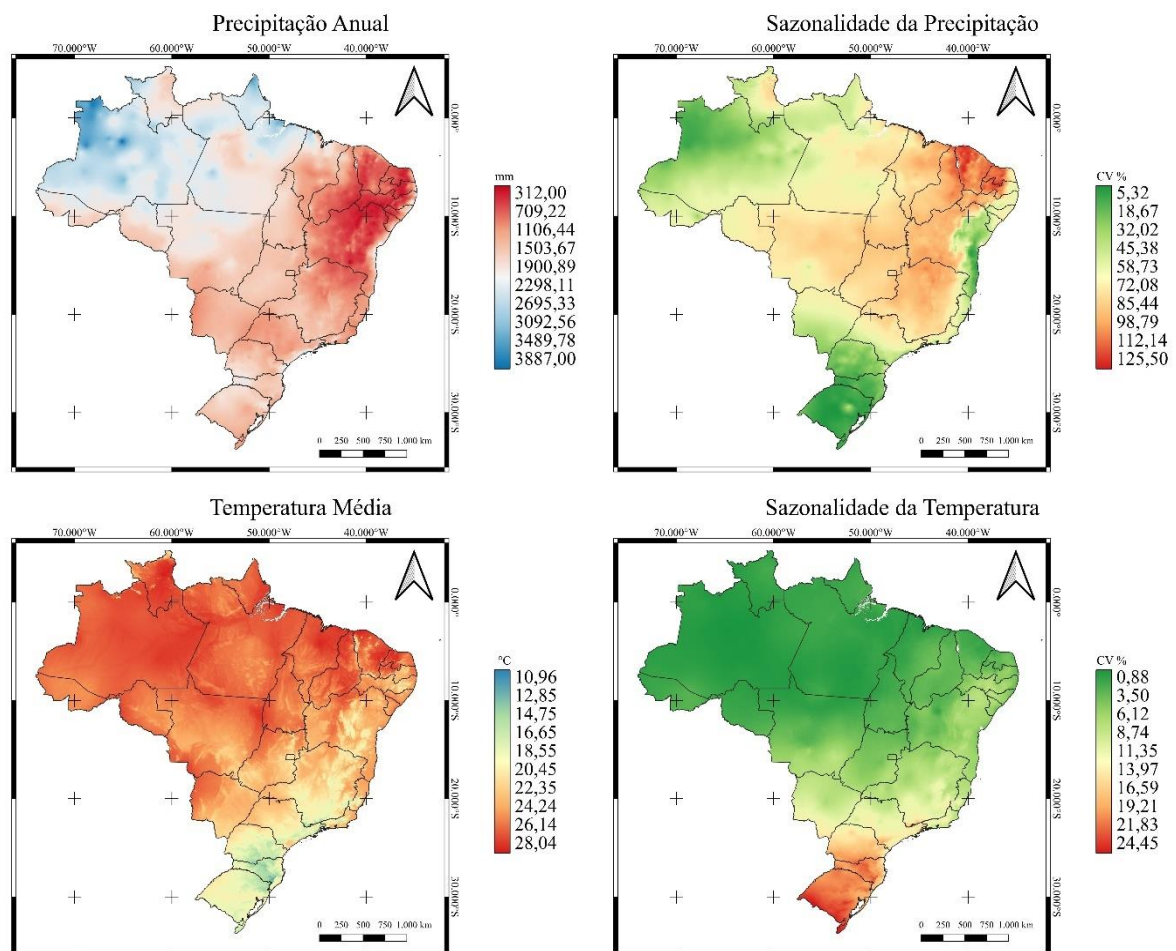
2.2.1. Heterogeneidade ambiental no Brasil

O caráter continental do Brasil contempla uma ampla gama de fatores edafoclimáticos que lhe confere diversas possibilidades de usos do solo. Nesse contexto, com o avanço de novas tecnologias e perspectivas produtivas, áreas consideradas como inviáveis para o agronegócio, décadas atrás, se consolidam gradativamente no cenário produtivo agrícola. Um exemplo desse fenômeno é a região definida como MATOPIBA, compreendida por municípios dos estados do Maranhão, Tocantins, Piauí e Bahia (BRASIL, 2015).

Essa região, mesmo concentrada em uma pequena parcela de Cerrado nas regiões Norte e Nordeste, possui uma grande variação edafoclimática não usual comparada às regiões de cultivo tradicionais (Centro-Oeste, Sul e Sudeste), demandando novas pesquisas e o uso intensivo de tecnologia (MINGOTI et al., 2014). Esse cenário de heterogeneidade se repete em escala nacional (Figura 2) e incentivou esforços para a compreensão da influência de fatores edafoclimáticos em diferentes culturas e cultivares, para fins de otimização (ARAUJO et al., 2019; ELLI et al., 2020; ROCHA et al., 2020).

Os resultados desses trabalhos mostram que a performance de um genótipo varia em função de gradientes macro e micro ambientais (ARAUJO et al., 2019; DA SILVA et al., 2019; ELLI et al., 2020). O macroambiente corresponde a todas as condições edafoclimáticas em diferentes latitudes, longitudes e tempo (anos ou estação agrícola). O microambiente corresponde à própria área de estudo ou área experimental. Esses padrões de desempenho frente a gradientes ambientais norteiam as decisões de plantios e de programa de melhoramento florestal.

Figura 2 – Normais climatológicas (1970-2000) mapeadas da precipitação pluviométrica e temperatura no Brasil.



Fonte: Os dados climáticos mapeados foram obtidos do banco WorldClim.

Na eucaliptocultura, em geral, a produção de madeira tende a diminuir em sítios próximos à linha do equador devido às altas temperaturas e chuvas mal distribuídas. Por outro lado, a produtividade tende a aumentar em latitudes maiores, onde as temperaturas são mais

amenas e as chuvas mais regulares (ELLI et al., 2020). De fato, a temperatura e a precipitação constituem as variáveis ambientais mais expressivas no desenvolvimento florestal (IVKOVIĆ et al., 2015; ARAUJO et al., 2019). Contudo, em ambientes tropicais a disponibilidade de água no solo tende a apresentar maior importância (ELLI et al., 2020; ROCHA et al., 2020). Essas são informações estratégicas para o melhoramento florestal, visto que representam variáveis de fácil obtenção e de grande impacto na produtividade.

2.2.2. Modelagem da produtividade utilizando covariáveis ambientais

O conhecimento das covariáveis de maior influência na produtividade pode melhorar a modelagem da interação genótipos x ambientes (GxA) e, conseqüentemente, a recomendação de cultivares em programas de melhoramento. Em virtude da natureza quantitativa de grande parte dos caracteres agrônômicos de interesse, a consideração dessas covariáveis em modelos biométricos pode resultar no incremento da capacidade preditiva (PIEPHO; DENIS; VAN EEUWIJK, 1998; MATHEWS et al., 2008; MALOSETTI et al., 2016; ELLI et al., 2020). Logo, as variáveis edafoclimáticas podem explicar grande parcela da variação produtiva.

As covariáveis ambientais podem ser utilizadas em diversas ferramentas para a predição da produtividade. Uma abordagem consiste na utilização de modelos de crescimento como o 3-PG (LANDSBERG; WARING, 1997), CABALA (BATTAGLIA et al., 2004) e APSIM (HOLZWORTH et al., 2018). Esses modelos são úteis para a predição e determinação de variáveis importantes relacionadas à produção. Por exemplo, Elli et al. (2020) utilizando o modelo APSIM, determinaram que a temperatura e precipitação pluvial exercem as maiores influências na produtividade de madeira, porém não da mesma forma em todo Brasil.

Outra vantagem da abordagem é a possibilidade de ampliar a recomendação de cultivares para ambientes não testados, reduzindo custos relacionados à implantação e avaliação de ensaios experimentais. Marcatti et al. (2017) utilizando um modelo de crescimento associado com a abordagem de modelos mistos (REML/BLUP) e sistema de informações geográficas (SIG), realizaram a recomendação mapeada de clones de eucalipto em uma região do estado da Bahia. Entretanto, esse tipo de abordagem também é possível com a utilização de outros modelos baseados em ensaios experimentais (ANNICCHIARICO; BELLAH; CHIARI, 2006; RESENDE et al., 2019; COSTA-NETO et al., 2020).

Na eucaliptocultura, dados climáticos diários e de longo prazo (i.e., normais, climatológicas ou provisórias) têm sido utilizados em associação a modelos para predição da

produtividade de madeira (MARCATTI et al., 2017; RYAN et al., 2020), volume específico (ROCHA et al., 2020) e altura dominante (SCOLFORO et al., 2017). Os esclarecimentos acerca da influência ambiental na produtividade obtida nos estudos dessa natureza fornecem informações importante para a incorporação de variáveis em modelos preditivos. Além disso, essas informações são úteis para inferências acerca do comportamento de genótipos frente a mudanças climáticas (SCOLFORO et al., 2017). Espera-se que o desempenho de diferentes clones no contexto dessas mudanças seja distinto, uma vez que estes apresentam constituição genética única, haja vista os níveis de expressão dos genes serão regulados de modos diferentes frente a variações do ambiente (BOYDEN; BINKLEY; STAPE, 2008; HODGINS-DAVIS; TOWNSEND, 2009).

As projeções do Painel Intergovernamental para a Mudança de Clima (IPCC) indicam um aumento de temperaturas médias e dias secos em todo o Brasil (GUTIÉRREZ et al., 2021). Embora se atribua certa resiliência a eucaliptocultura frente às mudanças climáticas (BOOTH, 2013), é provável que o agravamento das condições de estresse hídrico e térmico resulte no declínio da produtividade de madeira dos genótipos não tolerantes a essas condições. Assim, a identificação de fatores ambientais importantes relacionados adaptabilidade produtiva dos clones pode revelar algumas bases da interação GxA, melhorando a modelagem dessa interação e, conseqüentemente, a recomendação de cultivares específicos para diferentes cenários e ambientes.

2.2.3. Regressão via quadrados mínimos parciais (PLS) e suas aplicações

A regressão via quadrados mínimos parciais (PLS, do inglês *Partial Least Squares*) constitui uma ferramenta útil na predição de um conjunto de variáveis resposta (Y), em função de um grande conjunto de variáveis preditoras (X). A regressão PLS combina a regressão múltipla com a análise por componentes principais (PCA, do inglês *Principal Component Analysis*) para a obtenção de um conjunto de variáveis ortogonais, ou seja, livres de multicolinearidade (ABDI, 2010). Os componentes principais (vetores latentes) obtidos com essa análise diferem da PCA tradicional, visto que estes são decomposições conjuntas de X com Y que melhor explicam a possível covariância entre X e Y, ou seja, as variáveis latentes são componentes que melhor estimam Y.

As vantagens da utilização da regressão PLS vão além da eliminação dos problemas de colinearidade. O uso da ferramenta possibilita predições mais estáveis frente ao número de

observações e à identificação de variáveis importantes de maneira mais acurada que a técnicas usuais como regressão linear e regressão por componentes principais (CARRASCAL; GALVÁN; GORDO, 2009). Assim, têm-se utilizado a regressão PLS para a predição de produtividade de culturas utilizando covariáveis ambientais (COSTA-NETO, 2017; PORKER et al., 2020) e índices de vegetação, obtidos a partir de sensoriamento remoto (KAWAMURA et al., 2018; ZHANG et al., 2018).

Para a eucaliptocultura, os estudos utilizando a regressão PLS baseiam-se, majoritariamente, em técnicas de química computacional e espectrometria (FERREIRA et al., 2018; CASTRO et al., 2019). Dessa forma, o uso da ferramenta utilizando covariáveis ambientais constitui uma oportunidade para a predição da produtividade de madeira, identificação de covariáveis importantes e maior compreensão da interação GxA.

2.3. Interação genótipos x ambientes (GxA)

A recomendação e o desenvolvimento de novos clones e cultivares é dificultada pela interação GxA, devido à baixa consistência do desempenho de um genótipo em diferentes localidades, evidenciando uma resposta relativa e sensível à variação ambiental (CARGNIN et al., 2006). Essa instabilidade no desempenho relativo dos clones representa um grande entrave na identificação de materiais genéticos com ampla adaptabilidade e nas predições de desempenho dos clones, bem como no ganho de seleção (CRUZ; REGAZZI; CARNEIRO, 2012).

A natureza dessa interação pode ser atribuída a diferenças na resposta fisiológica e bioquímica de cada genótipo frente aos diferentes ambientes em que são submetidos (CRUZ; REGAZZI; CARNEIRO, 2012; KING, 2015). Cabe observar que, em programas de melhoramento, os genótipos são submetidos e avaliados em uma ampla gama de locais e épocas, em que, as diferentes combinações desses fatores conferem condições ambientais particulares (JAT et al., 2017; CASTRO et al., 2018). Devido ao fenômeno da interação GxA e dos efeitos ambientais médios, é esperado que ocorram respostas fenotípicas variadas entre os genótipos nos vários ambientes.

Essas respostas fenotípicas podem ser descritas pelo modelo simplificado: $F = G + A + GxA$. Por meio deste modelo, o fenótipo (F) é resultante da adição dos efeitos genéticos (G), ambientais (A) e da interação de genótipos com ambientes (GxA). A análise do conjunto de valores fenotípicos fornece informações de grande importância, como a influência relativa

da variação genética e ambiental expressa por meio da estimativa da herdabilidade (h^2) (MAZER; SCHICK, 1991). De maneira geral, quanto maior a influência do ambiente menor a contribuição genética para a expressão do caráter e, portanto, menor é a herdabilidade. Além disso, a interação GxA também contribui com a redução da herdabilidade, já que compõe o denominador da equação que estima esse parâmetro. Dessa forma, é necessário que o melhorista adote estratégias especiais para a seleção de genótipos superiores baseadas na natureza dessa interação (NUNES et al., 2002; CARVALHO et al., 2018; CASTRO et al., 2018).

Segundo Kang (1997), as interações GxA podem ser agrupadas em duas categorias: interações simples (*Noncrossover* ou interação quantitativa) e interações complexas (*Crossover* ou interação qualitativa). As interações simples ocorrem quando a classificação do desempenho dos genótipos não se altera entre os ambientes. Esse tipo de interação facilita a recomendação de cultivares, pois os melhores genótipos em um local também são os melhores em todos os demais locais analisados (NUNES et al., 2002; DE SOUZA et al., 2020). As interações complexas, por sua vez, ocorrem quando a classificação dos genótipos se altera nos ambientes. Este tipo de interação é um problema para o melhoramento genético, pois impossibilita a recomendação de genótipos em uma ampla gama de ambientes, uma vez que cada genótipo apresenta melhor adaptabilidade em um ou mais locais específicos (NUNES et al., 2002; CASTRO et al., 2018).

Uma vez identificada a resposta diferencial dos genótipos no gradiente de ambientes, deve-se adotar estratégias para a atenuação dos efeitos negativos da interação GxA. Os efeitos podem ser contornados com a recomendação dos genótipos mais adaptados para cada ambiente; com a seleção de genótipos com ampla estabilidade fenotípica entre os ambientes; e por meio da estratificação de ambientes, visando sub-regiões mais homogêneas e, conseqüentemente, com menor contribuição para a interação GxA (SANTOS et al., 2015).

Dentre tais estratégias, uma alternativa inicial interessante é a adoção da estratificação ambiental (CRUZ; REGAZZI; CARNEIRO, 2012; CASTRO et al., 2018; DE SOUZA et al., 2020). Com esse método, agrupam-se ambientes apresentam fatores edafoclimáticos macro ambientais mais uniformes, que resultam em interação GxA de baixa magnitude (ISIK; HOLLAND; MALTECCA, 2017).

O primeiro relato de estratificação ambiental é atribuído à Horner e Frey (HORNER; FREY, 1957). No estudo, a obtenção de número de sub-regiões ótimas resultou na redução de

30% da interação GxA predita para variedades de aveia conduzidas em nove locais, no estado de Iowa (EUA). Os dados obtidos possibilitaram a identificação de ambientes fortemente correlacionados, permitindo a formação de grupos ambientais que proporcionaram resposta fenotípica similar. Essa informação representa importante vantagem na otimização das estratégias de melhoramento, pois permite a identificação de áreas prioritárias para a condução de ensaios e maior facilidade na recomendação de novas cultivares (YAN; HUNT, 2001; ISIK; HOLLAND; MALTECCA, 2017).

Dentre as abordagens para a definição de grupos ambientais, os métodos *biplots* apresentam uma vantagem quanto à possibilidade da análise visual. Considerando as ferramentas disponíveis, destacam-se dois métodos no melhoramento florestal: os modelos aditivos principais com interação multiplicativa (AMMI) (PUPIN et al., 2018; ZHANG et al., 2018) e os modelos de efeitos genotípicos principais + efeitos de interação genótipos por ambientes (GGE biplot) (JASTRZEŹBOWSKI et al., 2018; ZHANG et al., 2018; ARAUJO et al., 2019). As técnicas utilizam a combinação de ANOVA e a análise de componentes principais (PCA) para o ajuste dos efeitos principais (genotípicos e ambientais) e de interação genótipos por ambientes (SILVA; BENIN, 2012; LI et al., 2017). Ao final da análise, obtém-se os efeitos G+GxA representados graficamente.

Uma das vantagens das análises *biplot* é a possibilidade de analisar, conjuntamente, a estratificação, a existência de genótipos mais adaptados em cada local e de genótipos mais estáveis nos mega-ambientes definidos. Dessa forma, foi possível inferir que, dentre as variáveis mais influentes nos efeitos GGE de espécies florestais, a temperatura e precipitação pluviométrica, são determinantes para a discriminação de mega-ambientes (IVKOVIĆ et al., 2015; ARAUJO et al., 2019). Essa é uma informação estratégica para o melhoramento florestal, visto que, estas representam covariáveis ambientais de fácil obtenção. Entretanto, apesar de uma parcela das variações macro ambientais serem controladas pela estratificação ambiental, as variações micro ambientais e aquelas causadas por variações imprevisíveis entre anos corroboram para persistência de problemas causados pelos efeitos de interação GxA. Sendo assim, as análises de estabilidade e adaptabilidade de genótipos são muito importantes em programas de melhoramento (EBERHART; RUSSELL, 1966; ARAUJO et al., 2019).

As estratégias de seleção por estabilidade e adaptabilidade podem ser executadas com o auxílio de diversas metodologias. Rezende et al. (2020) destacam que a utilização conjunta de abordagens com bases distintas é essencial para o incremento informacional sobre a estabilidade e adaptabilidade dos atributos estudados. Pode-se destacar que as ferramentas

com utilização crescente para esse objetivo no melhoramento florestal são: biplots (JASTRZEŹBOWSKI et al., 2018; PUPIN et al., 2018; ZHANG et al., 2018; ARAUJO et al., 2019); e Média Harmônica da Produtividade Relativa dos Valores Genéticos (MHPRVG) (CANUTO et al., 2016; DE OLIVEIRA et al., 2018; DE SOUZA et al., 2020).

As análises da estabilidade e adaptabilidade com base em coeficientes de regressão também são possíveis. A abordagem inicialmente desenvolvida por Eberhart e Russel (EBERHART; RUSSELL, 1966) utiliza a estimação de um coeficiente de regressão linear individual (β_i) associado a um índice ambiental para a determinação da adaptabilidade e da variância do desvio do modelo (σ_{ai}^2) para descrever a estabilidade de genótipos. No melhoramento florestal há alguns relatos do uso de um índice ambiental como base para a avaliação da adaptabilidade e estabilidade de genótipos, seja por meio de modelos lineares (NUNES et al., 2002; KUMAR et al., 2012; PEIXOTO et al., 2018) ou não lineares (MARCATTI et al., 2017).

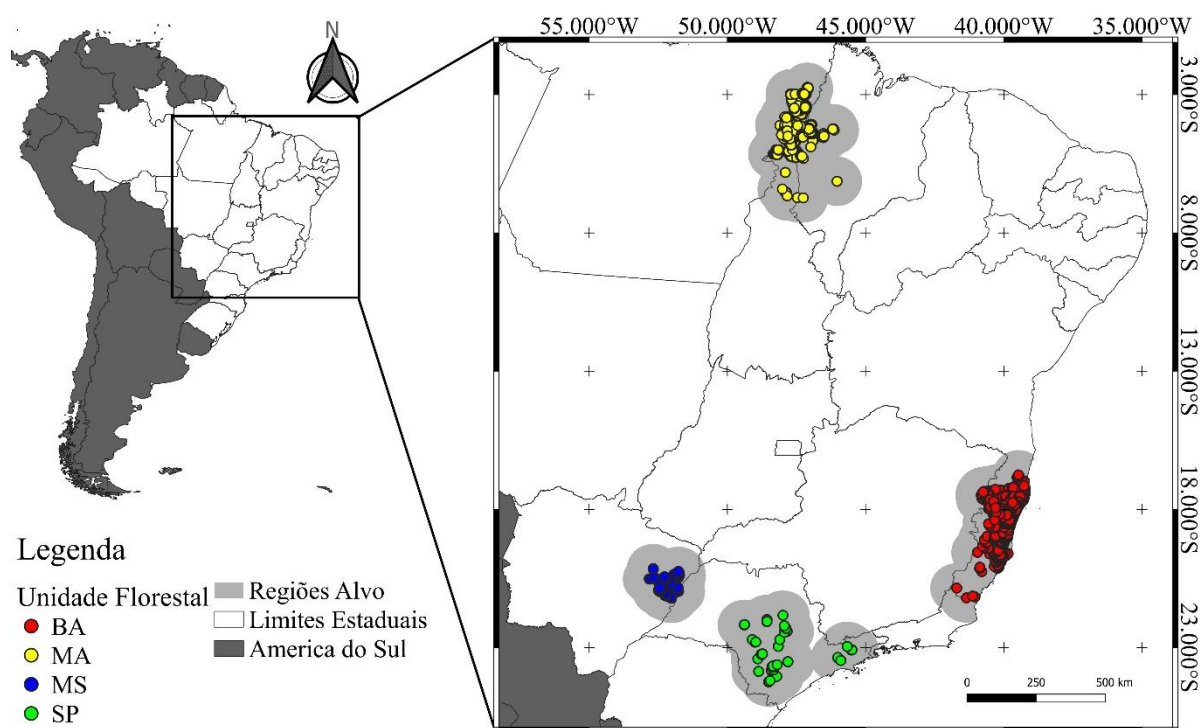
De fato, com os avanços computacionais tem-se empregado abordagens tradicionais com outras tecnologias como o sistema de informações geográficas (SIG) para a determinação da produtividade e adaptabilidade de materiais genéticos em nível espacial (ANNICCHIARICO; BELLAH; CHIARI, 2006; MARCATTI et al., 2017; COSTA-NETO et al., 2020). A abordagem baseia-se no emprego da ambientipagem para ajuste de modelos de predição de fenótipos. Xu (2016), defende que a ambientipagem deva ser usada para a compreensão da influência do ambiente utilizando caracteres de interesse (edáficos, climáticos, geográficos etc.), visto que esta compreende uma descrição detalhada do ambiente. Entretanto, por se tratar de um conceito relativamente novo, a incorporação da nova “tipagem” ainda é tímida nas mais diversas culturas.

