



**GABRIEL THALES DA CUNHA**

**UMA NOVA ABORDAGEM PARA INVENTÁRIO DE  
BIOMASSA EM PLANTAÇÕES DE EUCALIPTO NO  
BRASIL**

**LAVRAS – MG**

**2023**

**GABRIEL THALES DA CUNHA**

**UMA NOVA ABORDAGEM PARA INVENTÁRIO DE BIOMASSA EM  
PLANTAÇÕES DE EUCALIPTO NO BRASIL**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de concentração em Manejo Florestal, para a obtenção do título de Mestre.

Prof. Dr. Lucas Rezende Gomide  
Orientador

Dr. Henrique Ferraço Scolforo  
Coorientador

**LAVRAS – MG  
2023**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca  
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

da Cunha, Gabriel Thales.

Uma nova abordagem para inventário de biomassa em  
plantações de eucalipto no Brasil / Gabriel Thales da Cunha. - 2023.  
63 p.

Orientador(a): Lucas Rezende Gomide.

Coorientador(a): Henrique Ferraço Scolforo.

Dissertação (mestrado acadêmico) - Universidade Federal de  
Lavras, 2023.

Bibliografia.

1. Densidade Básica da Madeira. 2. Programação Genética. 3.  
Modelo de Efeito Misto. I. Gomide, Lucas Rezende. II. Scolforo,  
Henrique Ferraço. III. Título.

**GABRIEL THALES DA CUNHA**

**UMA NOVA ABORDAGEM PARA INVENTÁRIO DE BIOMASSA EM  
PLANTAÇÕES DE EUCALIPTO NO BRASIL  
A NEW APPROACH FOR BIOMASS INVENTORY IN EUCALIPT STAND IN  
BRAZIL**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de concentração em Manejo Florestal, para a obtenção do título de Mestre.

APROVADA em 29 de agosto de 2023.

Prof. Dr. Lucas Rezende Gomide

Dr. Henrique Ferraço Scolforo

Prof. Dr. José Roberto Soares Scolforo

UFLA

SUZANO S/A

UFLA

Prof. Dr. Lucas Rezende Gomide - UFLA

Orientador

Dr. Henrique Ferraço Scolforo – SUZANO S/A

Coorientador

**LAVRAS – MG  
2023**

## AGRADECIMENTOS

Esse trabalho de dissertação foi fruto de esforço, empenho, discussões e indagações. E se tornou possível somente com ajuda e influência de pessoas iluminadas e instituições que tive a oportunidade de cruzar os caminhos ao longo dos últimos anos. Os agradecimentos aqui contidos expressam uma pequena parte da gratidão que tenho por todos.

Em primeiro lugar, gostaria de agradecer verdadeiramente à Deus e à Nossa Senhora Aparecida por mais essa oportunidade a mim concedida, por iluminarem e guiarem na caminhada.

Agradeço aos meus pais por todo ensinamento, carinho, amor e incentivo que sempre me deram. Tenham certeza que o ser humano que me tornei é um espelho da humildade e fé do Seu Dito e fruto do afinco e afeto da Dona Vânia. Fica aqui registrado meu singelo obrigado por absolutamente tudo que já fizeram, fazem e continuarão fazendo por mim. Vocês são fundamentais em cada linha deste trabalho.

Agradeço também à minha irmã Janaina pelos constantes conselhos, conversas e cuidar para que tudo sempre esteja bem desde quando resolvi me aventurar em uma vida longe da nossa casa. Saiba que sem você nada é possível e que meu carinho por você é imensurável.

Agradeço à minha companheira de vida Alana, por estar sempre ao meu lado em cada um dos momentos bons e ruins dos últimos 11 anos. Obrigado pelo amor, carinho, paciência e por ouvir meus desabafos em dias ruins. Admiro a mulher e a profissional que você é, e sabia que a cada luta sua também te apoiarei incondicionalmente.