3. MATERIAL E MÉTODOS

3.1. Áreas de estudo e material vegetal

Os dados utilizados foram cedidos gentilmente pela empresa Suzano S.A. e compreendem estimativas de incremento médio anual em volume de madeira (IMA) aos 7 anos. Ao todo, o inventário compreende 13.483 talhões implantados entre os anos 2006 e 2021. As florestas estão situadas nas unidades florestais (UNF) do Mato Grosso do Sul (MS), Maranhão (MA), Bahia (BA) e São Paulo (SP). A localização geográfica dos talhões utilizados em cada unidade florestal está representada na Figura 3.

Figura 3 – Distribuição geográfica das localidades consideradas para estudo através de quatro regiões no Brasil.



Fonte: Do autor (2022)

Como regiões alvo das previsões de produtividade (ver seção 3.4.2.) considerou-se apenas o conjunto de áreas referentes ao raio arbitrário de 100km de cada talhão inventariado. Além disso, para melhorar a modelagem da interação GxA, reduzindo a variação aleatória e a complexidade dos dados, foram obtidos valores médios de IMA em talhões próximos, considerando uma área quadrada de 10km² e plantados no mesmo semestre. Dessa forma, o conjunto total de dados (13.483 talhões) foi sumarizado em 2.626 valores médios de produtividade, expressa em m³.ha⁻¹.ano⁻¹.

Os genótipos utilizados no estudo compreendem seis clones comerciais amplamente utilizados pela empresa. As espécies que compõem os clones e o número de pontos geográficos referentes aos valores médios de produtividade por unidade florestal encontram-se na Tabela 1.

Tabela 1 – Número de pontos geográficos referentes aos valores médios de produtividade por clone e unidade florestal.

Clone	Espécie/Híbrido	Unidade Florestal				Total
		BA	MA	MS	SP	
CLZ001	<i>E. urophylla</i>	1028	123	70	25	1246
CLZ002	<i>E. urophylla</i>	111	20	0	4	135
CLZ003	<i>E. grandis</i> x <i>E. urophylla</i>	145	136	1	7	289
CLZ004	<i>E. urophylla</i>	523	55	0	0	578
CLZ005	<i>E. urophylla</i>	62	1	0	5	68
CLZ006	<i>E. grandis</i> x <i>E. urophylla</i>	309	0	1	0	310
Total		2178	335	72	41	2626

3.2. Bancos de dados edafoclimáticos

Os dados edáficos foram obtidos no banco SoilGrids (<https://soilgrids.org/>), com uma resolução espacial de 2,5 minutos, utilizando o pacote *rgdal* v.1.5-27 (BIVAND; KEITT; ROWLINGSON, 2021). Com exceção da capacidade troca catiônica (CTC), optou-se pela utilização de covariáveis referentes a física do solo, uma vez que as características químicas são alteradas pela empresa a cada novo ciclo de plantio, por meio de corretivos e fertilizantes. Assim, as variáveis edáficas extraídas do SoilGrids foram: teor de argila (%), teor de areia (%), densidade do solo e CTC.

Para a escolha dos dados climáticos foram comparados três bancos: dados das estações meteorológicas do Instituto Nacional de Meteorologia - INMET (<https://portal.inmet.gov.br/dadoshistoricos>), espalhadas pelo Brasil; dados do banco NASA Power (<https://power.larc.nasa.gov/>), obtidos de sensores em satélites; e dados do banco WorldClim (<https://www.worldclim.org>), baseados em normais climatológicas referentes aos anos 1970-2000. Diferentemente dos dados do WorldClim, os dois primeiros bancos possuem informações diárias.

Os três bancos foram comparados com relação às 19 biovariáveis ambientais descritas na Tabela 2. As biovariáveis visam maximizar a utilização da temperatura e precipitação para explicar fenômenos biológicos. As 19 biovariáveis já estão prontamente disponíveis no banco

WorldClim. Para os dados do INMET e NASA Power, elas foram estimadas pelo pacote *dismo* v.1.3-3 (HIJMANS et al., 2017), a partir das médias mensais do período de plantio ao período de medição.

Tabela 2 – Covariáveis bioclimáticas.

Variável	Descrição
BIO1	Temperatura média anual
BIO2	Amplitude térmica diária
BIO3	Isotermalidade (BIO2/BIO7) x100
BIO4	Sazonalidade da temperatura (desvio Padrão ×100)
BIO5	Temperatura máxima do mês mais quente
BIO6	Temperatura mínima do mês mais frio
BIO7	Amplitude térmica anual (BIO5-BIO6)
BIO8	Temperatura média do trimestre mais úmido
BIO9	Temperatura média do trimestre mais seco
BIO10	Temperatura média do trimestre mais quente
BIO11	Temperatura média do trimestre mais frio
BIO12	Precipitação pluviométrica anual
BIO13	Precipitação pluviométrica do mês mais úmido
BIO14	Precipitação pluviométrica do mês mais seco
BIO15	Sazonalidade da precipitação pluviométrica (coeficiente de variação)
BIO16	Precipitação pluviométrica do trimestre mais úmido
BIO17	Precipitação pluviométrica do trimestre mais seco
BIO18	Precipitação pluviométrica do trimestre mais quente
BIO19	Precipitação pluviométrica do trimestre mais frio

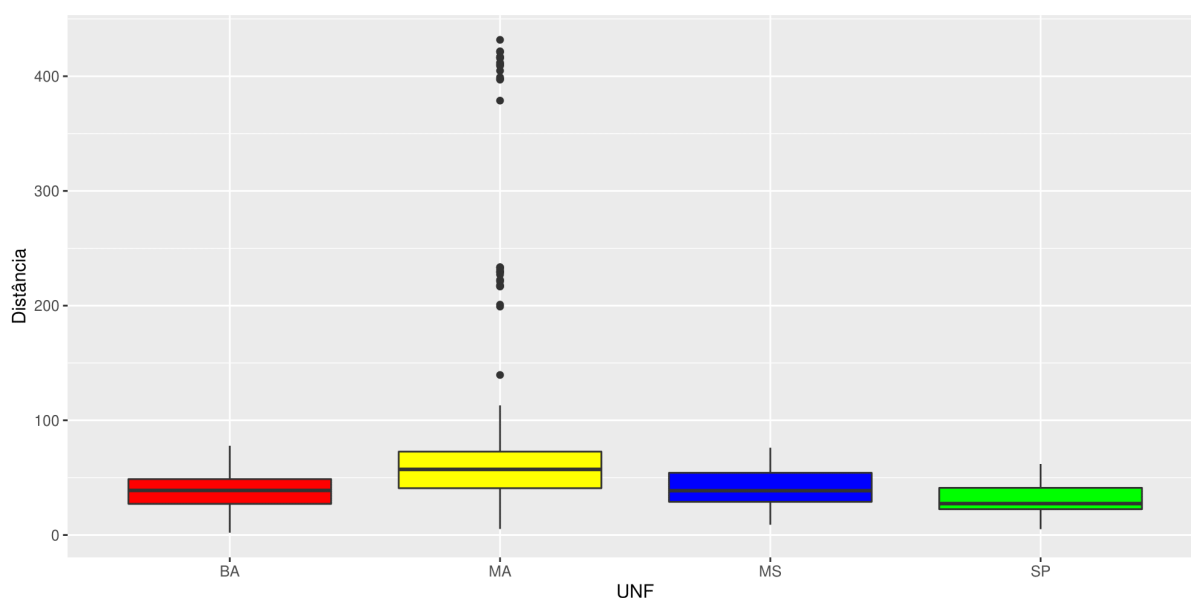
3.2.1. Dados do banco INMET

As covariáveis obtidas no banco de dados do Instituto Nacional de Meteorologia (INMET) são referentes ao período entre o plantio e a medição de cada talhão inventariado. Os dados foram obtidos em 43 estações meteorológicas automáticas presentes nos oito estados em que os talhões estão situados. A associação das covariáveis com os dados do

inventário foi realizada a partir da estação mais próxima contendo as medições relativas ao ano de implantação do talhão florestal. Essa exigência visou evitar que pontos fossem associados a estações implantadas apenas em anos seguintes ao do plantio.

Com o intuito de aumentar a fidelidade dos dados ambientais, realizou-se uma análise das distâncias entre as estações meteorológicas e seus pontos associados (Figura 4). Na unidade florestal do Maranhão foram identificados talhões bastante distantes ($> 100\text{km}$) de qualquer estação meteorológica do INMET. Optou-se por utilizar apenas os pontos distantes em até no máximo 50km de alguma estação meteorológica. Esse critério resultou na perda de aproximadamente 25% dos dados e, com isso, restaram 1913 pontos para esse banco de dados.

Figura 4 – Histograma e boxplots das distâncias (km) entre as estações e as florestas inventariadas para esse estudo.



Legenda: BA= Bahia, MA= Maranhão, MS= Mato Grosso do Sul, SP= São Paulo.

Fonte: Do autor (2022)

3.2.2. Dados do NASA Power

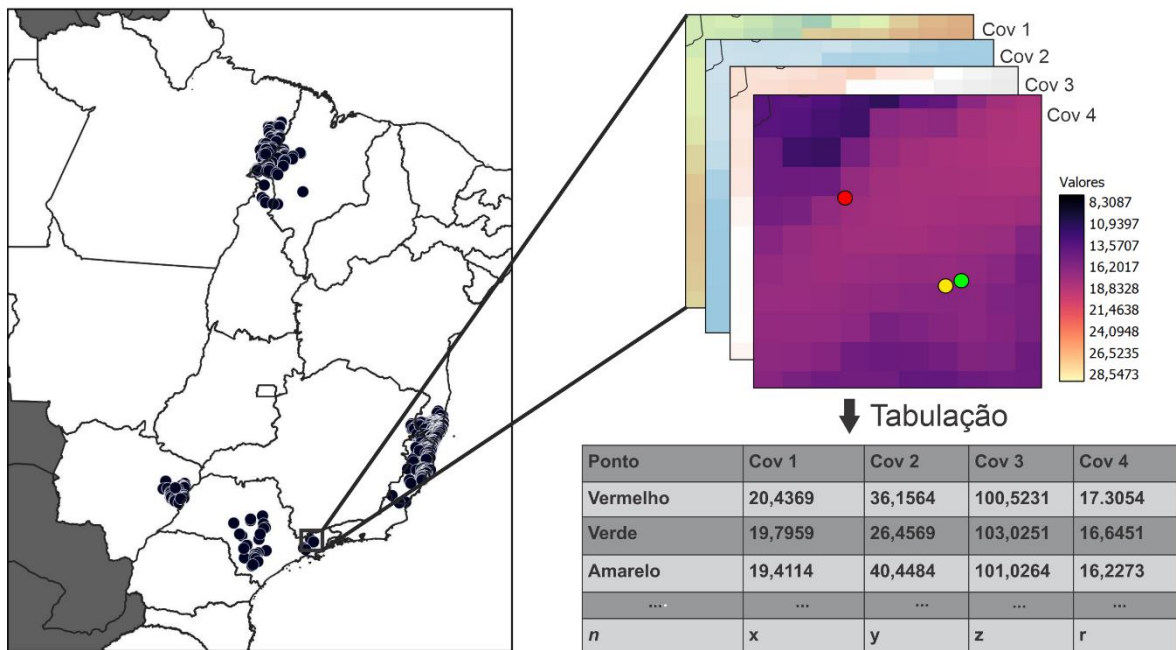
As covariáveis obtidas no banco de dados do NASA Power são referentes ao período entre o plantio e a medição de cada talhão inventariado. Apesar de possuir uma resolução temporal diária, a maior desvantagem desses dados é a sua baixa resolução espacial de $0,5^\circ$ ($\sim 55\text{km}^2$), podendo comprometer a variabilidade dos dados ambientais. Os dados foram obtidos

a partir do pacote *nasapower* v.4.0.0 (SPARKS, 2018). Com o pacote foram extraídos dados climáticos de 2372 pontos, que compuseram os dados para esse banco.

3.2.3. Dados do WorldClim

Os dados disponíveis no WorldClim compreendem normais climatológicas definidas por observações dos anos 1970 até 2000. Visando um balanço entre detalhamento e desempenho computacional, foi escolhida a resolução espacial de 2,5 minutos ($\sim 5 \text{ km}^2$). A obtenção dos dados climáticos foi realizada para cada ponto, considerando o pixel em que este está situado (Figura 5). Para esse banco, todos os 2.626 pontos puderam ser caracterizados a partir da extração de covariáveis utilizando o pacote *raster* v3.4-13 (HIJMANS, 2020).

Figura 5 – Esquema da extração de covariáveis hipotéticas (COV1 e COV2) por ponto a partir de arquivos raster.



Fonte: Do autor (2022)

O banco, juntamente ao SoilGrids, foi utilizado para a determinação da matriz de similaridade ambiental entre as várias florestas amostradas para fins de caracterização das regiões. Os bancos foram utilizados pois foram escolhidos para a predição e produção de mapas de recomendação (seção 3.4.2). A matriz de variâncias-covariâncias foi obtida a partir do método do kernel gaussiano, utilizando o pacote *EnvRtype* (COSTA-NETO et al., 2021).

3.3. Avaliação da capacidade preditiva com os diferentes bancos de dados

Para a avaliação das vantagens da inclusão dos dados de solo, foram ajustados modelos com diferentes combinações de dados para os seis clones do estudo. As combinações comparadas foram constituídas pelos diferentes bancos climáticos, com e sem os dados de solo. Além das variáveis edafoclimáticas, em todos os modelos testados foram incluídas as três variáveis geográficas (elevação, latitude e longitude). Os ajustes dos modelos de regressão de quadrados mínimos parciais (PLS) foram realizados utilizando o pacote *plsdepot* do R, com duas variáveis latentes (seção 3.4.1). Para validação da capacidade preditiva dos modelos obtidos utilizou-se o método *leave-one-out* em que cada valor é deixado de fora do ajuste do modelo PLS para, em seguida, ser estimado com a equação de regressão obtida. Devido ao interesse na elaboração de mapas de produtividade para recomendação de cultivares (seção 3.4.2), o *leave-one-out* foi conduzido a partir do ajuste com os diferentes conjuntos e a predição realizada com os valores das covariáveis ambientais extraídos dos arquivos de imagem disponível no WorldClim. A escolha desse banco de dados para a plotagem deu-se pela disponibilidade de dados espaciais interpolados, prontos para uso e devido à imprevisibilidade dos fatores climáticos de ano para ano. Dessa forma, considerou-se que a utilização das normais climatológicas contidas no WorldClim (1970-2000) é mais adequada para a predição de dados.

Após a validação, as capacidades preditivas dos modelos com os diferentes bancos de dados puderam ser avaliadas. Essa avaliação foi realizada pela comparação dos valores observados e preditos e dos valores do coeficiente de determinação (R^2), erro predito médio (RMSE), correlação de Spearman (ρ) e índice de concordância (c). O índice de concordância foi definido como a porcentagem de valores preditos nas mesmas classes arbitrárias de produtividade dos valores observados. As classes de produtividade foram definidas com um intervalo de $10 \text{ m}^3 \cdot \text{ha}^{-1} \cdot \text{ano}^{-1}$.

3.4. Influência das covariáveis ambientais na produtividade dos clones de eucalipto

3.4.1. Modelagem da produtividade de madeira via regressão PLS

A regressão via quadrados mínimos parciais, constitui uma metodologia útil na predição de um conjunto de variáveis resposta (Y) em função de um grande conjunto de variáveis predictoras (X) (ABDI, 2010). Nesse estudo, o vetor Y representa a produtividade

dos clones dentro dos ambientes de estudo. A decomposição da matriz X via mínimos quadrados parciais iterativos não-lineares (NIPALS), foi realizada por meio dessas três operações:

$$X = TP'$$

$$T'T = I$$

$$\hat{Y} = TBC'$$

Em que, I : matriz identidade; T : matriz contendo os escores dos vetores latentes; P' : transposta da matriz de coeficientes lineares (P); \hat{Y} : matriz coluna da produtividade de madeira (IMA, $m^3 \cdot ha^{-1} \cdot ano^{-1}$); B : matriz diagonal contendo os pesos da regressão; C' : transposta da matriz de pesos das variáveis preditoras (C). A definição do número de vetores latentes adequados contidos na matriz T para a predição ocorrerá pela métrica Q_L^2 :

$$Q_L^2 = 1 - \frac{PRESS_L}{RESS_{L-1}} \quad (1)$$

Em que, **RESS**: Soma de quadrados residual dos valores estimados (modelo fixo); **PRESS**: Soma de quadrados residual dos valores preditos (modelo aleatório); L : número de variáveis latentes. Como critério de seleção serão considerados todas as variáveis latentes que possuírem um $Q_L^2 > 0,095$. Utilizando L variáveis latentes ($L = 1, 2, \dots, n$), a predição dos valores de produtividade se dá conforme a equação: $Y = P^+BC'$. Em que, Y : produtividade predita; e P^+ : inversa generalizada de Moore-Penrose da matriz P . Enfim, a predição dos valores de IMA não padronizados foi realizada com a seguinte equação:

$$\hat{Y}_{ij} = (m + a) + \sum b_L c_L + \varepsilon_{ij} \quad (2)$$

Em que, $\mathbf{m} + \mathbf{a}$: soma da média geral ao intercepto; $\sum \mathbf{b}_L \mathbf{c}_L$: somatório do produto dos pesos da regressão (\mathbf{b}) com os pesos das covariáveis (\mathbf{c}); e ε_{ij} : diferença do valor observado pelo predito da observação j do clone i (resíduo). Além disso, visando a identificação das tendências de erros preditivos ao longo das quatro unidades florestais foi adotado seguinte critério: $\varepsilon_{ij} > \overline{RMSE}$: valores subestimados; $\varepsilon_{ij} < -\overline{RMSE}$: valores superestimados; e $-\overline{RMSE} < \varepsilon_{ij} < \overline{RMSE}$: valores adequados. Em que, \overline{RMSE} : Erro predito médio entre todos os clones.

Por fim, as influências de cada covariável ambiental sobre a variável resposta foram avaliadas via análise visual do *biplot* ou círculo de correlações e do ranqueamento das cinco covariáveis mais influentes. O ranqueamento foi realizado a partir dos coeficientes de regressão normalizados. A modelagem PLS foi realizada com o pacote *plsdepot* v.0.6.17 (SANCHEZ, 2012).

3.4.2. Elaboração de mapas temáticos sob valores preditos

O modelo ajustado pela regressão PLS foi utilizado para a elaboração dos mapas com os valores preditos obtidos “pixel a pixel” para as regiões alvo. As estimativas de produtividade foram realizadas com base no valor das covariáveis ambientais (ou ambiótipo) de cada pixel, obtidas a partir do WorldClim com resolução espacial de 2,5 minutos (~ 5km²). Assim, utilizando o software QGIS (QGIS DEVELOPMENT TEAM, 2021) e os pacotes *raster* e *sp* v. 1.4-2 do programa R (PEBESMA; BIVAND, 2005) foram gerados dois tipos de mapas: 1) adaptabilidade predita regionalizada; e 2) clone vencedor “pixel a pixel”.

Os mapas do tipo 1 foram gerados a partir das estimativas de produtividade (IMA) de cada clone em cada pixel. Nesse mapa foram atribuídas diferentes cores a cada uma das classes arbitrárias, variando de 10 em 10 m³.ha⁻¹.ano⁻¹. A partir disso, pôde-se identificar também genótipos mais estáveis, como aqueles que possuem o menor número de classes de produtividade diferentes. Quantitativamente, essa definição foi feita a partir do cálculo do coeficiente de variação dos valores de produtividade (CV%).

Os mapas do tipo 2 foram gerados para cada região alvo a partir da comparação pixel a pixel dos valores de produtividade preditos para cada um dos seis clones. Para a distinção, cada clone foi representado por uma cor, possibilitando a associação de cada pixel (ou área) ao clone mais produtivo. Esses mapas têm como objetivo a visualização das regiões em que

cada clone foi vencedor, em nível de pixel, facilitando as recomendações dos clones em cada região alvo dos plantios.

3.5. Recursos computacionais

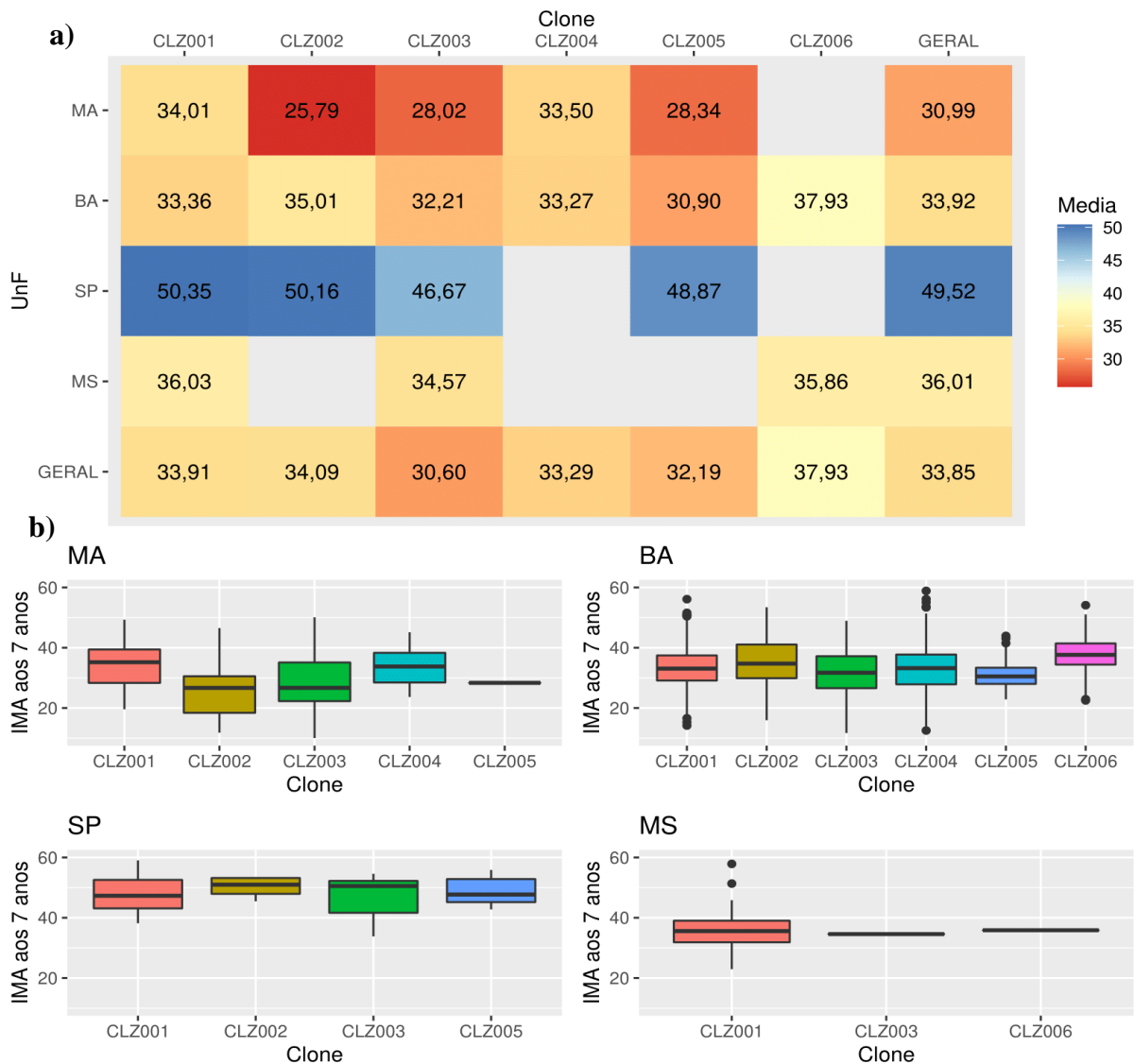
As análises foram realizadas no *software* R v. 4.0.5 (R CORE TEAM, 2021), instalado em um servidor de alta performance com sistema operacional Linux. As análises gráficas foram realizadas com os pacotes *ggplot2* v.3.31 (WICKHAM., 2016), *Ggally* v.2.1.2 (SCHLOERKE et al., 2021), *ggforce* v. 0.3.3 (PEDERSEN, 2021), *gridExtra* v.2.3 (AUGUIE; ANTONOV, 2017). A manipulação de dados tabulares foi realizada com o pacote *dplyr* v. 1.8.6 (WICKHAM et al., 2021).

4. RESULTADOS

4.1. Caracterização dos dados de inventário e ambientais

As médias de produtividade e suas distribuições obtidas a partir dos inventários florestais encontram-se na Figura 6. Os maiores valores de IMA foram observados para a unidade florestal de São Paulo ($49,52 \text{ m}^3 \cdot \text{ha}^{-1} \cdot \text{ano}^{-1}$), enquanto os menores foram observados para o estado do Maranhão ($30,99 \text{ m}^3 \cdot \text{ha}^{-1} \cdot \text{ano}^{-1}$).

Figura 6 – Caracterização da produtividade de eucalipto (IMA, $\text{m}^3 \cdot \text{ha}^{-1} \cdot \text{ano}^{-1}$) de seis clones (CLZ001, CLZ002, CLZ003, CLZ004, CLZ005, CLZ006) e geral em quatro unidades florestais (MA, BA, SP, MS). **a)** Médias de produtividades. **b)** Distribuição dos valores de produtividade.



Fonte: Do autor (2022)

O CLZ001 se destaca nas unidades de São Paulo (SP), Mato Grosso do Sul (MS) e Maranhão (MA). Os clones CLZ006 no MS, o CLZ002 em SP e o CLZ004 em MA apresentaram o segundo melhor desempenho dentro de cada região. As diferenças de IMA médio desses clones em relação a CLZ001 não ultrapassam $0,51 \text{ m}^3 \cdot \text{ha}^{-1} \cdot \text{ano}^{-1}$. Quanto à unidade da Bahia (BA), o CLZ006 foi o clone de melhor desempenho. Nos talhões inventariados na Bahia, o genótipo apresentou uma produtividade média aproximadamente $3 \text{ m}^3 \cdot \text{ha}^{-1} \cdot \text{ano}^{-1}$ superior ao segundo clone mais produtivo (CLZ002).

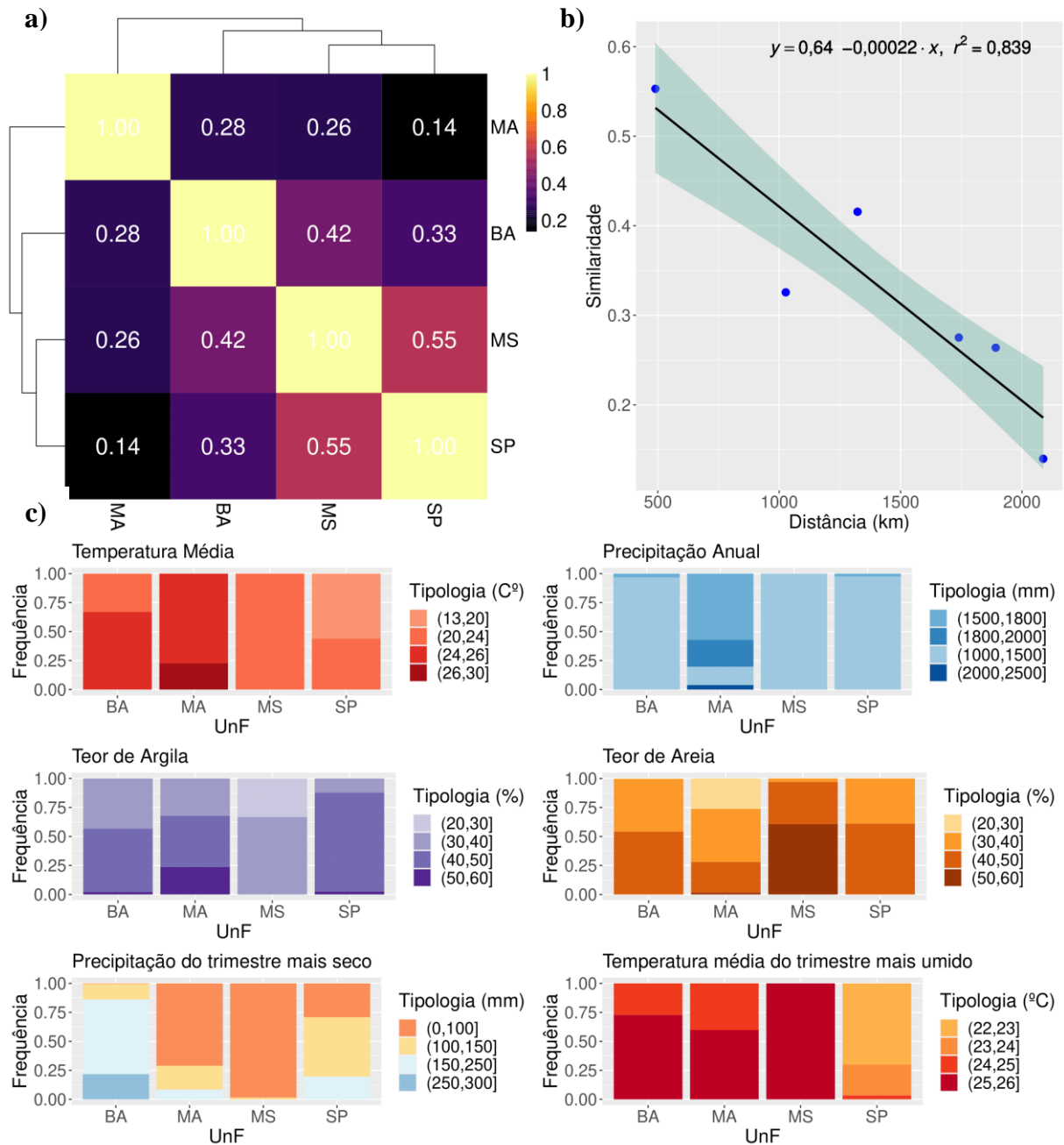
Quanto à variação na produção, a unidade BA apresentou a maior quantidade de *outliers* e as maiores amplitudes de produtividade (Figura 6.b). O comportamento inverso ocorre em São Paulo, sendo a região com as menores flutuações quanto ao desempenho dos clones. Esses padrões de resposta distintos ao longo dos ambientes resultam dos efeitos do ambiente e da interação GxA. Dessa forma, também foi realizada a caracterização ambiental das regiões do estudo utilizando os dados do WorldClim e SoilGrids (Figura 7).

As unidades SP e MS apresentaram a maior similaridade (0,55), seguido de BA e MS (0,42) e BA e SP (0,33) (Figura 7.a). O conjunto de talhões situados na unidade MA apresentou as menores similaridades com todas as demais regiões, sendo o ambiente mais distinto. De forma geral, a similaridade entre as unidades se correlaciona com a distância geográfica entre elas ($r = -0,916$). A relação linear entre a similaridade e a distância dos pontos médios das unidades florestais (Figura 7.b) relevou redução de similaridade de aproximadamente 0,022 a cada aumento de 100km ($R^2 = 0,839$, p-valor $< 0,001$).

As unidades da Bahia e Maranhão apresentam as maiores temperaturas médias e as maiores amplitudes de latitude ($2.74^\circ\text{S} - 6.73^\circ\text{S}$ e $16.75^\circ\text{S} - 21.20^\circ\text{S}$, respectivamente). Além disso, essas unidades apresentam condições de estresse hídrico e térmico mais rigorosas que as demais. Entretanto, Maranhão apresenta meses secos mais severos e maior variabilidade quanto à textura do solo e precipitação anual (Figura 7.c).

As unidades de São Paulo e Mato Grosso do Sul apresentam as menores temperaturas médias anuais, porém, há diferenças intra-anuais. As temperaturas médias do trimestre mais chuvoso e os volumes de precipitação pluviométrica no trimestre mais seco em Mato Grosso do Sul são semelhantes às do Maranhão e Bahia. Em contrapartida, os talhões de São Paulo apresentam as menores temperaturas médias durante todo ano e embora não contenham os maiores volumes de chuva no período seco, a região possui solos com maiores teores de argila, garantindo maiores reservas hídricas para o desenvolvimento vegetal.

Figura 7 – Caracterização ambiental. **a)** Matriz de similaridade ambiental utilizando um conjunto de 26 covariáveis ambientais. **b)** Relação entre distância geográfica e similaridade para as quatro unidades florestais. **c)** Frequência de tipologias de covariáveis ambientais por unidade florestal.



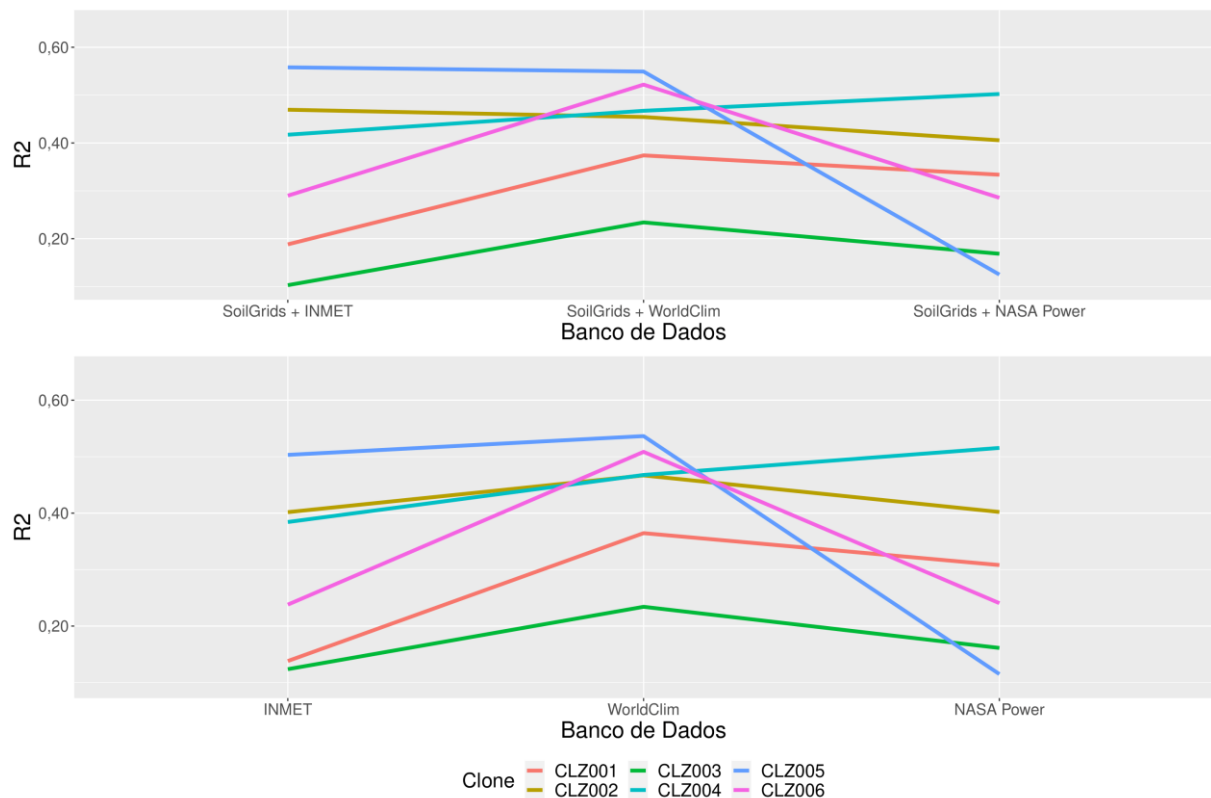
Fonte: Do autor (2022)

4.2. Definição dos bancos de dados ambientais

Para avaliar se há vantagem na inclusão de covariáveis de solo, os valores de R^2 da capacidade preditiva foram comparados graficamente (Figura 8). Observou-se que os dados de solo resultaram alguma melhoria de modelagem apenas para os conjuntos que utilizam

covariáveis em nível diário (INMET e NASA Power). Além disso, não houve prejuízos quanto à capacidade preditiva para o WorldClim. Assim, optou-se pela inclusão das covariáveis edáficas na modelagem. As métricas da capacidade preditiva, visando a escolha do conjunto de dados climáticos para a modelagem encontram-se na Tabela 3.

Figura 8 – Comparação dos coeficientes de determinação (R^2) da capacidade preditiva utilizando modelos ajustados a partir de diferentes combinações de bancos de dados, com e sem dados edáficos do SoilGrids.



Fonte: Do autor (2022)

Não houve um banco de dados vencedor para todos os clones quanto ao R^2 . Com os dados do WorldClim obtiveram-se maiores valores de R^2 para os clones CLZ001 (0,374), CLZ003 (0,234) e CLZ006 (0,522). Para os demais clones foram obtidos valores ligeiramente superiores para os clones CLZ002 (0,469) e CLZ005 (0,558) com o banco do INMET e para o clone CLZ004 (0,502) com o NASA Power. Com o WorldClim, foram obtidos os valores mais adequados quanto ao erro predito médio (RMSE), não ultrapassando $7,6 \text{ m}^3 \cdot \text{ha}^{-1} \cdot \text{ano}^{-1}$. Com os demais bancos de dados, os valores de RMSE comumente apresentaram valores superiores a $10 \text{ m}^3 \cdot \text{ha}^{-1} \cdot \text{ano}^{-1}$. Além disso, foram obtidas as porcentagens de valores preditos nas mesmas classes de produtividade (variando de 10 em $10 \text{ m}^3 \cdot \text{ha}^{-1} \cdot \text{ano}^{-1}$) dos valores

observados. A métrica denominada índice de concordância foi desenvolvida para medir a fidelidade dos mapas de produtividade gerados, que utilizam as classes para a visualização das tendências espaciais da variação do IMA.

Tabela 3 – Métricas da capacidade preditiva dos modelos obtida com o método *leave-one-out* para cada conjunto de dados ambientais. *c*: Índice de concordância; *c+*: porcentagem de valores acima do intervalo (10 em 10 m³.ha⁻¹.ano⁻¹); *c-*: porcentagem de valores abaixo do intervalo; *R*²: coeficiente de determinação; *RMSE*: erro preditivo médio; *ρ*: correlação de Spearman; e *n*: número de observações.

Clone	<i>c</i>	<i>c+</i>	<i>c-</i>	<i>R</i> ²	<i>RMSE</i>	<i>ρ</i>	<i>n</i>
WorldClim							
CLZ001	0,600	0,201	0,199	0,374	5,446	0,599	1246
CLZ002	0,570	0,259	0,170	0,454	6,721	0,688	135
CLZ003	0,519	0,311	0,170	0,234	7,575	0,363	289
CLZ004	0,609	0,202	0,189	0,467	5,191	0,692	578
CLZ005	0,471	0,338	0,191	0,549	4,517	0,381	68
CLZ006	0,732	0,142	0,126	0,522	3,734	0,722	310
NASA Power							
CLZ001	0,000	1,000	0,000	0,334	39,403	0,566	1133
CLZ002	0,000	1,000	0,000	0,406	61,556	0,706	133
CLZ003	0,367	0,529	0,104	0,169	11,745	0,324	289
CLZ004	0,000	1,000	0,000	0,502	31,098	0,718	520
CLZ005	0,000	1,000	0,000	0,125	49,977	0,357	68
CLZ006	0,026	0,974	0,000	0,286	15,911	0,595	229
INMET							
CLZ001	0,260	0,726	0,015	0,188	10,258	0,451	940
CLZ002	0,063	0,937	0,000	0,469	16,436	0,669	111
CLZ003	0,319	0,605	0,076	0,103	11,040	-0,054	157
CLZ004	0,156	0,833	0,010	0,417	10,919	0,698	384
CLZ005	0,392	0,255	0,353	0,558	4,744	-0,043	51
CLZ006	0,293	0,700	0,007	0,290	7,994	0,566	270

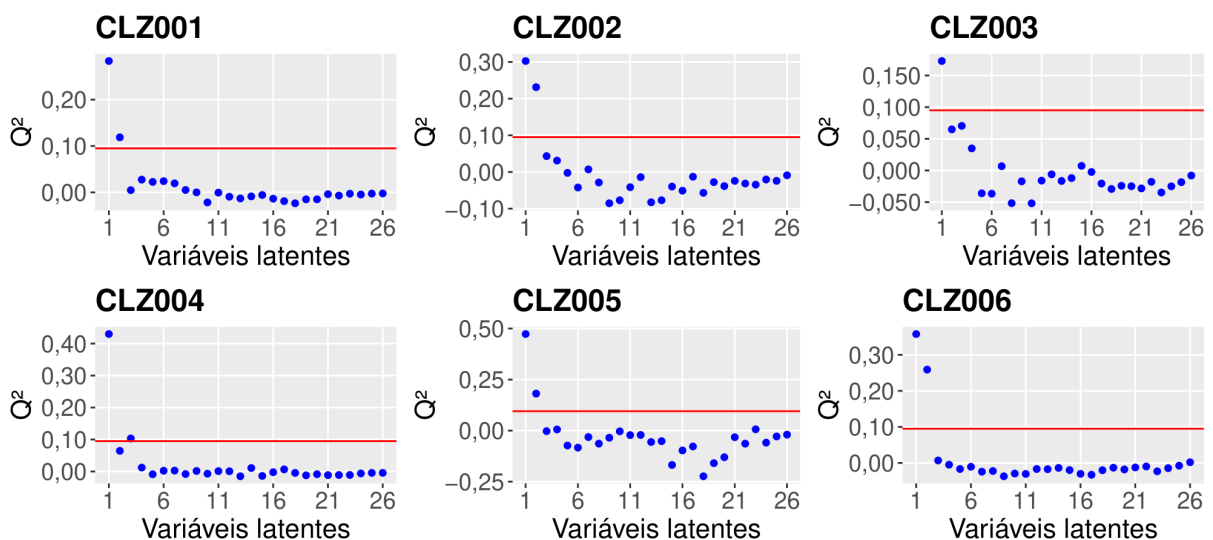
Os maiores valores de concordância (*c*) dos valores preditos nas classes de produtividade dos valores observados também foram obtidos a partir do modelo ajustado com WorldClim (0,471 – 0,732), enquanto os menores foram obtidos com o modelo do NASAPower (0,00 – 0,367). É importante notar que, os modelos ajustados para outros bancos de dados resultaram valores de produtividade preditos em classes superiores aos dos valores observados, indicando superestimativas. Os menores valores de *c* indicam que a modelagem

PLS com os dados do INMET e NASA Power não são adequados para elaboração de mapas utilizando os arquivos de imagem (*rasters*) contidos no WorldClim. Logo, com base nas métricas estabelecidas, optou-se pela utilização das variáveis climáticas presentes no WorldClim com os dados edáficos do SoilGrids para o ajuste e predição da produtividade dos clones.

4.3. Modelagem da produtividade de madeira via regressão PLS

Utilizando o conjunto de 19 biovariáveis obtidas no WorldClim, quatro variáveis de solo (teor de argila, teor de areia, densidade do solo e CTC) obtidas no SoilGrids e as três variáveis geográficas (latitude, longitude e altitude), realizou-se a modelagem PLS para a predição dos valores de IMA aos 7 anos. Optou-se por utilizar duas variáveis latentes visto que este é o número ótimo de acordo com o critério da razão entre as somas de quadrados residual dos valores preditos e estimados (Figura 9). Para o clone CLZ003, o número ótimo foi de apenas um componente, entretanto, a metodologia exige o uso de no mínimo dois. Logo, para esse clone também foram utilizadas duas variáveis latentes.

Figura 9 – Valores da razão de soma de quadrados dos valores estimados (RESS) pela soma de quadrados residual dos valores preditos (PRESS) para cada variável latente (Q^2).

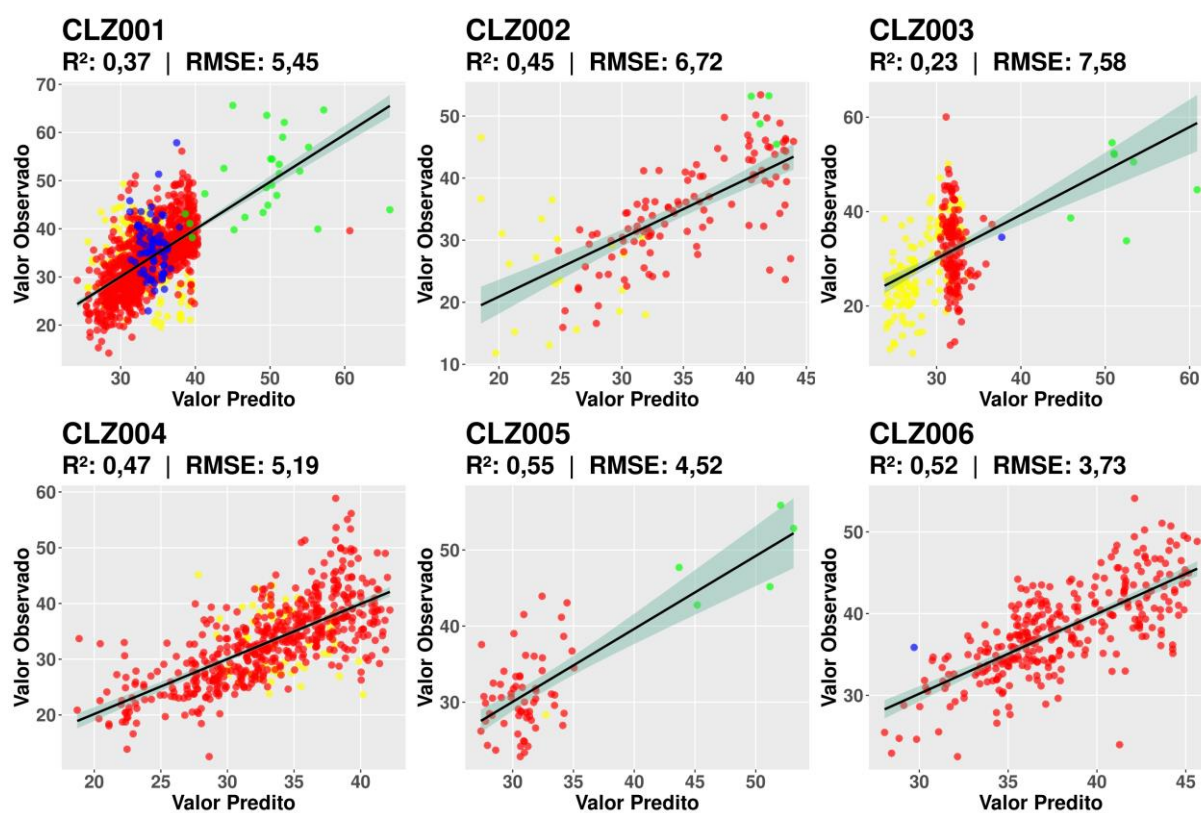


Legenda: Em azul, os pontos referentes ao valor de Q^2 (Equação 1). Em vermelho, a linha do limite pré-estabelecido ($Q^2 > 0,095$). Levou-se em conta um número mínimo de duas variáveis latentes.