Meu agradecimento ao Professor Lucas Gomide, que aceitou mais uma vez o desafio de me orientar, porém agora em um momento diferente da minha vida. Minha eterna gratidão pela compreensão nos momentos em que as demandas corporativas competiram com as disciplinas e momentos de estudo do mestrado. Por todo incentivo e orientações nos momentos difíceis e principalmente pela paciência que tem comigo desde o período de graduação, saiba que grande parte do profissional que me tornei hoje é fundamentado em seus ensinamentos e conselhos.

Ao meu amigo, coorientador, gestor, e grande incentivador, Henrique Ferraço. Todo esse sonho só foi viável pelas suas indagações e provações desde o processo de inscrição para a pós-graduação, passando pelos puxões de orelha no meio do percurso

e por toda orientação e auxílio na realização do projeto. Não tenho como agradecer toda ajuda e tempo dedicados ao projeto. Gostaria de deixar registrado aqui minha admiração e gratidão pela pessoa e pelo profissional que você é.

Aos meus amigos Evandro e Talles, tanto pelo suporte dentro da empresa como também em todo processo de geração de cada uma das etapas desse trabalho. Admiro muito a amizade de vocês e a dedicação em meu ajudar mesmo com todas as demandas do mundo corporativo.

Agradeço também aos membros da banca de Defesa de Mestrado, que puderam abdicar de um tempo em suas agendas para participar e contribuir no processo de desenvolvimento desta dissertação.

Agradeço a Universidade Federal de Lavras (UFLA) e ao Laboratório de Estudos em Manejo Florestal (LEMAF) pela oportunidade de continuar adquirindo conhecimento desde o período de Graduação.

Agradeço a Suzano Celulose e Papel, aos meus gestores e a equipe de Inventário Florestal, pela oportunidade de aprimoramento pessoal e profissional concedidos a mim. Além de todo suporte que me foi dado no período de Pós-Graduação.

Por fim, a todos aqueles que contribuíram, direta ou indiretamente, para a realização desta dissertação, o meu sincero e singelo agradecimento.

## RESUMO GERAL

Tradicionalmente, os inventários florestais auxiliam na determinação dos estoques volumétricos e em sua estrutura em plantios comerciais. O volume sempre foi a principal variável estudada, sendo um paradigma científico a ser mudado. Hoje, diante das inúmeras aplicações da biomassa, a indústria 4.0 passou a requisitar informações mais detalhadas do ativo florestal, não apenas o volume, o que impacta diretamente na forma como se estruturam os inventários e o consumo de madeira. Considerando essa premissa, a densidade básica é a principal variável a ser devidamente explorada e modelada, em virtude de sua relação positiva com a biomassa produzida através do crescimento das árvores. Assim, o presente estudo propôs uma nova abordagem para o inventário de biomassa por meio de uma metodologia híbrida que utiliza métodos não-destrutivos para a obtenção de densidade básica da madeira, programação genética para seleção de variáveis e geração de um modelo linear, e modelagem de efeito misto para aumentar a precisão das estimativas da variável dependente. O experimento considerou o uso de 55 parcelas amostrais distribuídas em 3 clones comerciais de *Eucalyptus* localizados em São Paulo, Brasil. As variáveis clássicas ao nível da árvore foram adotadas e somadas à dureza (Pilodyn), nas amostras extraídas. Em um segundo momento, modelou-se a densidade (proveniente do modelo anterior) apenas com a utilização de variáveis coletadas no inventário tradicional. Por fim, no último passo, estimou-se a biomassa florestal por meio do produto entre densidade básica e volume individual. A metodologia implementada se apresentou eficaz, com alta aderência à densidade laboratorial. Conclui-se que a proposta metodológica viabiliza a implementação em escala empresarial, permitindo informações precisas de biomassa ao nível da árvore. Além disso, a aplicação da programação genética demonstrou ser um método robusto e prático na modelagem.