Fonte: Do autor (2022).

A capacidade preditiva dos modelos ajustados foi avaliada utilizando a abordagem *leave-one-out*. Os valores de R^2 , RMSE e o gráfico de valores preditos e observados encontram-se na Figura 10. Entre os clones avaliados, CLZ006 teve a melhor modelagem da sua produtividade ($R^2 = 0,52$ e $RMSE = 3,73$), enquanto o modelo ajustado para o clone CLZ003 apresentou o pior desempenho ($R^2 = 0,23$ e $RMSE = 7,58$).

Figura 10 – Gráfico de dispersão dos valores preditos e observados para cada clone avaliado no estudo.



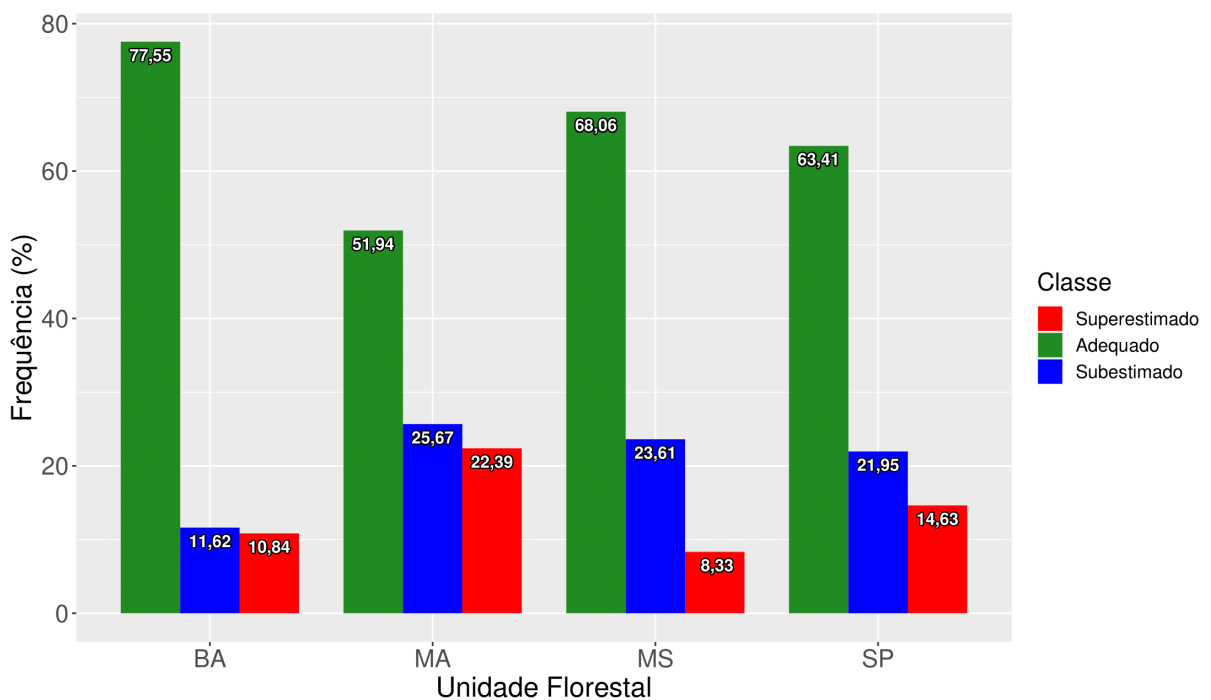
Legenda: Em vermelho, talhões referentes a unidade florestal da Bahia. Em amarelo, talhões referentes a unidade florestal do Maranhão. Em verde, talhões referentes a unidade florestal de São Paulo. Em azul, talhões referentes a unidade florestal do Mato Grosso do Sul. Em preto, a linha de tendência.

Fonte: Do autor (2022).

Para todos os clones é possível observar uma maior quantidade de dados obtidos das unidades florestais Bahia e Maranhão (Tabela 1). Essas regiões são bastante afetadas por estresses abióticos como a ocorrência de seca e do distúrbio fisiológico do eucalipto. Entretanto, a unidade florestal do Maranhão apresentou a menor proporção de valores adequados considerando RMSE médio para os clones, enquanto a Bahia apresentou a maior (Figura 11). Essa maior proporção de bons ajustes na Bahia pode ter ocorrido devido à participação majoritária da unidade no conjunto de dados. Com isso, os desempenhos

preditivos dos modelos nas demais regiões parecem seguir o grau de similaridade com os ambientes ali situados. Dessa forma, a unidade Maranhão foi prejudicada, mesmo apresentando o segundo maior volume de dados, já que esta possui as condições ambientais mais distintas (Figura 7.a). Além disso, considerando todos os clones juntos, uma maior proporção de predições classificadas como adequadas (~90%) foi obtida no raio ótimo de transporte de 100km até as fábricas.

Figura 11 – Frequência das classes de valores preditos por unidade florestal.



Legenda: $\overline{RMSE}_{clones} = 5,53 \text{ m}^3 \cdot \text{ha}^{-1} \cdot \text{ano}^{-1}$; As predições foram classificadas como subestimadas quando $e_{ij} > \overline{RMSE}_{clones}$; superestimados quando $e_{ij} < -\overline{RMSE}_{clones}$ e adequadas quando $-\overline{RMSE}_{clones} < e_{ij} < \overline{RMSE}_{clones}$.

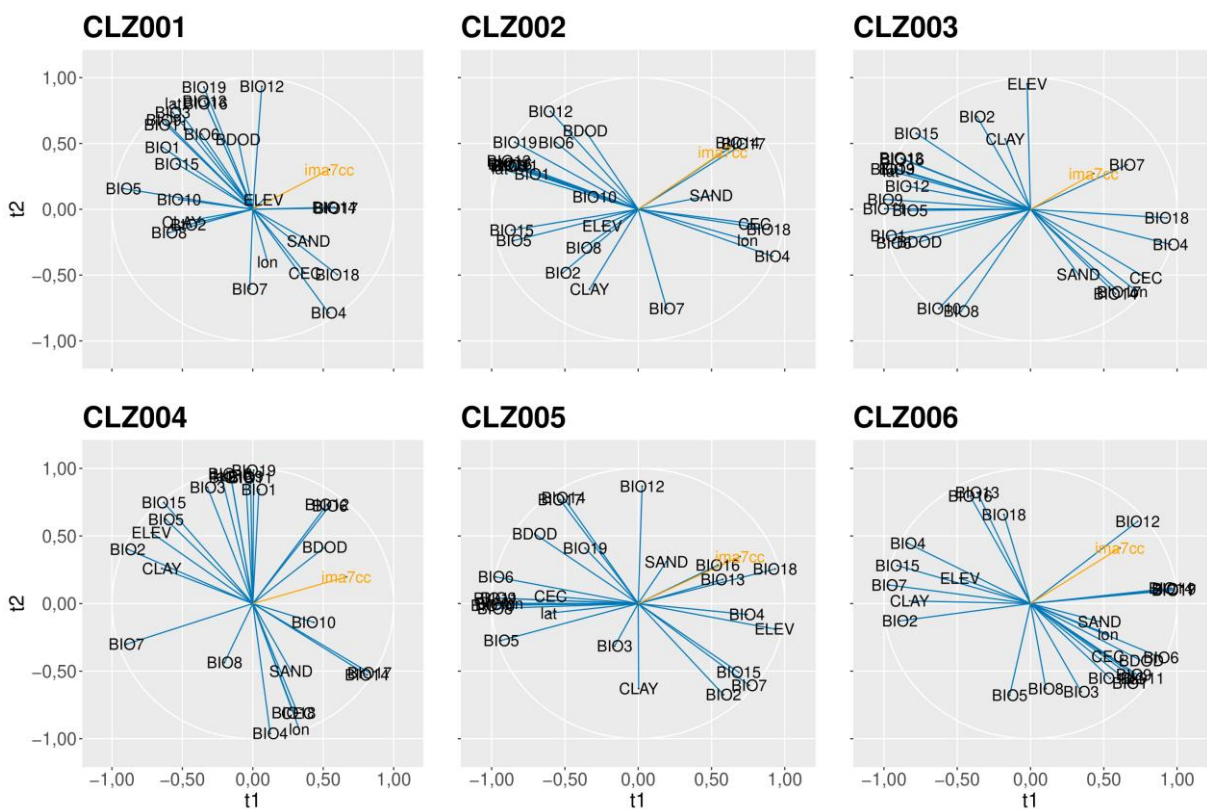
Fonte: Do autor (2022)

4.4. Análises dos efeitos das covariáveis ambientais na produtividade dos clones

A avaliação do efeito de cada covariável ambiental na produção de madeira pode ser realizada a partir da interpretação de gráficos biplot, que mostram as correlações de cada variável com as duas primeiras variáveis latentes obtidas a partir da modelagem PLS (Figura 12). Os gráficos conhecidos como círculos de correlação ou *biplots*, permitem a interpretação da relação entre as variáveis preditoras, bem como delas com a variável resposta (IMA). Nesse gráfico, os ângulos entre os vetores de cada variável indicam suas correlações. Vetores

com angulação entre si superior a 90° indicam uma correlação negativa entre as variáveis, enquanto angulação inferior a 90° indica uma correlação positiva. Já ângulos de 90° indicam correlação nula. Como esperado, devido à interação GxA, os clones apresentaram diferentes respostas em relação às covariáveis ambientais. Isso ocorre devido a diferenças nas constituições genéticas dos clones e a diferentes combinações de ambientes em que cada clone está submetido.

Figura 12 – Círculos de correlação entre as covariáveis ambientais e a variável resposta (IMA) para cada um dos seis clones do estudo.



Legenda: t1= primeira variável latente; t2= segunda variável latente.

Fonte: Do autor (2022)

O grande número de covariáveis ambientais utilizadas dificulta as inferências acerca da influência de cada uma delas sobre a produtividade, diretamente pelo *biplot* da Figura 12. Para a definição quantitativa das cinco variáveis de maior influência na produtividade dos clones de eucalipto foram comparados os coeficientes de regressão normalizados (Tabela 4). Conforme esperado, há uma tendência de que as variáveis climáticas relacionadas à perda de água (temperatura) apresentem influências negativas enquanto as variáveis relacionadas ao ganho de água (precipitação) apresentem influências positivas.

Três biovariáveis relacionadas à precipitação pluviométrica se destacaram estando presentes entre as covariáveis mais influentes no IMA de pelo menos quatro clones: precipitação anual (BIO12), precipitação do mês mais seco (BIO14), precipitação do trimestre mais seco (BIO17). Covariáveis relacionadas à temperatura também apresentaram destaque em pelo menos três clones: temperatura média do trimestre mais úmido (BIO8) e temperatura máxima do mês mais quente (BIO5). Essas covariáveis estão relacionadas à causa do estresse hídrico em diferentes culturas florestais e agrônomicas.

Tabela 4 – Coeficientes de regressão normalizados das cinco covariáveis de maior influência na produtividade de cada um dos seis clones avaliados.

Posição	CLZ001		CLZ002		CLZ003	
	Variável	Valor	Variável	Valor	Variável	Valor
1°	BIO5	-0,1498	BIO8	-0,1468	BIO8	-0,1155
2°	BIO12	0,1389	BIO5	-0,1268	BIO10	-0,0903
3°	BIO8	-0,1387	BIO17	0,1241	ELEVAÇÃO	0,0691
4°	BIO17	0,1022	BIO14	0,1230	BIO1	-0,0552
5°	BIO14	0,0989	ARGILA	-0,0999	AREIA	-0,0465

Posição	CLZ004		CLZ005		CLZ006	
	Variável	Valor	Variável	Valor	Variável	Valor
1°	BIO7	-0,1265	BIO12	0,1073	BIO12	0,1417
2°	BIO12	0,1261	BIO5	-0,0954	BIO2	-0,1053
3°	BIO2	-0,1056	BIO18	0,0942	BIO14	0,1015
4°	BIO17	0,0918	ARGILA	-0,0749	BIO19	0,0980
5°	BIO14	0,0914	BIO10	-0,0708	BIO17	0,0968

BIO5= Temperatura máxima do mês mais quente; BIO8= Temperatura média do trimestre mais úmido; BIO12= Precipitação anual; BIO14 = precipitação do mês mais seco; BIO17= precipitação do trimestre mais seco; BIO1= Temperatura média anual; BIO2= Amplitude térmica diária; BIO7= Amplitude térmica anual; BIO10= Temperatura média do trimestre mais quente; BIO18= Precipitação do trimestre mais quente; BIO19= Precipitação do trimestre mais frio.

O clone CLZ003, foi o único que apresentou uma covariável geográfica (Elevação) entre as cinco covariáveis mais influentes no seu crescimento. Esse resultando, não necessariamente, indica a ausência de influência das demais covariáveis importantes no clone em questão ou da elevação nos demais clones.

A partir da análise das correlações entre as covariáveis mais importantes com a elevação é possível verificar se há alguma influência indireta da covariável geográfica no desempenho de cada clone. Assim, uma vez que a elevação representa uma covariável de

fácil obtenção, foram obtidas as correlações par a par das cinco covariáveis mais importantes com essa covariável geográfica (Figura 13).

Considerando apenas as correlações entre a elevação e as biovariáveis selecionadas, os maiores valores foram obtidos para precipitação do mês mais seco ($r = -0,672$, p-valor $< 0,001$) e precipitação do trimestre mais seco ($r = -0,671$, p-valor $< 0,001$). Com exceção do CLZ006, todos os clones apresentaram correlações negativas moderadas ($0,5 < |r| < 0,7$) e altas ($0,7 < |r| < 0,9$) para as duas covariáveis. Além disso, é possível verificar que, de maneira geral, as duas biovariáveis apresentam altas correlações de Pearson entre si ($r \geq 0,98$), sendo significativas para todos os clones. Dada a elevada magnitude da correlação entre essas variáveis, seria possível considerar apenas uma destas variáveis nas futuras análises a serem realizadas.

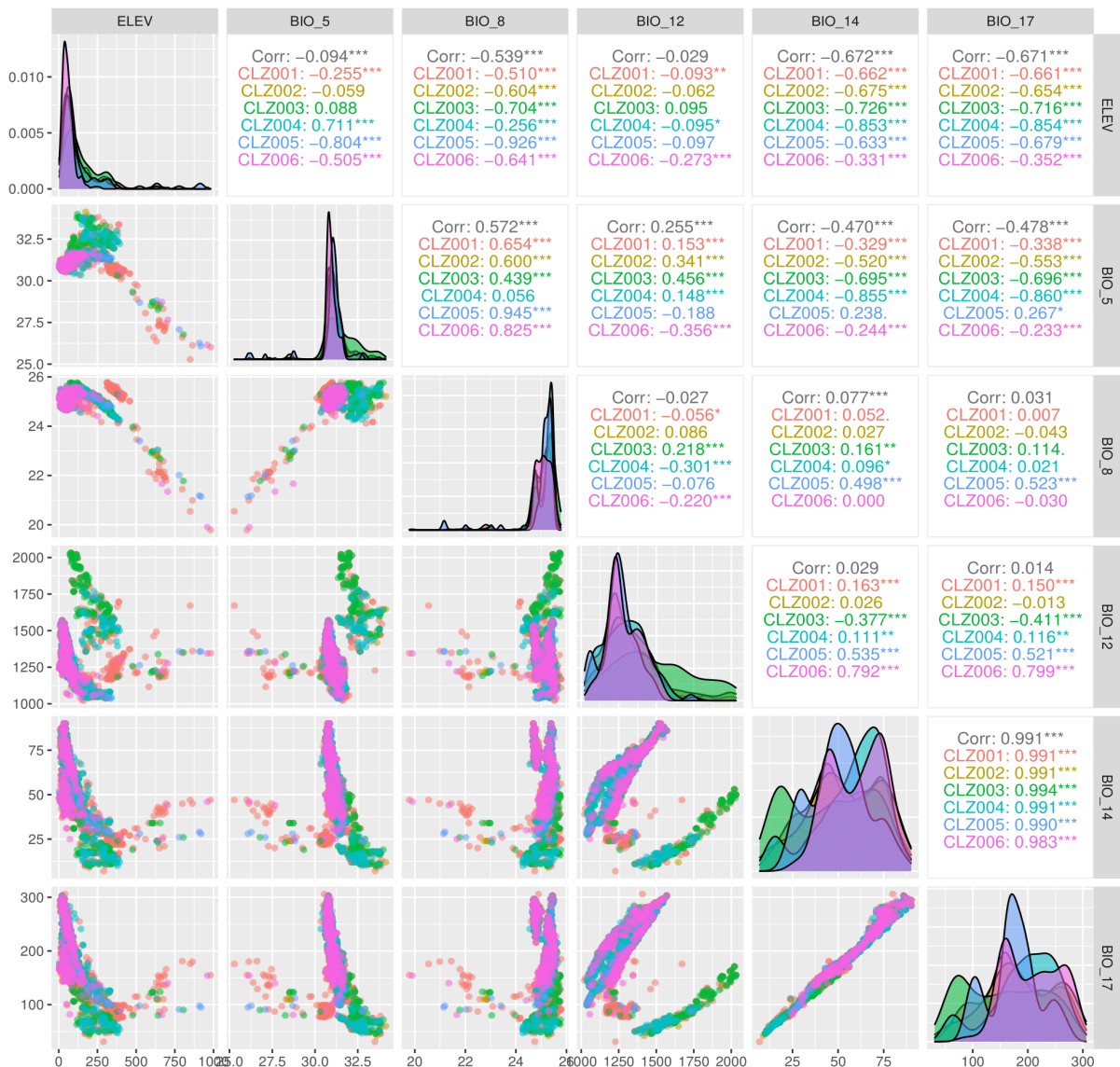
Para a temperatura máxima do mês mais quente (BIO5), a elevação apresentou correlações negativas de alta magnitude para o clone CLZ005 ($r = -0,804$, p-valor $< 0,001$), desprezível ($0 < |r| < 0,3$) para o clone CLZ001 ($r = -0,255$, p-valor $< 0,001$) e não significativa para o clone CLZ002 ($r = -0,059$). Já a temperatura média do trimestre mais úmido (BIO8) apresenta correlações negativas moderadas para os clones CLZ001 ($r = -0,501$, p-valor $< 0,001$), CLZ002 ($r = -0,604$, p-valor $< 0,001$) e CLZ003 ($r = -0,704$, p-valor $< 0,001$). A precipitação anual (BIO12) apresentou valores desprezíveis para todos os clones que apresentam a covariável em seu ranking.

Essas correlações indicam que a elevação exerce influência nas predições de IMA mesmo dos clones em que essa covariável não aparece entre as cinco de maior influência. Considerando a magnitude das correlações da elevação com as covariáveis BIO5 (temperatura máxima do mês mais quente), BIO14 (precipitação do mês mais seco) e BIO17 (precipitação do trimestre mais seco), infere-se que estas também representam alguma importância na produtividade do clone CLZ003. Além disso, esse clone apresenta a maior proporção de talhões situados na unidade do Maranhão (MA) e é possível que as características da região influenciem no ranqueamento das covariáveis mais importantes para o genótipo. Assim, a investigação acerca da influência da elevação nas covariáveis mais importantes também foi realizada a nível de unidade florestal (Figura 14).

Para BIO5 a unidade MA foi a única que não apresentou correlação significativa com a elevação e apresentou um valor desprezível para BA ($0,285$, p-valor $< 0,001$), fraco ($0,3 < |r| < 0,5$) para MS ($-0,482$, p-valor $< 0,001$) e alto para SP ($-0,813$, p-valor $< 0,001$). Já para

BIO8, todas as correlações foram significativas e negativas, com um valor desprezível para BA (-0,179, p-valor <0,001), altos para MS (-0,879, p-valor < 0,001) e SP (-0,894, p-valor <0,001), e muito alto para MA (-0,981, p-valor < 0,001) (Figura 14). É importante notar que BIO8 foi a biovariável que apresentou a maior influência na produtividade do CLZ003 (Tabela 4).

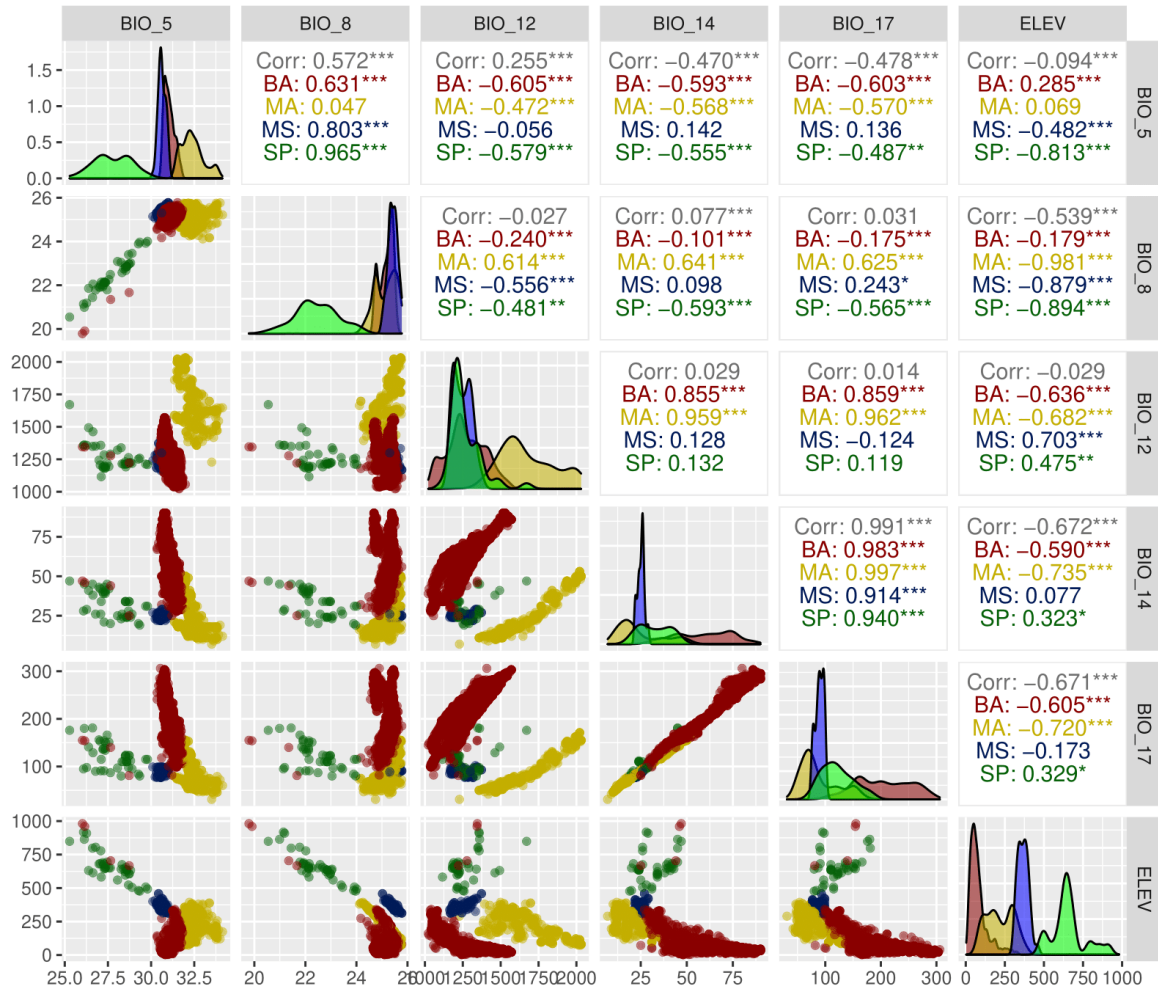
Figura 13 – Correlações par a par das cinco covariáveis mais influentes na produtividade dos clones (BIO12, BIO14, BIO17, BIO5 e BIO8) e elevação (ELEV) por clone. Na primeira linha, os valores de correlação geral e separado por clone. Na última linha, os gráficos de dispersão (pontos coloridos por clone).



Legenda: BIO5= Temperatura máxima do mês mais quente; BIO8= Temperatura média do trimestre mais úmido; BIO12= Precipitação anual; BIO14 = precipitação do mês mais seco; BIO17= precipitação do trimestre mais seco; ELEV = Elevação.

Fonte: Do autor (2022)

Figura 14 – Correlações par a par das cinco covariáveis mais influentes na produtividade dos clones (BIO12, BIO14, BIO17, BIO5 e BIO8) e elevação (ELEV) por unidade florestal. Na primeira linha, os valores de correlação geral e separado por unidade florestal. Na última linha, os gráficos de dispersão (pontos coloridos por unidade florestal).



Legenda: BIO5= Temperatura máxima do mês mais quente; BIO8= Temperatura média do trimestre mais úmido; BIO12= Precipitação anual; BIO14 = precipitação do mês mais seco; BIO17= precipitação do trimestre mais seco; ELEV = Elevação.

Fonte: Do autor (2022)

Em todas as unidades florestais a BIO12 apresentou estar correlacionada com a elevação de forma significativa a pelo menos 1% de significância. Entretanto, para MA e BA foram obtidas correlações negativas (-0,636 e -0,682, respectivamente) enquanto MS e SP apresentam correlações positivas (0,703 e 0,475, respectivamente). Inversões na direção das correlações (positiva ou negativa) também ocorrem nas biovariáveis BIO14 e BIO17, também relacionadas com a precipitação. Para essas covariáveis, a unidade MA apresenta as correlações de maior magnitude, indicando que a precipitação no período seco da região sofre o maior declínio com o aumento da altitude. Esses resultados, reforçam as condições distintas

da unidade MA das demais regiões, além da maior influência da elevação nas condições relacionadas ao estresse hídrico.

De forma geral, as diferenças de magnitude e até de sinal nas correlações entre as covariáveis nas diferentes unidades florestais, reforçam a complexidade dos efeitos ambientais e das inter-relações entre os seus componentes (covariáveis ambientais). O entendimento das relações entre os componentes da variação ambiental pode ser importante para melhorar a compreensão e modelagem da interação GxA.

4.4.1. Mapeamento da produtividade predita dos clones

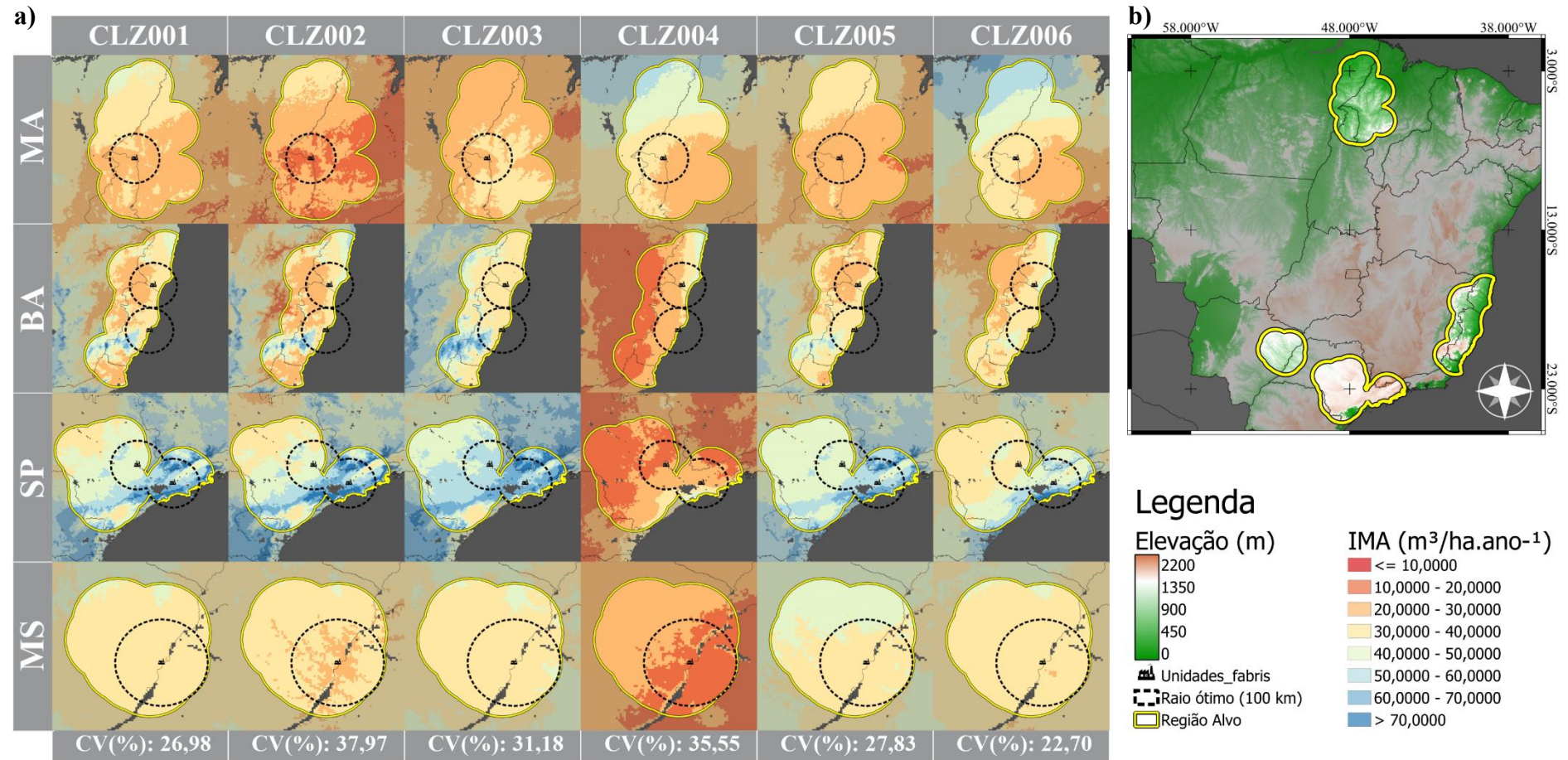
Para uma interpretação mais acurada dos valores de produtividade, as predições a partir do modelo PLS foram realizadas apenas nas áreas que compreendem raios de 100km dos talhões inventariados. Essa restrição no mapeamento das predições visa evitar extrapolações com um conjunto de covariáveis de ambientes muito distantes daqueles testados, uma vez que isso pode resultar em predições irreais. Os mapas de produtividade e os modelos preditivos utilizados encontram-se na Figura 15.a e Tabela 1 do Apêndice A, respectivamente.

A partir da análise visual dos mapas é possível inferir que a unidade florestal de São Paulo apresentou a maior proporção de área contendo as classes mais altas de produtividade, enquanto a unidade do Maranhão apresentou o comportamento inverso. Além disso, é possível notar que os maiores volumes de IMA na unidade da Bahia foram obtidas nas regiões próximas da faixa litorânea e no Sul do Espírito Santo. Quanto à unidade de Mato Grosso do Sul, houve uma tendência de aumento de IMA com o aumento da distância do rio Paraná.

Esses padrões de produtividade nas diferentes unidades florestais apresentam alguma concordância com os padrões de elevação (Figura 15.b). Essas tendências são mais evidenciadas no Sul do Espírito Santo e em São Paulo, onde há as áreas de maior altitude dentre os ambientes alvo e apresentam os maiores valores de IMA.

Ao contrário dos demais clones, o CLZ004 não apresentou maior produtividade na unidade de São Paulo, mas sim na unidade do Maranhão. Já o clone CLZ006, apresentou bons níveis de produtividade predita em todo mapa e a menor variação (CV = 22,70%). Portanto, esse é o clone de maior estabilidade. O clone CLZ002 (CV = 37,97%) apresentou a menor estabilidade ao longo dos ambientes.

Figura 15 – Adaptabilidade mapeada. **a)** Produtividade por classe para cada clone em cada unidade florestal. **b)** Elevação do Brasil. Produtividades previstas com covariáveis ambientais utilizando os modelos ajustados com a regressão PLS.

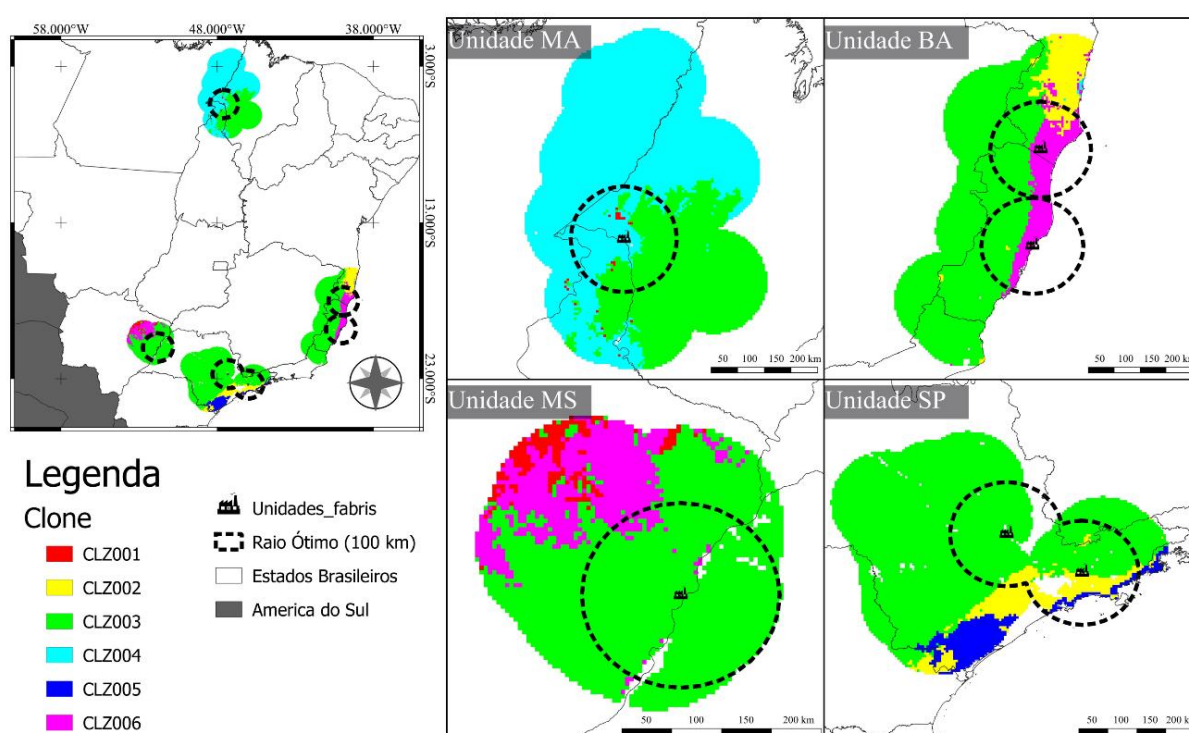


Fonte: Do autor (2022)

4.4.2. Mapas dos clones vencedores

Para esses mapas, foram comparados apenas clones cuja produtividade foi observada dentro de cada unidade florestal (Tabela 1). Essa restrição visa evitar que clones com alto desempenho em locais testados apresentem um desempenho predito irreal em locais onde não foram testados e, potencialmente, com condições edafoclimáticas muito distintas. Os mapas dos clones vencedores encontram-se na Figura 16.

Figura 16 – Mapas dos clones vencedores a partir de valores de produtividades previstas com modelos PLS utilizando covariáveis ambientais.



Fonte: Do autor (2022)

As proporções de áreas de recomendação para cada clone encontram-se na Tabela 5. Os clones mais recomendados por produtividade foram CLZ003 (SP: 78,94%; BA: 70,21%; MS: 47,60%) e CLZ004 (MA: 63,79%). Os padrões de recomendação se mantiveram no raio ótimo de transporte até a fábrica.

Embora os valores preditos representem a combinação linear dos efeitos de cada uma das 26 covariáveis, aquelas mais importantes desempenharam um papel determinante na predição de produtividade dos clones e, conseqüentemente, nos mapas do genótipo vencedor. Esses padrões de recomendação são mais evidentes em alguns clones. Por exemplo, para os clones CLZ004 e CLZ006, as regiões de maior precipitação pluviométrica foram

determinantes (Figura 2), enquanto as regiões mais vantajosas para o clone CLZ003, compreendem as regiões de maiores altitudes (Figura 15.b).

Tabela 5 – Clones vencedores e seus percentuais de área de adaptação produtiva superior aos demais clones nas regiões alvo.

Unidade Florestal	Clone	Todos os clones		Sem o CLZ003	
		Área Total	Raio ótimo*	Área Total	Raio ótimo*
BA	CLZ001	-	-	19,80%	6,00%
	CLZ002	13,57%	10,06%	21,44%	13,93%
	CLZ003	70,21%	48,54%	x	x
	CLZ004	0,01%	-	0,18%	-
	CLZ005	-	-	29,99%	9,72%
	CLZ006	16,20%	41,40%	28,59%	70,35%
SP	CLZ001	-	-	5,73%	13,05%
	CLZ002	12,97%	17,22%	36,92%	51,87%
	CLZ003	78,94%	79,35%	x	x
	CLZ005	8,08%	3,43%	57,35%	35,08%
MS	CLZ001	23,07%	5,64%	50,18%	63,99%
	CLZ003	47,60%	71,88%	x	x
	CLZ006	29,33%	22,48%	49,82%	36,01%
MA	CLZ001	0,16%	1,10%	21,64%	25,95%
	CLZ002	-	-	-	-
	CLZ003	36,05%	52,86%	x	x
	CLZ004	63,79%	46,04%	77,29%	72,06%
	CLZ005	-	-	1,06%	2,00%

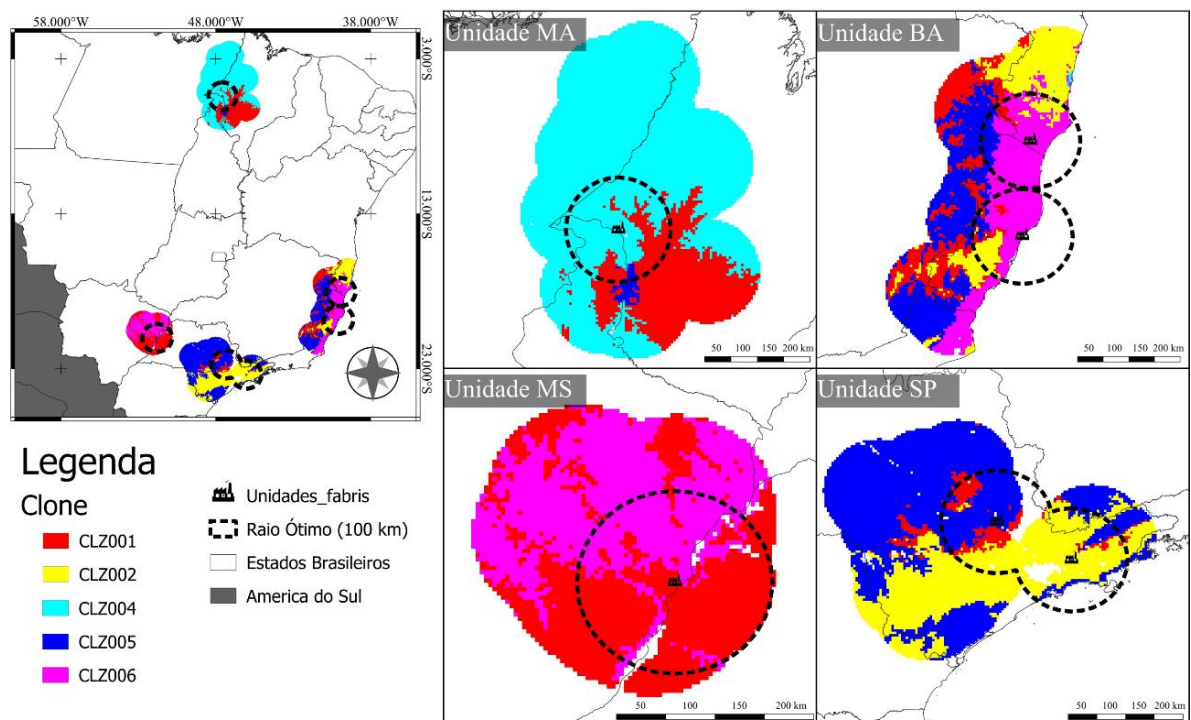
* Raio de 100km (SILVA et al., 2007)

É pouco provável que o clone CLZ003 apresente vantagem tão grande para tantas regiões distintas, uma vez que este genótipo não apresentou superioridade quanto à produtividade média em quaisquer das unidades florestais nos dados de inventário. A ampla recomendação deste clone pode estar fortemente influenciada pela baixa capacidade preditiva do modelo específico ajustado ($R^2 = 0,23$, Figura 10). Assim, foi elaborado um novo mapa do genótipo vencedor sem o clone CLZ003 (Figura 17). Nesse novo mapa, todos os clones apresentaram aumento na área de recomendação, com destaque para o clone mais plantado (CLZ001), que apresentou “pixels” em todas as regiões e o clone CLZ005 que agora apresenta uma ampla recomendação nas unidades de São Paulo e Bahia.

Em relação ao mapa anterior, na Figura 17 as proporções de recomendação para área total foram mais consistentes com as médias de produtividade dos dados de inventário (Figura

7). Entretanto, para regiões em que o CLZ001 apresentou a superioridade média (SP, MA e MS), houve também ampla recomendação dos segundos colocados, que apresentam pequenas diferenças médias de produtividade com o melhor clone dentro das unidades. Para a unidade da Bahia, o clone CLZ006 apresentou umas das maiores recomendações na área total e a maior para os raios ótimos de transporte até as fábricas.

Figura 17 – Mapas dos clones vencedores a partir de valores de produtividade predita excluindo o clone CLZ003. As produtividades foram preditas com as covariáveis ambientais utilizando o modelo PLS.



Fonte: Do autor (2022)

5. DISCUSSÃO

A interação GxA impõem grandes dificuldades na recomendação de cultivares. Com isso, há demanda por uma maior compreensão da influência de fatores ambientais na produtividade de diversas culturas. Nesse sentido, a partir da utilização de bancos de dados edafoclimáticos e geográficos foram ajustados modelos preditivos para o mapeamento da adaptabilidade de seis clones utilizados em escala comercial. A partir desses modelos, foi possível verificar que a influência de cada covariável ambiental difere entre clones distintos de *Eucalyptus*. Essa variação é, provavelmente, uma das causas da interação GxA.

Um das dificuldades enfrentadas para estudos nesse aspecto é a escolha da melhor fonte de covariáveis edafoclimáticas para o ajuste dos modelos. O melhor banco variou entre os clones quanto ao R^2 . Porém, nesse estudo, o WorldClim foi considerado o banco mais adequado para todos os genótipos avaliados, considerando o erro predito médio (RMSE) e, portanto, escolhido para as predições. Essa vantagem pode ter ocorrido devido à imprevisibilidade de condições climáticas futuras, tornando a utilização de normais climatológicas estimadas com dados de longo prazo mais adequada para a realização das predições. Além disso, é possível que a qualidade da modelagem entre os clones tenha sido afetada pela similaridade das normais climatológicas com as condições climáticas dos anos de desenvolvimento das florestas (ANNICCHIARICO; BELLAH; CHIARI, 2006).

A caracterização ambiental e do inventário revelou padrões importantes para a modelagem da produtividade de madeira (Figura 8). A partir das similaridades ambientais entre as unidades florestais, evidenciou-se que a região mais distinta das demais é a unidade do Maranhão, onde predominam maiores temperaturas e maior sazonalidade de precipitação pluviométrica ao longo do ano. Essas características são os principais fatores da elevada incidência de estresse hídrico e térmico na região (CAMPOE et al., 2020; ELLI et al., 2020). Nessas condições, a produção de madeira é prejudicada devido ao fechamento de estômatos resultando em menores taxas fotossintéticas (MARTIN-STPAUL; DELZON; COCHARD, 2017). Além disso, os produtos fotossintéticos já reduzidos tendem a ser alocados prioritariamente para as raízes, visando a exploração de recursos hídricos em camadas mais profundas do solo (CAMPOE et al., 2020).

As condições de estresse tendem a ser amenizadas em maiores latitudes e altitudes. A cada aumento de 1° de latitude, há aumento médio da produção madeireira do eucalipto em $0,91 \text{ m}^3 \cdot \text{ha}^{-1} \cdot \text{ano}^{-1}$ (ELLI et al., 2020). A unidade do Maranhão é a região mais próxima da

linha do Equador e está situada cerca de 10° de latitude de distância da outra unidade mais próxima, que é a da Bahia. As unidades de Mato Grosso do Sul e de São Paulo apresentaram a maior parte de seus talhões em faixa de maiores latitudes (22°S – 30° S), o que resulta em maior produção. A unidade da Bahia, por sua vez, está situada em uma faixa de latitude com presença de estresse hídrico (11° S – 21°S). Essa unidade apresenta a maior proporção de talhões inventariados (~83%) e a maior variabilidade quanto à produtividade média. A variação observada na região pode ser atribuída ao grande número de combinações de fatores geográficos, edafoclimáticos e temporais. Além disso, sabe-se que na região ocorre o distúrbio fisiológico do *Eucalyptus*, cuja incidência depende do genótipo e da localização geográfica do plantio. Alguns dos talhões de menor produtividade na Bahia podem ter sofrido com esse distúrbio.

O maior volume de dados para Bahia favoreceu a modelagem para a região (Figura 11). A qualidade da modelagem seguiu o grau de similaridade com a unidade, em que, a unidade de Mato Grosso do Sul apresentou a segunda melhor modelagem, seguida de São Paulo e do Maranhão. Dessa forma, a capacidade preditiva dos modelos ajustados para cada clone foi afetada pela proporção de talhões inventariados presentes nas regiões favorecidas. O clone CLZ003 apresentou, além da pior capacidade preditiva, a menor proporção de dados na região da Bahia (50,2%), enquanto o comportamento inverso foi notado para o CLZ006, que apresenta cerca de 99,7% dos dados na unidade florestal. Assim, é natural que os fatores ambientais importantes também sigam as proporções de talhões inventariados em cada unidade florestal. Dentre as variáveis de maior influência, aquelas relacionadas com a precipitação apresentaram as maiores frequências nos ranqueamentos de coeficientes de regressão normalizados para os seis clones (Tabela 4). Na Bahia, a precipitação anual (BIO12) é um dos recursos determinantes para o desenvolvimento vegetal e exerce um papel importante no fornecimento de água para os sistemas produtivos (STAPE; BINKLEY; RYAN, 2004; MARCATTI et al., 2017; SCOLFORO et al., 2017). Nesse sentido, os quatro clones com maiores proporções de talhões situados na unidade foram os mais afetados pela precipitação pluviométrica anual.

Conforme esperado, os clones apresentaram diferenças quanto à magnitude da resposta à precipitação pluviométrica. Esse fenômeno também foi observado por Scolforo et al. (2017), na mesma região do estudo, ao incluir covariáveis relacionadas com a precipitação em um modelo preditivo para altura dominante. No estudo, a maior parte dos clones tiveram preferência pela regularidade da chuva ao longo do ano, ao invés da pluviometria total,

implicando em uma maior sensibilidade a períodos secos na região. Essa sensibilidade é evidente em regiões tropicais, onde as condições de estresse são mais pronunciadas, para as regiões subtropicais que possuem uma distribuição de chuvas uniformes, a variação de temperatura é o fator mais limitante para a produção (CAMPOE et al., 2016; ELLI et al., 2020). Dessa forma, maiores volumes de precipitação no mês mais seco (BIO14) e no trimestre mais seco (BIO17) atuam atenuando o estresse hídrico. Em contrapartida, a temperatura máxima do mês mais quente (BIO5) e a temperatura média do mês mais úmido (BIO8) atuam na severidade do estresse hídrico e térmico.

Projetos amplos de investigação da influência ambiental na produtividade do eucalipto no Brasil, como o TECHS (Tolerância de Eucalyptus Clonais aos Estresses Hídrico e Térmico) e o BEPP (Brasil Eucalyptus Produtividade Potencial), indicam que a temperatura desempenha a maior influência na produtividade de madeira (BINKLEY et al., 2017, 2020; RYAN et al., 2020). Esse fenômeno é atribuído a fortes indícios de inflação do efeito da covariável devido à existência de interação com a precipitação (RYAN et al., 2020). Além disso, embora a temperatura afete o desenvolvimento vegetal pelo aumento das taxas de respiração e redução de taxas fotossintéticas (ASPINWALL et al., 2016), a atuação da covariável sobre a produção é mais importante devido à sua influência indireta na umidade do solo (LIM et al., 2020; RYAN et al., 2020).

Quanto às demais variáveis, apenas um clone apresentou uma variável geográfica (altitude) entre as mais importantes preditoras do IMA (Tabela 4). Entretanto, foi demonstrado que há influência indireta de outras biovariáveis (BIO8, BIO17 e BIO14) através da correlação com a elevação, assim como o inverso nos demais genótipos (Figura 13). A correlação negativa da temperatura com a elevação já é conhecida e utilizada para estimativas de covariáveis e classificações climáticas (ALVARES et al., 2013; FICK; HIJMANS, 2017). A influência indireta da elevação sobre a precipitação (BIO12, BIO14 e BIO17), por sua vez, apresentou direções distintas para as unidades situadas em menores latitudes (Maranhão e Bahia) e para as situadas em maiores latitudes (São Paulo e Mato Grosso do Sul) (Figura 14). Essa diferença se dá principalmente devido ao padrão pluviométrico da região nordeste, onde os maiores índices e regularidade da chuva estão no litoral da região. Dessa forma, com o aumento da distância do mar e, conseqüentemente, da elevação, há um decréscimo na pluviometria. Naturalmente, esse gradiente associado a maiores temperaturas resulta em um baixo estoque de água no solo, na maior parte do ano no interior do nordeste (APARECIDO et al., 2021).

Essa tendência na variação do IMA também foi observada nos mapas de produtividade (Figura 15). As regiões costeiras tropicais e as regiões subtropicais de maior altitude apresentaram as maiores produtividades, reforçando as relações do fornecimento de água e temperatura na produção de madeira. Nesse sentido, as maiores produtividades em São Paulo podem ser atribuídas a distribuição pluviométrica uniforme ao longo do ano e a temperaturas próximas do ótimo produtivo para o eucalipto (20 – 22 °C) (BINKLEY et al., 2020). Apenas o CLZ004 apresentou comportamento inverso dos demais, ou seja, com melhor desempenho na unidade do Maranhão. Isso pode ter ocorrido devido à ausência de talhões na região subtropical, levando a estimativas errôneas, uma vez que as influências ambientais captadas são baseadas nos padrões de variação ambiental das regiões tropicais, onde a precipitação se comporta de maneira distinta das regiões subtropicais (Figura 14).

Além disso, o clone CLZ006 apresentou altas produtividades em todo mapeamento. Novamente, esse padrão de resposta deve ser observado com cautela, uma vez que o clone também não foi observado em todas as unidades florestais. Uma das razões para isso é a consideração da média geral para o ajuste do modelo na regressão PLS (Equação 2). Dessa forma, é possível que a não representatividade do clone na região menos produtiva (MA) tenha interferido em sua média geral, beneficiando suas estimativas para todas as demais regiões. Logo, para clones restritos, os mapas de recomendação devem ser interpretados apenas a nível regional ou de tipos climáticos semelhantes.

Para evitar comparações “desleais”, os mapas do clone vencedor foram elaborados por unidade florestal considerando apenas os genótipos plantados dentro de cada região. Na primeira versão (Figura 16), o clone CLZ003 apresentou a maior área de recomendação para três das quatro unidades florestais, mesmo não tendo a sua superioridade observada em nenhuma delas com os dados de inventário. Esse clone em questão, possui o modelo com a pior capacidade preditiva devido à quantidade majoritária de talhões na unidade mais distinta (Maranhão). Dessa forma, os erros preditivos muito elevados ($7,575 \text{ m}^3 \cdot \text{ha}^{-1} \cdot \text{ano}^{-1}$) podem ter resultado essa falsa vantagem do clone CLZ003. Logo, com sua exclusão, obteve-se um mapa de recomendação mais acurado (Figura 17).

A maior consistência da nova versão pode ser atribuída aos menores erros produtivos médios para os clones com modelos mais acurados, assim, as produtividades estimadas são, em média, mais próximas das observadas. Além disso, a ampla recomendação dos segundos colocados para as unidades do Mato Grosso do Sul, São Paulo e Maranhão foi favorecida

visto que as diferenças produtivas são pequenas ($< 0.51 \text{ m}^3.\text{ha}^{-1}.\text{ano}^{-1}$). Assim, as proporções de recomendação foram semelhantes ao ranqueamento das produtividades médias nos dados de inventário (Figura 7). Essas tendências são mais claras no raio ótimo econômico, onde há a maior proporção de talhões inventariados (~86%) e, portanto, uma maior acurácia das predições.

As influências das covariáveis mais importantes foram determinantes para a recomendação dos clones. Consequentemente, investigações da influência ambiental na produção de madeira de eucalipto podem fornecer pistas acerca dos efeitos das mudanças climáticas. Segundo o IPCC (Painel Intergovernamental para a Mudança de Clima), espera-se que ocorra uma maior frequência de temperaturas médias elevadas e dias secos no Brasil (GUTIÉRREZ et al., 2021). As previsões indicam que as condições de estresse hídrico e térmico podem se agravar nas regiões em que os fenômenos já são observados e surgir em novas. Logo, esse cenário indica mudanças no padrão de recomendação e a necessidade de adaptação de tratamentos culturais, uma vez que os genótipos avaliados apresentaram alta sensibilidade a covariáveis relacionadas ao estresse hídrico e térmico (BOOTH, 2013; SCOLFARO et al., 2017).

O mapeamento da adaptabilidade forneceu a compreensão da resposta dos diferentes clones comerciais frente a variação ambiental no Brasil. Assim, os mapas de recomendação representam uma ferramenta útil para a seleção de genótipos que melhor capitalizem em seus fenótipos os efeitos positivos de ambiente e interação GxA na produção de madeira.

6. CONCLUSÃO

A modelagem via Quadrados Mínimos Parciais (PLS) revelou padrões de resposta dos clones avaliados frente a variação ambiental e a interação GxA. Para ambientes tropicais, as maiores produtividades foram alcançadas no litoral, enquanto para ambientes subtropicais as maiores elevações foram determinantes. Assim, a variação ambiental atua de maneira distinta na produção de madeira de eucalipto ao longo das unidades florestais. Entretanto o maior volume de dados de inventário para a unidade da Bahia, influenciou fortemente a qualidade da modelagem. Dessa forma, os modelos ajustados para os genótipos com maior proporção de observações na unidade, apresentaram as melhores capacidades preditivas.

As covariáveis que mais afetaram o desempenho dos clones foram a precipitação pluviométrica anual (BIO12), precipitação pluviométrica do mês mais seco (BIO14), precipitação pluviométrica do trimestre mais seco (BIO17), temperatura máxima do mês mais quente (BIO5) e temperatura média do trimestre mais úmido (BIO8). Esse resultado indica que, a produtividade de madeira é fortemente influenciada pelo estresse hídrico e térmico. Logo, em um panorama de mudanças climáticas, espera-se um declínio no desempenho dos genótipos.

REFERÊNCIAS

- ABDI, H. Partial least squares regression and projection on latent structure regression (PLS Regression). **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 2, n. 1, p. 97–106, 2010. doi: 10.1002/wics.51.
- ALVARES, C. A. *et al.* Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711–728, 2013. doi: 10.1127/0941-2948/2013/0507.
- ANNICCHIARICO, P.; BELLAH, F.; CHIARI, T. Repeatable genotype x location interaction and its exploitation by conventional and GIS-based cultivar recommendation for durum wheat in Algeria. **European Journal of Agronomy**, v. 24, n. 1, p. 70–81, 2006. doi: 10.1016/j.eja.2005.05.003.
- APARECIDO, L. E de O. *et al.* Soil water seasonal and spatial variability in Northeast Brazil. **Environment, Development and Sustainability**, 2021. doi: 10.1007/s10668-021-01695-4.
- ARAÚJO, M. J. de *et al.* Adaptability and stability of eucalypt clones at different ages across environmental gradients in Brazil. **Forest Ecology and Management**, v. 454, n. June, p. 117631, 2019. doi: 10.1016/j.foreco.2019.117631.
- ASPINWALL, M. J. *et al.* Convergent acclimation of leaf photosynthesis and respiration to prevailing ambient temperatures under current and warmer climates in *Eucalyptus tereticornis*. **The New phytologist**, v. 212, n. 2, p. 354–367, 1 out. 2016. doi: 10.1111/nph.14035.
- ASSIS, T. F.; MAFIA, R. G. Hibridação e clonagem. In: BORÉM, A. (Ed.). **Biotecnologia florestal** Viçosa, MG: Suprema, 2007. p.93-121.
- AUGUIE, B.; ANTONOV, A. Miscellaneous Functions for “Grid” Graphics [R package gridExtra version 2.3]. **CRAN Repository**, 2017.
- BATTAGLIA, M. *et al.* CABALA: A linked carbon, water and nitrogen model of forest growth for silvicultural decision support. **Forest Ecology and Management**, v. 193, n. 1–2, 2004. doi: 10.1016/j.foreco.2004.01.033.
- BINKLEY, D. *et al.* The interactions of climate, spacing and genetics on clonal eucalyptus plantations across Brazil and Uruguay. **Forest Ecology and Management**, v. 405, n. July, p. 271–283, 2017. doi: 10.1016/j.foreco.2017.09.050.
- BINKLEY, D. *et al.* Variation in whole-rotation yield among eucalyptus genotypes in response to water and heat stresses: The TECHS project. **Forest Ecology and Management**, v. 462, 15 abr. 2020. doi: 10.1016/j.foreco.2020.117953.
- BIVAND, R.; KEITT, T.; ROWLINGSON, B. Package ‘rgdal’ - Bindings for the “Geospatial” Data Abstraction Library. **CRAN Repository**, 2021.
- BOOTH, T. H. Eucalypt plantations and climate change. **Forest Ecology and Management**, v. 301, p. 28–34, 1 ago. 2013. doi: 10.1016/j.foreco.2012.04.004.

- BOYDEN, S.; BINKLEY, D.; STAPE, J. L. Competition among eucalyptus trees depends on genetic variation and resource supply. **Ecology**, v. 89, n. 10, p. 2850–2859, 2008. doi: 10.1890/07-1733.1.
- BRASIL. Portaria no. 244, de 12 de novembro de 2015. **Diário Oficial da União**, 2015.
- CÂMARA, A. P. *et al.* Physiological disorders affecting dendrometric parameters and eucalyptus wood quality for pulping wood. **Cerne**, v. 24, n. 1, p. 27–34, 1 jan. 2018. doi: 10.1590/01047760201824012480.
- CAMPOE, O. C. *et al.* Climate and genotype influences on carbon fluxes and partitioning in *Eucalyptus* plantations. **Forest Ecology and Management**, v. 475, 1 nov. 2020. doi: 10.1016/j.foreco.2020.118445.
- CAMPOE, O. C. *et al.* Meteorological seasonality affecting individual tree growth in forest plantations in Brazil. **Forest Ecology and Management**, v. 380, p. 149–160, 15 nov. 2016. doi: 10.1016/j.foreco.2016.08.048.
- CANUTO, D. S. D. O. *et al.* De. Estabilidade e adaptabilidade em testes de progênies de *Myracrodruon urundeuva* sob quatro sistemas de plantio. **Cerne**, v. 22, n. 2, p. 171–180, 2016. doi: 10.1590/01047760201622021978.
- CARGNIN, A. *et al.* Interação entre genótipos e ambientes e implicações em ganhos com seleção em trigo. **Pesquisa Agropecuária Brasileira**, v. 41, n. 6, p. 987–993, 2006. doi: 10.1590/s0100-204x2006000600014.
- CARRASCAL, L. M.; GALVÁN, I.; GORDO, O. Partial least squares regression as an alternative to current regression methods used in ecology. **Oikos**, v. 118, n. 5, p. 681–690, maio 2009. doi: 10.1111/j.1600-0706.2008.16881.x.
- CARVALHO, I. R. *et al.* Biometric models and maize genetic breeding: A review. **Australian Journal of Crop Science**, v. 12, n. 11, p. 1796–1805, 2018. doi: 10.21475/ajcs.18.12.11. p792.
- CASTRO, C. A. de O. *et al.* Brief history of *Eucalyptus* breeding in Brazil under perspective of biometric advances. **Ciência Rural**, v. 46, n. 9, p. 1585–1593, 2016. doi: 10.1590/0103-8478cr20150645.
- CASTRO, C. A. de O. *et al.* Comportamento da interação genótipos por locais aos três e nove anos em clones de eucalipto. **Scientia Forestalis**, v. 46, n. 120, p. 594–605, 2018. doi: 10.18671/scifor.v46n120.08.
- CASTRO, C. A. de O. *et al.* Optimization of *Eucalyptus benthamii* progeny test based on Near-Infrared Spectroscopy approach and volumetric production. **Industrial Crops and Products**, v. 141, 1 dez. 2019. doi: 10.1016/j.indcrop.2019.111786.
- COOPER, M. *et al.* Predicting the future of plant breeding: Complementing empirical evaluation with genetic prediction. **Crop and Pasture Science**, v. 65, n. 4, p. 311–336, 2014. doi: 10.1071/CP14007.

- COSTA-NETO, G. *et al.*, R. EnvRtype: a software to interplay enviromics and quantitative genomics in agriculture. **G3: Genes, Genomes, Genetics**, v. 11, n. 4, 2021. doi: 10.1093/g3journal/jkab040.
- COSTA-NETO, G. M. F. *et al.* A novel GIS-based tool to reveal spatial trends in reaction norm: upland rice case study. **Euphytica**, v. 216, n. 3, 2020. doi: 10.1007/s10681-020-2573-4.
- COSTA-NETO, G. M. F. **Integrando covariáveis ambientais e mapas temáticos na análise da interação “genótipo x ambiente” em arroz de terras altas**. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Escola de Agronomia, Universidade Federal de Goiás, Goiânia, p.125. 2017.
- CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. 4. ed. Viçosa: Ed. UFV, 2012.
- DA SILVA, P. H. M. *et al.* Selecting for stable and productive families of *Eucalyptus urophylla* across a country-wide range of climates in Brazil. **Canadian Journal of Forest Research**, v. 49, n. 1, p. 87–95, 2019. doi: 10.1139/cjfr-2018-0052.
- DE OLIVEIRA, T. W. G. *et al.* Stability and adaptability for wood volume in the selection of *Eucalyptus saligna* in three environments. **Pesquisa Agropecuaria Brasileira**, v. 53, n. 5, p. 611–619, 2018. doi: 10.1590/S0100-204X2018000500010.
- DE SOUZA, B. M. *et al.* Genotype-by-environment interaction in *Corymbia citriodora* (Hook.) K.D. Hill, & L.A.S. Johnson progeny test in Luiz Antonio, Brazil. **Forest Ecology and Management**, v. 460, n. December 2019, p. 117855, 2020. doi: 10.1016/j.foreco.2019.117855.
- EBERHART, S. A.; RUSSELL, W. A. Stability Parameters for Comparing Varieties. **Crop Science**, v. 6, n. 1, p. 36–40, 1966. doi: 10.2135/cropsci1966.0011183x000600010011x.
- ELLI, E. F. *et al.* Gauging the effects of climate variability on Eucalyptus plantations productivity across Brazil: A process-based modelling approach. **Ecological Indicators**, v. 114, n. March, p. 106325, 2020. doi: 10.1016/j.ecolind.2020.106325.
- FERREIRA, G. W. D. *et al.* Temporal decomposition sampling and chemical characterization of eucalyptus harvest residues using NIR spectroscopy and chemometric methods. **Talanta**, v. 188, p. 168–177, 1 out. 2018. doi: 10.1016/j.talanta.2018.05.073.
- FERREIRA, M. A Aventura dos eucaliptos. In: SCHUMACHER, M. V.; VIERA, M (org.). **Silvicultura do Eucalipto no Brasil**. Santa Maria: Editora UFSM, 2016. p. 11–47. 2016.
- FICK, S. E.; HIJMANS, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. **International Journal of Climatology**, v. 37, n. 12, p. 4302–4315, 1 out. 2017. doi: 10.1002/joc.5086.
- GUTIÉRREZ, J.M *et al* (eds.). **Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change**. Cambridge: Cambridge University Press, 2021. Disponível em:< <http://interactive-atlas.ipcc.ch/>>. Acesso em: 10 de jan. de 2022.

HIJMANS, R. J. *et al.* dismo: Species distribution modeling. R version 1.1-4. **Reference Module in Life Sciences**, 2017.

HIJMANS, R. J. Package ‘raster’ - Geographic Data Analysis and Modeling. **CRAN Repository**, 2020.

HODGINS-DAVIS, A.; TOWNSEND, J. P. Evolving gene expression: from G to E to $G \times E$. **Trends in Ecology and Evolution**, v. 24, n. 12, p. 649–658, 2009. doi: 10.1016/j.tree.2009.06.011.

HOLZWORTH, D. *et al.* APSIM Next Generation: Overcoming challenges in modernising a farming systems model. **Environmental Modelling and Software**, v. 103, 2018. doi: 10.1016/j.envsoft.2018.02.002.

HORNER, T. W.; FREY, K. J. Methods for Determining Natural Areas for Oat Varietal Recommendations. **Agronomy Journal**, v. 49, n. 6, p. 313, 1957. doi: 10.2134/agronj1957.00021962004900060009x.

IBÁ. **Relatório Anual IBÁ 2015**. 2016. Disponível em: < <https://www.iba.org/publicacoes> >. Acesso em: 19 de jul. de 2021.

IBÁ. **Relatório Anual IBÁ 2019**. 2020. Disponível em: < <https://www.iba.org/publicacoes> >. Acesso em: 20 de jul. de 2021.

IBÁ. **Relatório Anual IBÁ 2021.**, 2022. Disponível em: < <https://www.iba.org/publicacoes> >. Acesso em: 26 de jan. de 2022.

IBGE. **Produção da Extração Vegetal e da Silvicultura**. 2021. Disponível em: < <https://sidra.ibge.gov.br/tabela/289> >. Acesso em: 25 de jul. de 2021.

ISIK, F.; HOLLAND, J.; MALTECCA, C. **Genetic Data Analysis for Plant and Animal Breeding**. Springer, Cham, 2017.

IVKOVIĆ, M. *et al.* Pattern of genotype by environment interaction for radiata pine in southern Australia. **Annals of Forest Science**, v. 72, n. 3, p. 391–401, 2015. doi: 10.1007/s13595-014-0437-6.

JASTRZĘBOWSKI, S. *et al.* Assessment of the height stability in progeny of *Fagus sylvatica* L. Populations using the GGE biplot method. **Dendrobiology**, v. 79, n. 3, p. 34–46, 2018. doi: 10.12657/denbio.079.004.

JAT, M. L. *et al.* Predicting Yield and Stability Analysis of Wheat under Different Crop Management Systems across Agro-Ecosystems in India. **American Journal of Plant Sciences**, v. 08, n. 08, p. 1977–2012, 2017. doi: 10.4236/ajps.2017.88133.

KAWAMURA, K. *et al.* Canopy hyperspectral sensing of paddy fields at the booting stage and PLS regression can assess grain yield. **Remote Sensing**, v. 10, n. 8, 1 ago. 2018. doi: 10.3390/rs10081249.

KING, G. J. Crop epigenetics and the molecular hardware of genotype \times environment interactions. **Frontiers in Plant Science**, v. 6, n. NOVEMBER, p. 1–19, 2015. doi: 10.3389/fpls.2015.00968.

- KUMAR, A. *et al.* Genotype \times environmental interactions for analyzing adaptability and stability in different clones of *Dalbergia sissoo* Roxb. **Journal of Forestry Research**, v. 23, n. 1, p. 65–74, 2012. doi: 10.1007/s11676-012-0234-8.
- LANDSBERG, J. J.; WARING, R. H. A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. **Forest Ecology and Management**, v. 95, n. 3, 1997. doi: 10.1016/S0378-1127(97)00026-1.
- LI, Y. *et al.* Genotype by environment interactions in forest tree breeding: review of methodology and perspectives on research and application. **Tree Genetics and Genomes**, v. 13, n. 3, 2017. doi: 10.1007/s11295-017-1144-x.
- LIM, H. *et al.* Assessing the cross-site and within-site response of potential production to atmospheric demand for water in Eucalyptus plantations. **Forest Ecology and Management**, v. 464, 15 maio 2020. doi: 10.1016/j.foreco.2020.118068.
- MALOSETTI, M. *et al.* Predicting responses in multiple environments: Issues in relation to genotype \times Environment interactions. **Crop Science**, v. 56, n. 5, p. 2210–2222, 2016. doi: 10.2135/cropsci2015.05.0311.
- MARCATTI, G. E. *et al.* GIS-based approach applied to optimizing recommendations of Eucalyptus genotypes. **Forest Ecology and Management**, v. 392, p. 144–153, 2017. doi: 10.1016/j.foreco.2017.03.006.
- MARTIN-STPAUL, N.; DELZON, S.; COCHARD, H. Plant resistance to drought depends on timely stomatal closure. **Ecology Letters**. Blackwell Publishing Ltd, 1 nov. 2017. doi: 10.1111/ele.12851.
- MATHEWS, K. L. *et al.* Multi-environment QTL mixed models for drought stress adaptation in wheat. **Theoretical and Applied Genetics**, v. 117, n. 7, p. 1077–1091, 2008. doi: 10.1007/s00122-008-0846-8.
- MAZER, S. J.; SCHICK, C. T. Constancy of population parameters for life history and floral traits in *raphanus sativus* L. I. norms of reaction and the nature of genotype by environment interactions. **Heredity**, v. 67, n. 2, p. 143–156, 1991. doi: 10.1038/hdy.1991.74.
- MINGOTI, R. *et al.* **Matopiba: caracterização das áreas com grande produção de culturas anuais**. Campinas: Embrapa gestão territorial, 2014. Nota Técnica
- 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 Agropecuaria Brasileira**, v. 56, 2021. doi: 10.1590/S1678-3921.pab2021.v56.02398.
- MORI, E. S.; KAGEYAMA, P. Y.; FERREIRA, M. Variação genética e interação progênies \times locais em *eucalyptus urophylla*. **Ipef**, v. 1, p. 53–63, 1988.
- MORI, E. S.; LELLO, L. R. B. DE; KAGEYAMA, P. Y. Efeitos da interação genótipo \times ambiente em progênies de *Eucalyptus saligna* Smith. **Ipef**, n. 33, p. 19–25, 1986.
- NUNES, S. *et al.* Implications of the Genotype-Environment Interaction Effects on Eucalyptus Clone Selection. **Cerne**, v. 8, n. 1, p. 49–58, 2002.

PEBESMA, E.; BIVAND, R. S. S Classes and Methods for Spatial Data: the sp Package. **Economic Geography**, v. 50, n. 1, 2005.

PEDERSEN, T. L. ggforce: Accelerating “ggplot2”. **CRAN Repository**, 2021.

PEIXOTO, L. de A. *et al.* L. *Jatropha* half-sib family selection with high adaptability and genotypic stability. **PLoS ONE**, v. 13, n. 7, p. 1–19, 2018. doi: 10.1371/journal.pone.0199880.

PIEPHO, H. P.; DENIS, J. B.; VAN EEUWIJK, F. A. Predicting Cultivar Differences Using Covariates. **Journal of Agricultural, Biological, and Environmental Statistics**, v. 3, n. 2, p. 151–162, 1998. doi: 10.2307/1400648.

PORKER, K. *et al.* Using a novel PLS approach for envirotyping of barley phenology and adaptation. **Field Crops Research**, v. 246, 1 fev. 2020. doi: 10.1016/j.fcr.2019.107697.

PUPIN, S. *et al.* Genotype x Environment interaction, stability, and adaptability in progenies of *Eucalyptus urophylla* S.T. BLAKE using the AMMI model. **Silvae Genetica**, v. 67, n. 1, p. 51–56, 2018. doi: 10.2478/sg-2018-0007.

QGIS DEVELOPMENT TEAM. **QGIS Geographic Information System. Open Source Geospatial Foundation Project**. 2021

R CORE TEAM. **R: A language and environment for statistical computing**. Viena, 2021.

RESENDE, R. T. *et al.* Enviromics in breeding: Applications and perspectives on envirotypic-assisted selection. **Theoretical and Applied Genetics**, p. 1–23, 2019. doi: 10.1101/726513.

REZENDE, W. S. *et al.* Half a century of studying adaptability and stability in maize and soybean in Brazil. **Scientia Agricola**, v. 78, n. 3, 2020. doi: 10.1590/1678-992x-2019-0197.

ROCHA, S. M. G. *et al.* A. Influence of climatic variations on production, biomass and density of wood in eucalyptus clones of different species. **Forest Ecology and Management**, v. 473, n. 82, p. 118290, 2020. doi: 10.1016/j.foreco.2020.118290.

RYAN, M. G. *et al.* Cross-site patterns in the response of Eucalyptus plantations to irrigation, climate and intra-annual weather variation. **Forest Ecology and Management**, v. 475, 1 nov. 2020. doi: 10.1016/j.foreco.2020.118444.

SANCHEZ, G. plsdepot: Partial Least Squares (PLS) Data Analysis Methods. **CRAN Repository**. 2012.

SANTOS, G. A. *et al.* Interação genótipos x ambientes para produtividade de clones de *Eucalyptus* l’hér. No estado do rio grande do sul. **Revista Árvore**, v. 39, n. 1, p. 81–91, 2015. doi: 10.1590/0100-67622015000100008.

SCHLOERKE, B. *et al.* Ggally: Extension to ggplot2. **CRAN Repository**. 2021

SCOLFORO, H. F. *et al.* Incorporating rainfall data to better plan eucalyptus clones deployment in eastern Brazil. **Forest Ecology and Management**, v. 391, p. 145–153, 1 maio 2017. doi: 10.1016/j.foreco.2017.02.025.

- SILVA, M. L. da. *et al.* Análise do custo e do raio econômico de transporte de madeira de reflorestamentos para diferentes tipos de veículos. **Revista Árvore**, v. 31, n. 6, dez. 2007. doi: 10.1590/S0100-67622007000600012.
- SILVA, R. R.; BENIN, G. Análises biplot: Conceitos, interpretações e aplicações Biplot analysis: concepts, interpretations and uses Raphael. **Ciencia Rural**, v. 42, n. 8, p. 1404–1412, 2012. doi: 10.1590/s0103-84782012000800012.
- SPARKS, A. nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Climatology Data Client for R. **Journal of Open Source Software**, v. 3, n. 30, 2018. doi: 10.21105/joss.01035.
- STAPE, J. L. *et al.* The Brazil Eucalyptus Potential Productivity Project: Influence of water, nutrients and stand uniformity on wood production. **Forest Ecology and Management**, v. 259, n. 9, p. 1684–1694, 2010. doi: 10.1016/j.foreco.2010.01.012.
- STAPE, J. L.; BINKLEY, D.; RYAN, M. G. Eucalyptus production and the supply, use and efficiency of use of water, light and nitrogen across a geographic gradient in Brazil. **Forest Ecology and Management**, v. 193, n. 1–2, p. 17–31, 17 maio 2004. doi: 10.1016/j.foreco.2004.01.020.
- WICKHAM, H. *et al.* dplyr: A Grammar of Data Manipulation. **CRAN Repository**, 2021.
- WICKHAM., H. ggplot2: Elegant Graphics for Data Analysis. **Springer-Verlag New York**, 2016.
- XU, Y. Envirotyping for deciphering environmental impacts on crop plants. **Theoretical and Applied Genetics**, v. 129, n. 4, p. 653–673, 2016. doi: 10.1007/s00122-016-2691-5.
- YAN, W.; HUNT, L. A. Genotype × Environment Interaction for Winter Wheat Yield in Ontario. **Crop Science**, v. 41, n. 1, p. 19–25, 2001. doi: 10.2135/cropsci2001.41119x.
- ZHANG, W. *et al.* One compound approach combining factor-analytic model with AMMI and GGE biplot to improve multi-environment trials analysis. **Journal of Forestry Research**, v. 31, n. 1, p. 123–130, 2018. doi: 10.1007/s11676-018-0846-8.

APÊNDICE A – Modelos Preditivos

Tabela A.1 – Coeficientes de regressão obtidos via Quadrados Mínimos Parciais (PLS) para os seis clones avaliados

Parâmetro	CLZ001	CLZ002	CLZ003	CLZ004	CLZ005	CLZ006
Intercepto	155,682	190,810	151,956	81,580	85,709	91,318
BIO_1	-0,453	-0,493	-0,433	0,298	-0,390	-0,104
BIO_2	-0,678	-0,919	0,084	-0,939	-0,235	-0,871
BIO_3	-0,030	-0,041	-0,031	-0,035	-0,106	-0,196
BIO_4	-0,001	0,001	0,001	-0,006	0,012	-0,004
BIO_5	-1,385	-1,445	-0,329	-0,919	-0,595	-1,187
BIO_6	-0,003	0,040	-0,208	0,750	-0,165	0,152
BIO_7	-0,338	-0,537	0,204	-1,019	-0,082	-0,346
BIO_8	-1,826	-2,816	-2,059	-1,538	-0,535	-0,531
BIO_9	-0,096	-0,135	-0,115	0,173	-0,277	0,005
BIO_10	-1,106	-1,162	-1,315	-0,410	-0,452	-0,157
BIO_11	-0,131	-0,176	-0,160	0,230	-0,291	-0,006
BIO_12	0,005	0,004	0,000	0,006	0,006	0,007
BIO_13	0,007	0,000	-0,001	0,006	0,015	0,026
BIO_14	0,035	0,056	0,001	0,035	0,027	0,036
BIO_15	-0,016	-0,037	-0,002	-0,018	-0,007	-0,020
BIO_16	0,003	0,000	0,000	0,002	0,007	0,007
BIO_17	0,011	0,017	0,001	0,011	0,009	0,010
BIO_18	0,004	0,009	0,003	0,003	0,010	0,015
BIO_19	0,002	0,001	0,000	0,002	0,002	0,011
lon	-0,131	-0,064	-0,023	-0,009	-0,170	0,117
lat	0,011	-0,034	-0,020	0,025	-0,149	-0,127
ELEV	0,002	0,001	0,005	-0,005	0,002	-0,002
BDOD	7,000	12,532	-4,294	13,098	-0,193	2,039
CEC	-0,018	0,070	-0,012	-0,017	-0,064	-0,038
CLAY	-0,114	-0,184	0,031	-0,112	-0,132	-0,068
SAND	0,015	-0,015	-0,066	0,005	-0,013	0,052

APÊNDICE B – Script em linguagem R

```
#####
#   Modelagem e Mapeamento de adaptabilidade de clones de eucalipto
#                               Versão 1.0
#
# Elaborado por: Leonardo Oliveira Silva da Costa
#
#####
```

```
#####
#####          Download SoilGrids
#####
```

```
library('rgdal')
library('XML')
library('gdalUtils')
library('sf')
library('dplyr')
```

```
pontos<- st_read('./Shapefiles/BrasilWGS.shp') #ShapeFile do Brasil
tar<- pontos #pontos[pontos$UF %in% c('MS','SP','TO','BA','MA','PA','MG','ES'), ]
#Selecionando estados
```

```
#Transformando o DATUM
igh='+proj=igh +lat_0=0 +lon_0=0 +datum=WGS84 +units=m +no_defs'
tar_igh <- st_transform(tar, igh)
```

```
#Obtendo as extremidades
(bbox <- st_bbox(tar_igh))
```

```
#Organizando para entrar com o Gdal
ulx = bbox$xmin
uly = bbox$ymax
lrx= bbox$xmax
lry = bbox$ymin
(bb <- c(ulx, uly, lrx, lry))
```

```
#layers <- c("5-15cm_mean", "15-30cm_mean", "30-60cm_mean", "60-100cm_mean", "100-200cm_mean")
```

```
#vetores com os atributos e profundidades
```



```

attributes <- c("bdod","cec","silt","clay","sand")
layers <- c("30-60cm_mean")
sg_url="/vsicurl/https://files.isric.org/soilgrids/latest/data/"

#Baixando cada combinação
for (i in 1:length(attributes)){
  for(j in 1:length(layers)){
    dados = paste(attributes[i],"/",attributes[i],"_",layers[j],".vrt", sep="")
    lfile = paste("./SoilGrids/",attributes[i],layers[j],"_RES_",".tiff",sep=")

    gdal_translate(paste0(sg_url,dados), lfile ,
                  tr=c(4500,4640),
                  projwin=bb,
                  projwin_srs =igh,
                  verbose=TRUE)

  }
}

#####
#####          Tipagem Ambiental
#####          WORLCLIM e SOILGRIDS

#Adaptado de COSTA-NETO(2017)
#Download WorldClim - https://worldclim.org/

#Selecionando os estados do estudo

WGS84<-CRS("+proj=longlat +datum=WGS84 +no_defs ") #DATUM
uf<-shapefile('./Shapefiles/BrasilWGS.shp') #Shapefile do Brasil
uf.sub<-uf[as.character(uf@data$UF) %in% c('MS','SP','TO','BA',"MA", "PA", 'MG', "ES"), ]
#Selecionando as UNF
projection(uf.sub) <- WGS84 # ajustando o DATUM

#Importando pontos

trees<- read.csv("./Dados_Inventario.csv", header=T,sep=';') #Dados de inventário
coordinates(trees)<- ~ lon + lat #forma que a longitude e latitude estao nos dados

```

```

# importa e corta para o tamanho da região alvo

(rasters<-list.files(path = "./Rasters_2.5m/", pattern = ".tif" ))
for(i in 1:length(rasters)){
  a<-raster(paste0("./Rasters_2.5m/",rasters[i]))
  projection(a)<-WGS84
  a.sub <- crop(a, extent(uf.sub))
  a.sub <- raster::mask(a.sub, uf.sub)
  rm(a)
  if(i == 1) { e <- a.sub } else { e <- stack(e,a.sub)} }

#Extração

for (j in 1:length(names(e))){
  if(i==1) vars<- vector() #criando vetor auxiliar
  a<- e[[j]]
  for ( i in 1:length(trees@data$id)){
    indv<- trees[trees@data$id == trees@data$id[i],]
    pp<- raster::extract(a,indv) # extrai o valor do pixel do raster j em que esta o ponto da
arvore i
    vars<- c(vars, pp)
  }

  cat(j*100/length(names(clim)),"%")
}

#Montando Tabela com covariáveis na ordem dos dados de inventario

mt<-matrix(vars, nrow=nrow(trees), ncol=length(rasters))
edacli<- data.frame(mt); names(edacli)<- names(e)
edacli$id<-trees@data$id
write.table(edacli, './Covariaveis.csv', sep=';',row.names = F)

#####
#####          INMET para cada talhao
#####

##### 1º Achar a estação mais proxima no ano de implantação

INMET<- read.csv(".INMET_2004-2021.csv", header=T,sep=";") #Dados INMET por dia de
2004-2021

```

```

#Convertendo data
INMET$DATA..YYYY.MM.DD.<-
as.Date(paste(INMET$DIA,INMET$MES,INMET$ANO,sep='-'),"%d-%m-%Y")

#Selecionando os estados
INMET<- INMET[INMET$UF %in%
c('MS','SP','TO','BA',"MA","PA",'MG',"ES","PR","GO","MT","RJ","PI"),]
INMET$ID

estacoes<- read.csv("estacoes.csv", header=T,sep=";") #Montar arquivo com as estacoes e
suas coordenadas

#importando dados
inv<-read.csv('.Inventario.csv', header=T, sep=';')

#Adicionando uma etiqueta
inv$Ponto<- 1:nrow(inv)
str(inv)

#Informação de Ano, mes e semestre

inv$data_plantio<- as.Date(inv$data_plantio, "%d/%m/%Y")
inv$data_medicao<- as.Date(inv$data_medicao, "%d/%m/%Y")
inv$Ano_plantio<- format(inv$data_plantio, "%Y")
inv$Ano_medicao<- format(inv$data_medicao, "%Y")
inv$MESi<- format(inv$data_plantio, "%m")
inv$DIAi<- format(inv$data_plantio,"%d")
inv$MESf<- format(inv$data_medicao, "%m")
inv$DIAf<- format(inv$data_medicao,"%d")

for (i in 1:nrow(inv)){
  if(i==1){ponto_est<-data.frame()}

  inv_i<-inv[i,]
  ano<-inv_i$Ano_plantio
  est_i<-estacoes[estacoes$DATA..YYYY.MM.DD. == inv_i$data_plantio,]

#Simplificando os dados
ponto<- inv_i[c(46,27:28)]

#Coordenadas e identificação dos pontos
CompID<-ponto$Ponto
sitexy<- data.matrix(ponto[2:3])

```



```

str(INMET)
str(ponto_est[1,])

unfs<-c("SP","MA","MS","BA")

#Loop por unidade florestal

for (j in 1:length(unfs)) {

  ponto_unf <- ponto_est[ponto_est$unf == unfs[j], ]

  for (i in 1:nrow(ponto_unf)) {

    pt <- ponto_unf[i, ]

    dias <-seq(as.Date(pt$datai),by = 'days',length = (pt$dataf - pt$datai) + 1)

    if (i == 1) {
      cat("\nINICIANDO EXTRAÇÃO DA UNIDADE FLORESTAL:', unfs[j],'\n')
      EST <-INMET[(INMET$ID == pt$ID) & (INMET$DATA..YYYY.MM.DD. %in%
dias), ]
      EST$Ponto <- pt$Ponto
      EST$daysFromStart<-1:nrow(EST)

    } else{
      est <- INMET[(INMET$ID == pt$ID) & (INMET$DATA..YYYY.MM.DD. %in% dias),
]
      est$Ponto <- pt$Ponto
      est$daysFromStart<-1:nrow(est)

      EST <- rbind(EST, est)

      if (i %% 100 == 0) {cat("\nEXTRAÇÃO",(" ,unfs[j],") : ",round(i * 100 /
nrow(ponto_unf), 2),"%")}
      if (i == nrow(ponto_unf)) {
        cat("\nSALVANDO ARQUIVO...")
        write.table(EST,paste("./Tabelas/INMET_DIARIO_", unfs[j], ".csv", sep = "),sep =
';',row.names = F)
        cat("\n\n##### EXTRAÇÃO COMPLETA #####\n\n")
      }
    }
  }
}

```

```
}
```

```
#Variaveis bio para cada talhão
```

```
rm(list=ls())
```

```
library(dismo)
```

```
library(dply)
```

```
#antes unir as tabelas de cada unidade
```

```
INMET<-read.csv('./Tabelas/INMET_DIARIO_.csv', sep=';', header=T)
```

```
str(INMET)
```

```
#Tabela dinamica com as médias do período de plantio-medição
```

```
pivot_INMET<- INMET %>% group_by(Ponto,unf,ANO,MES) %>% summarize(temp_min
= mean(temp_min, na.rm=T),
temp_max = mean(temp_max ,na.rm=T),
prec = sum(precipitacao,na.rm=T),
temp_noite = mean(temp_noite ,na.rm=T),
temp_max_noite = mean(temp_max_noite,na.rm=T),
temp_min_noite = mean(temp_min_noite,na.rm=T),
umidade_dia = mean(umidade_dia, na.rm=T),
umidade_min_dia = mean(umidade_min_dia,na.rm=T),
umidade_max_dia = mean(umidade_max_dia,na.rm=T),
vento_vel_hor = mean(vento_vel_hor, na.rm=T)) %>% group_by(Ponto,unf,MES) %>%
summarise(temp_min = mean(temp_min, na.rm=T),
temp_max = mean(temp_max,na.rm=T),
prec = mean(prec,na.rm=T),
temp_noite = mean(temp_noite ,na.rm=T),
temp_max_noite = mean(temp_max_noite,na.rm=T),
temp_min_noite = mean(temp_min_noite,na.rm=T),
umidade_dia = mean(umidade_dia, na.rm=T),
umidade_min_dia = mean(umidade_min_dia,na.rm=T),
umidade_max_dia = mean (umidade_max_dia,na.rm=T),
vento_vel_hor = mean(vento_vel_hor, na.rm=T))
```

```
(ponto<-unique(pivot_INMET$Ponto))
```

```
total <- length(ponto)
```

```
for (i in 1:total){
```

```
  if(i==1) BIO.INMET_DIA<- data.frame()
```

```
  #ponto
```

```
INMET_ponto<- pivot_INMET[pivot_INMET$Ponto == i,]
```

```
#Variáveis BIO
```

```
temp_min<- INMET_ponto$temp_min
```

```
temp_max<- INMET_ponto$temp_max
```

```
prec<- INMET_ponto$prec
```

```
df.na<-na.exclude(data.frame(temp_min,temp_max,prec))
```

```
bio.my<-data.frame(biovars(df.na$prec, df.na$temp_min,df.na$temp_max))
```

```
bio.my$t_min<-mean(INMET_ponto$temp_min,na.rm=T)
```

```
bio.my$t_max<-mean(INMET_ponto$temp_max,na.rm=T)
```

```
bio.my$temp_noite<-mean(INMET_ponto$temp_noite,na.rm=T)
```

```
bio.my$temp_max_noite <-mean(INMET_ponto$temp_max_noite,na.rm=T)
```

```
bio.my$temp_min_noite<-mean(INMET_ponto$temp_min_noite,na.rm=T)
```

```
bio.my$umidade_dia<-mean(INMET_ponto$umidade_dia,na.rm=T)
```

```
bio.my$umidade_min_dia<-mean(INMET_ponto$umidade_min_dia,na.rm=T)
```

```
bio.my$umidade_max_dia<-mean(INMET_ponto$umidade_max_dia,na.rm=T)
```

```
bio.my$prec<-sum(INMET_ponto$prec,na.rm=T)
```

```
bio.my$unf<-unique(INMET_ponto$unf)
```

```
bio.my$Ponto<- unique(INMET_ponto$Ponto)
```

```
bio.my$DSC<- sum(INMET_ponto$DSC,na.rm=T)
```

```
bio.my$varDSC<- var(INMET_ponto$DSC,na.rm=T)
```

```
bio.my$vento_vel_hor <- mean(INMET_ponto$vento_vel_hor,na.rm=T)
```

```
bio.my$VSDC<- mean(INMET_ponto$VSDC,na.rm=T)
```

```
BIO.INMET_DIA<- rbind(BIO.INMET_DIA, bio.my)
```

```
if(i%%100 ==0) cat("\n",round(i*100/length(ponto),2),"%")
```

```
}
```

```
#Filtro de ate 50km
```

```
BIO.INMET_DIA50 <- BIO.INMET_DIA [ BIO.INMET_DIA $DIST<50,]
```

```
write.table(BIO.INMET_DIA50,'./Tabelas/BIO.INMET_DIARIO_50km.csv',sep=';',row.names = F)
```

```
#####
```

```
##### Capacidade preditiva ( leave-onde-out)
```

```
#####
```

```
rm(list=ls())
```

```

#Importando os banco de dados com covariáveis e dados de inventário

WORLD<- read.csv("./Banco_WorldClim.csv header=T, sep=";")
INMET<- read.csv("./Banco_INMET.csv header=T, sep=";")
NASA<- read.csv("./Banco_Nasa.csv", header=T, sep=";")

#Padronizando as posições das covariáveis

cov<- c("id","unf" ,paste0("bio", 1:19), "lon", "lat", "alt", "ima7cc", "MatGen_oficial",
"BDOD","CEC","CLAY","SAND")

INMET<-INMET[cov]
NASA<-NASA[cov]
WORLD<-WORLD[cov]

pred<- c(3:24,27:30) #Posição das covariáveis ambientais
resp<-25 #Posição da variável resposta

##### PLS
library(plsdepot)
clones<- sort(unique(WORLD$MatGen_oficial)) #Obtendo os nomes dos clones

#Leave-One-Out
BANK<- list(WORLD,NASA,INMET) #Lista com os conjuntos de dados
names(BANK)<-c("WORLD","NASA","INMET") # Nomeando cada conjunto na lista

##Loop para realizar o leave-one-out por banco e por clone

for( k in 1:length(BANK)){

  if( k==1){
    r2.leave<-vector()
    ima.pred<-vector()
    predito<-vector()
    observ<-vector()
    clon<-vector()
    bank<-vector()
    uf<-vector()
    gg.ima<-list()
    lat<-vector()
    point<-vector()
  }
}

```



```

banco<-BANK[[k]]

cat("\n#####INICIANDO LEAVE ONE OUT COM BANCO DE DADOS:",
names(BANK)[k], "#####\n")

for (c in c(2)){
  for (j in 1:length(clones)){

    cat("\nINICIANDO LEAVE ONE OUT COM CLONE:",clones[j],"\t','BANCO DE
DADOS:', names(BANK)[k],"\t','COMPS:',c,'\n')

    banco.clone<-banco[banco$MatGen_oficial == clones[j],]
    total <- nrow(banco.clone)

    WORLD.clone<- WORLD[WORLD$id %in% banco.clone$id,]

    banco.clone<-banco.clone[order(banco.clone$id),]
    WORLD.clone<-WORLD.clone[order(WORLD.clone$id),]
    # Barra de progresso
    pb <- txtProgressBar(min = 0, max = total, style = 3)

    for( i in 1:total){

      lat<-c(lat,c)
      pls <-plsreg1(predictors = banco.clone[-i,][var],response = banco.clone[-i,][resp],comps
= c,crosval = T)
      r2.leave<-c(r2.leave,round(cor(pls$y.pred, pls$y$Sima7cc) ^ 2,4))

    pls.coefs<-pls$reg.coefs
    ima.pred<-
    pls.coefs[1]+
    pls.coefs[2]*WORLD.clone[i,][3]+
    pls.coefs[3]*WORLD.clone[i,][4]+
    pls.coefs[4]*WORLD.clone[i,][5]+
    pls.coefs[5]*WORLD.clone[i,][6]+
    pls.coefs[6]*WORLD.clone[i,][7]+
    pls.coefs[7]*WORLD.clone[i,][8]+
    pls.coefs[8]*WORLD.clone[i,][9]+
    pls.coefs[9]*WORLD.clone[i,][10]+
    pls.coefs[10]*WORLD.clone[i,][11]+
    pls.coefs[11]*WORLD.clone[i,][12]+
    pls.coefs[12]*WORLD.clone[i,][13]+
    pls.coefs[13]*WORLD.clone[i,][14]+
    pls.coefs[14]*WORLD.clone[i,][15]+
    pls.coefs[15]*WORLD.clone[i,][16]+
    pls.coefs[16]*WORLD.clone[i,][17]+
    pls.coefs[17]*WORLD.clone[i,][18]+
    pls.coefs[18]*WORLD.clone[i,][19]+

```

```

pls.coefs[19]*WORLD.clone[i,][20]+
pls.coefs[20]*WORLD.clone[i,][21]+
pls.coefs[21]*WORLD.clone[i,][22]+
pls.coefs[22]*WORLD.clone[i,][23]+
pls.coefs[23]*WORLD.clone[i,][24]+
pls.coefs[24]*WORLD.clone[i,][27]+
pls.coefs[25]*WORLD.clone[i,][28]+
pls.coefs[26]*WORLD.clone[i,][29]+
pls.coefs[27]*WORLD.clone[i,][30]

clon<-c(clon, clones[j])
bank<-c(bank, names(BANK)[k])
predito<-c(predito, ima.pred[,1])
observ<- c(observ, banco.clone[i,][25]$ima7cc)
uf<-c(uf, banco.clone[i,]$unf)
point<-c(point, banco.clone[i,]$id)

#Atualizando Barra de Progresso
setTxtProgressBar(pb, i)

}

close(pb)

if(i == total){

df.plot<-data.frame(uf,clon,bank,predito,observ)

df.plot <- df.plot %>%
mutate(Cor = case_when(
  uf == "BA" ~ 'red',
  uf == "MS" ~ 'blue',
  uf == 'MA' ~ 'yellow',
  uf == 'SP' ~ 'green'))

df.clon<-df.plot[(df.plot$bank == names(BANK)[k]) & (df.plot$clon == clones[j]),]

gg.pls<-ggplot(df.clon, aes(x=predito, y=observ)) +
geom_point(alpha=0.5,col=df.clon$Cor) +
geom_smooth(method=lm, color="black", fill="#69b3a2", se=TRUE) +
labs(
  x = 'Valor Predito',
  y = "Valor Observado",
  title = clones[j],
  subtitle = paste("R2:", round(
    cor(df.clon$predito, df.clon$observ, use = "complete.obs") ^ 2, 2
  ), " | ", "RMSE:", round(sqrt(mean((df.clon$predito - df.clon$observ) ^ 2, na.rm =
T)),2))) +
  theme_bw()

```

```

    gg.ima[[j]]<-gg.pls
  }

  if(j == length(clones)){
    pdf(paste("./",names(BANK)[k],"_BANK-
WORLD_SEMSG5m_", "_",c,"_IMA.pdf",sep="), width = 10, height = 7)
    grid.arrange(gg.ima[[1]],gg.ima[[2]],gg.ima[[3]],gg.ima[[4]],gg.ima[[5]],gg.ima[[6]],top
= paste(names(BANK)[k],"n IMA aos 7 anos\n", "COMPS:" ,c,sep="),nrow=2)
    dev.off()
  }
}
}
}
}

```

```

#Tabela com os valore4s observados e preditos
df<- data.frame(bank,Clone=clon,predito,observ,lat,r2.ajste =r2.leave,unf=uf, id =point)
write.table(df,"/LOO_pred(BANK-WORLD).csv",sep=',';) #Salvar os dados

```

```

#####
#####          Metricas
#####

```

```

#Tabela com R2 e RMSE

```

```

(DF<-df %>% group_by(Clone,bank,lat) %>% summarise(R2 = cor(predito,observ, use =
"complete.obs")^2, RMSE =sqrt( sum ( (predito-observ)^2/length(observ),na.rm=T ) ) )

```

```

clones<- paste0("CLZ00",1:6) #Vetor com clones
BANKS<-c("WORLD","INMET","NASA") #Vetor com os bancos

```

```

for(k in 1:length(BANKS)){

```

```

  bancos<-df[df$bank == BANKS[k],]
  me<-DF[DF$bank == BANKS[k],]

```

```

  for (i in 1:length(clones)){
    if(i==1) {pred.int<-vector(); n.obs<-vector(); pred.sup<-vector(); pred.inf<-vector(); cor.p<-
vector();cor.s<-vector();rmse<-vector();r2<-vector()}

```

```

    clone1<-bancos[bancos$Clone == clones[i],]
    vars<-clone1$predito
    df.clon<- clone1[c("predito","observ")]

```

```

df.clon<- df.clon %>%

```

```
mutate(int = case_when(predito < 0 ~ '< 0',
  (predito > 0) & (predito) <10 ~ '0-10',
  (predito > 10) & (predito) <20 ~ '10-20',
  (predito > 20) & (predito) <30 ~ '20-30',
  (predito > 30) & (predito) <40 ~ '30-40',
  (predito > 40) & (predito) <50 ~ '40-50',
  (predito > 50) & (predito) <60 ~ '50-60',
  (predito > 60) & (predito) <70 ~ '60-70',
  (predito > 70) & (predito) <80 ~ '70-80',
  (predito > 80) ~ '> 80'))
```

```
df.clon<- df.clon %>%
```

```
mutate(int.obs = case_when(observ < 0 ~ '< 0',
  (observ > 0) & (observ) <10 ~ '0-10',
  (observ > 10) & (observ) <20 ~ '10-20',
  (observ > 20) & (observ) <30 ~ '20-30',
  (observ > 30) & (observ) <40 ~ '30-40',
  (observ > 40) & (observ) <50 ~ '40-50',
  (observ > 50) & (observ) <60 ~ '50-60',
  (observ > 60) & (observ) <70 ~ '60-70',
  (observ > 70) & (observ) <80 ~ '70-80',
  (observ > 80) ~ '> 80'))
```

```
df.clon<- df.clon %>%
```

```
mutate(sup = case_when(observ < 0 ~ 0,
  (observ > 0) & (observ <10) ~ 10,
  (observ > 10) & (observ <20) ~ 20,
  (observ > 20) & (observ <30) ~ 30,
  (observ > 30) & (observ <40) ~ 40,
  (observ > 40) & (observ <50) ~ 50,
  (observ > 50) & (observ <60) ~ 60,
  (observ > 60) & (observ <70) ~ 70,
  (observ > 70) & (observ <80) ~ 80),
```

```
inf = case_when(
```

```
  (observ > 0) & (observ <10) ~ 0,
  (observ > 10) & (observ <20) ~ 10,
  (observ > 20) & (observ <30) ~ 20,
  (observ > 30) & (observ <40) ~ 30,
  (observ > 40) & (observ <50) ~ 40,
  (observ > 50) & (observ <60) ~ 50,
  (observ > 60) & (observ <70) ~ 60,
  (observ > 70) & (observ <80) ~ 70,
  (observ > 80) ~ 80))
```

```
df.clon$match<- df.clon$int==df.clon$int.obs
```

```

df.clon$match.sup<- df.clon$predito > df.clon$sup
df.clon$match.inf<- df.clon$predito < df.clon$inf
df.clon<-df.clon[is.na(df.clon$int)==F,]

pred.sup<-c(pred.sup,round(sum(df.clon$match.sup,na.rm=T)/length(df.clon$match),4))
pred.inf<-c(pred.inf,round(sum(df.clon$match.inf,na.rm=T)/length(df.clon$match),4))
pred.int<-c(pred.int,round(sum(df.clon$match,na.rm=T)/length(df.clon$match),4))
n.obs<- c(n.obs, nrow(df.clon))
cor.p<-c(cor.p, round(cor(df.clon$predito,df.clon$observ, use="complete.obs"),4))
cor.s<-c(cor.s, round(cor(df.clon$predito,df.clon$observ, use="complete.obs", method =
"spearman"),4))
rmse<- me$RMSE
r2<- me$R2
}

(
df.int <-
  data.frame(
    Clone = clones,
    Concordancia = pred.int,
    Sup = pred.sup,
    Infe = pred.inf,
    R2 = r2,
    RMSE = rmse,
    Correlacao.S = cor.s,
    N° = n.obs
  )
)

#Gera uma tabela para cada banco
write.table(df.int,paste0("./","Metricas_",BANKS[k],".csv"),sep=';',row.names = F)

}

#####
#####          Caracterizacao Ambiental
#####

library(EnvRtype)

names(WORLD[pred])

```

```

#dataf = data final
#unf = Unidade Florestal

env<-WORLD[c("unf","lat","lon",'dataf',names(WORLD[pred[pred]]))]

#Organizar arquivo com as colunas demandadas pela função
#Como o trabalho utilizou médias, teremos que adaptar considerando cada local como um dia

names(env)[1:4]<-c('env',"LAT","LON","YYYYMMDD")
env$YYYYMMDD<-as.Date(env$YYYYMMDD,'%Y-%m-%d')
env$YEAR<- format(env$YYYYMMDD,"%Y")
env$MM<- format(env$YYYYMMDD,"%m")
env$DD<- format(env$YYYYMMDD,"%d")
env$DOY<- 1:nrow(env)
env$daysFromStart<- 1:nrow(env)

str(env)

#Definicao de intervalos para o calculo de frequencia
inter<-list("BIO_1" = c(13,20,24,26,30),"BIO_12" = c(1000,1500,1800,2000,2500), "CLAY"
= c(20,30,40,50,60),"SAND" = c(20,30,40,50,60))

#Caracterizacao por frequencia de valores dos intervalos
typin_env<-env_typing(env.data = env, env.id = 'unf', var.id =
c("BIO_1","BIO_12","CLAY","SAND"), cardinals = inter)

#Similaridade entre ambientes
library(viridis)

W <- W_matrix(env.data = env, var.id = names(WORLD[pred]) , statistic = 'mean')

#Cálculo de similaridades
distances <- env_kernel(env.data = W, gaussian = T)[[2]]
heatmap(distances) #heatmap

#####
#####          Super/Subestimação
#####
rm(list=ls())

#Erros de predição
residuo<- read.csv("./Media_Pixel/Dispersao/Residuo.csv", header=T,sep=';')

#RMSE médio

```

```
df<- read.csv("./LOO_pred(BANK-WORLD).csv", header=T,sep=';') # Arquivo com os
valores observados e preditos
```

```
(DF<-df[df$bank == "WORLD",] %>% group_by(Clone,bank,lat) %>% summarise(R2 =
cor(predito,observ, use = "complete.obs")^2, RMSE =sqrt( sum ( (predito - observ) ^2
/length(observ),na.rm=T ) ) ) )
```

```
(rmse.mean<-mean(DF$RMSE))
```

```
#Definindo as classes
```

```
residuo<-residuo %>% mutate(classe = case_when(
  residuo < -rmse.mean ~ "Superestimado",
  residuo > rmse.mean ~ "Subestimado",
  residuo <= rmse.mean & ( residuo > - rmse.mean) ~ "Adequado"))
```

```
#Proporções por unf
```

```
unfs<-c("BA","MA","SP","MS")
```

```
for(j in 1:length(unfs)){
```

```
  c.unf<- c<-residuo[residuo$uf == unfs[j],]
  df.freq<-data.frame(round(table(c.unf$classe)*100/sum(table(c.unf$classe)),4))
  df.freq$unf<-unfs[j]
```

```
  if(j==1) DF.freq<- df.freq else DF.freq <- rbind(DF.freq, df.freq)
}
```

```
#Plotando o gráfico
```

```
names(DF.freq)[1]<- "Classe"
```

```
library(ggplot2)
```

```
library(shadowtext)
```

```
ggplot(data=DF.freq, aes(x=unf, y=Freq, fill=Classe)) +
  geom_bar(stat="identity", position=position_dodge(), width = 0.8)+
  geom_shadowtext(aes(label= round(Freq,2)), vjust=1.6, color="white",fontface =
"bold",position = position_dodge(0.8), size=3.5)+
  labs(x= "Unidade Florestal", y="Frequência (%)")+
  theme(plot.title = element_text(size = 26, face = "bold"), #Aqui vc ajusta os tamanhos das
fontes
```

```
  plot.subtitle = element_text(size = 22),
  axis.title =element_text(size = 18),
  axis.text =element_text(size = 16), #coloquei itálico aqui pq é espécie
  legend.text =element_text(size = 12),
  legend.title =element_text(size = 16)) +
```

```
scale_fill_manual("Classe", values = c("Superestimado" = "red", "Adequado" = "green",
"Subestimado" = "blue")) # Aqui você muda as cores das barras, muda so o red e blue
```

```
#####
#####                               Mapas
#####
```

```
#!!! Utilize o melhor banco para plotagem com a mesma ordem definida no objeto “cov”
```

```
#Montando o arquivo com os rasters
```

```
# variáveis WORLDclim
```

```
#Importando os rasters BIO, Longitude, Latitude e Elevacao (OBS: Nesse caso, situados na
pasta Rasters_2.5m)
```

```
(arq<-c(list.files(path = "./Rasters_2.5m/",pattern = "bio"), "LONG.tif", "LAT.tif",
"BR_wc2.1_2.5m_elev.tif"))
```

```
for (i in 1:length(arq)){
  b<- raster(paste("./Rasters_2.5m/", arq[i],sep=""))
```

```
  if(i==1) W<-b else W<-stack(W,b)
```

```
}
```

```
#Variáveis SoilGrids
```

```
#Importando os rasters do soil Grids
```

```
(arq<-c(list.files(path = "./Rasters_2.5m/",pattern = "SoilGrids")))
```

```
for (i in 1:length(arq)){
  b<- raster(paste("./Rasters_2.5m/",arq[i],sep=""))
```

```
  if(i==1) S<-b else S<-stack(S,b)
```

```
}
```

```
#Juntando WorldClim com SoilGrids
```

```
V<-stack(W,S)
```

```
#conferindo a ordem dos rasters com as covariáveis
```

```
names(V)
```

```
names(WORLD[var])
```

```
#Definindo o banco para plotagem
```



```

banco<- WORLD

#Ajuste do modelo e plotagem

for (j in 1:length(clones)) {
  if(j==1){ r2<-vector();clz<-vector();rmse<-vector()}

  cat("\n\nINICIANDO ESTIMATIVAS COM CLONE:",clones[j],"\n")

  banco.clone <- banco[banco$MatGen_oficial == clones[j], ]

  pls.q2 <-
  plsreg1(
    predictors = banco.clone[var],
    response = banco.clone[resp],
    comps = 20,
    crosval = T
  )
  df.q2<-data.table::data.table(pls.q2$Q2)

  pls <-
  plsreg1(
    predictors = banco.clone[var],
    response = banco.clone[resp],
    comps = 2#ifelse(sum(df.q2$Q2>=df.q2$LimQ2)<2,2,sum(df.q2$Q2>=df.q2$LimQ2))
  )
  rmse<-c(rmse,sqrt(mean((pls$y.pred-pls$y$ima7cc)^2)))
  r2<-c(r2,cor(pls$y.pred,pls$y$ima7cc)^2)
  clz<-c(clz,clones[j])

  cat("R^2:",cor(pls$y.pred,pls$y$ima7cc)^2,"\t","RMSE:", sqrt(mean((pls$y.pred-
pls$y$ima7cc)^2)), "\t", "Q^2:",
ifelse(sum(df.q2$Q2>=df.q2$LimQ2)<2,2,sum(df.q2$Q2>=df.q2$LimQ2)))
  pls.coefs <- pls$reg.coefs
  pls.stdcoefs<-pls$std.coefs

  if (j==1){ DF.coefs<-data.frame(pls.coefs)} else{ df.coefs<- data.frame(pls.coefs);
DF.coefs<- cbind(DF.coefs,df.coefs)}
  if (j==1){ DF.stdcoefs<-data.frame(pls.stdcoefs)} else{ df.stdcoefs<-
data.frame(pls.stdcoefs); DF.stdcoefs<- cbind(DF.stdcoefs,df.stdcoefs)}

  pred <-
  pls.coefs[1] +
  pls.coefs[2] * V[[1]] +
  pls.coefs[3] * V[[2]] +

```

```

pls.coefs[4] * V[[3]] +
pls.coefs[5] * V[[4]] +
pls.coefs[6] * V[[5]] +
pls.coefs[7] * V[[6]] +
pls.coefs[8] * V[[7]] +
pls.coefs[9] * V[[8]] +
pls.coefs[10] * V[[9]] +
pls.coefs[11] * V[[10]] +
pls.coefs[12] * V[[11]] +
pls.coefs[13] * V[[12]] +
pls.coefs[14] * V[[13]] +
pls.coefs[15] * V[[14]] +
pls.coefs[16] * V[[15]] +
pls.coefs[17] * V[[16]] +
pls.coefs[18] * V[[17]] +
pls.coefs[19] * V[[18]] +
pls.coefs[20] * V[[19]] +
pls.coefs[21] * V[[20]] +
pls.coefs[22] * V[[21]] +
pls.coefs[23] * V[[22]] +
pls.coefs[24] * V[[23]] +
pls.coefs[25] * V[[24]] +
pls.coefs[26] * V[[25]] +
pls.coefs[27] * V[[26]]

names(pred)<- clones[j]

if(j==1) PRED<-pred else PRED<-stack(PRED,pred)

}

plot(PRED) #plotando os mapas

#Quem-Vence-Onde
QVO<-whiches.max(PRED)

#Obs: O mapa resultante informa o vencedor numericamente, ou seja, a camada que
apresenta o maior valor para o pixel. Esse Script organiza o objeto PRED (stack raster que
possui todos os mapas) na ordem do objeto clones (objeto que possui o nome dos clones)

#####
#####          Salvando Rasters
#####

WGS84<-CRS("+proj=longlat +datum=WGS84 +no_defs ") #DATUM

```

```

#Produtividade
for (i in 1:length(names(PRED))) {
  a<- PRED[[i]]
  projection(a)<-WGS84
  writeRaster(a,file = paste0('./names(PRED[[i]]),"_PRED",".tif"))

#Quem-vence-onde
  writeRaster(QVO,file = "Quem-Vence-Onde_PRED.tif")

#####
#####          Ranqueamento de Coeficientes
#####

#Adicionando nome dos clones
names(DF.coefs)<-clones; DF.coefs$var<- rownames(DF.coefs)

names(DF.stdcoefs)<-clones; DF.stdcoefs$var<- rownames(DF.stdcoefs)

#Variáveis mais responsivas por clone

for( i in 1:(ncol(DF.stdcoefs)-1)){

  if(i==1){
    clz<-vector()
    vr<-vector()
    val<-vector()
  }

  DF.clz<-DF.stdcoefs[c(i,7)][order( abs(DF.stdcoefs[c(i,7)],[,1]),decreasing = T),]
  vr<- c(vr, rownames(DF.clz)[1:5])
  val<- c(val, DF.clz[,1][1:5])
  clz<-c(clz, rep(names(DF.stdcoefs)[i],5))

  if(i==(ncol(DF.stdcoefs)-1)) DF.sensc<-data.frame(Clone =
  clz,Variavel=vr,Valor=val,Posicao= rep(1:5,6))

}

```