**Palavras chaves:** Densidade básica da madeira. Programação genética. Modelo de efeito misto.

## GENERAL ABSTRACT

Traditionally, forest inventories assist in determining volumetric stocks and their structure in commercial plantations. Volume has always been the primary studied variable, representing a scientific paradigm to be changed. In the present moment, considering the manifold applications of biomass, Industry 4.0 has started to seek more comprehensive data regarding forest resources, extending beyond mere volume. Consequently, the arrangement of inventories and the management of wood consumption are directly affected by this necessity. With this concept in consideration, fundamental density emerges as the primary variable to be thoroughly investigated and modeled, owing to its favorable correlation with biomass generated through tree growth. Thus, this research proposes a new approach to biomass inventory through a hybrid methodology that employs non-destructive methods to obtain basic wood density, genetic programming for variable selection and linear model generation, and mixed-effects modeling to enhance the accuracy of dependent estimated variables. The experiment involved using 55 sample plots distributed across three commercial clones of Eucalyptus located in São Paulo, Brazil. Classic tree-level variables were adopted and combined with hardness (Pilodyn) in the extracted samples. In a subsequent step, density (derived from the previous model) was modeled using only variables collected in the traditional inventory. Finally, in the last stage, biomass forest was estimated by multiplying basic wood density and individual volume. The applied methodology has proven its efficacy, demonstrating a substantial correlation with laboratorial density. In conclusion, it can be inferred that the methodological proposition renders the implementation on an enterprise scale viable, thereby facilitating the provision of accurate tree-level biomass data. Additionally, the application of genetic programming demonstrated its robust and practical modeling capabilities.

**Keywords:** Basic wood density. Genetic programming. Linear mixed effects model.



## LISTA DE FIGURAS

### PRIMEIRA PARTE – REVISÃO BIBLIOGRÁFICA

Figura 1 - Distribuição das florestas plantadas no Brasil.....	13
Figura 2 - Valor da produção do setor florestal brasileiro.....	14
Figura 3 – Ciclo do Eucalyptus para indústria de celulose (Do plantio a produção de celulose). .....	15
Figura 4 - Tora de madeira de Eucalyptus (I) e sua composição (Cortes Longitudinais II e IV e Transversais II e III - Microscópio).....	18

### SEGUNGA PARTE – ARTIGO

Figura 1 - Distribuição espacial das unidades amostrais mensuradas para o desenvolvimento dos modelos e suas interações ambientais.....	45
Figura 2 - Coleta da dureza da árvore com uso do Pilodyn nas 3 posições de coleta.....	47
Figura 3 - Fluxograma conectando o passo a passo do trabalho.....	48
Figura 4 – Relação Iteração x RMSE na seleção e construção de expressões a serem utilizadas posteriormente para ajuste dos modelos PGM1(a) e PGM2(b).....	53
Figura 5 - Expressões geradas pela PG a serem utilizadas no ajuste do modelo PGM1.....	53
Figura 6 - Expressões geradas pela PG a serem utilizadas no ajuste do modelo PGM2.....	54
Figura 7 - DB observada em laboratório x DB estimada pelo PGM1 (a); DB observada em laboratório x DB estimada pelo PGM2(b).....	56
Figura 8 - Incerteza média através da simulação bootstrap de 5.000 bases de dados para os modelos PGM1 (Eq. 8) (a); e PGM2 (Eq. 9) (b).....	57
Figura 9 - Biomassa observada x biomassa estimada (DB estimada a partir da Eq. 9).....	57

## LISTA DE TABELAS

### SEGUNDA PARTE - ARTIGO

Tabela 1 - Resumo do volume total com casca (Vol., m <sup>3</sup> .ha <sup>-1</sup> ) e densidade básica da madeira (DB, kg.m <sup>-3</sup> ) para os diferentes clones, zonas climáticas, classes de idade e sítio em SP.....	46
Tabela 2 - Parâmetros fixos e aleatórios do modelo ajustado PGM1.....	55
Tabela 3 - Parâmetros fixos e aleatórios do PGM2.....	56

## SUMÁRIO

<b>PRIMEIRA PARTE – REVISÃO BIBLIOGRÁFICA.....</b>	<b>11</b>
<b>1</b>	<b>INTRODUÇÃO.....11</b>
<b>2</b>	<b>REVISÃO DE LITERATURA.....12</b>
<b>2.1</b>	<b>SETOR FLORESTAL E A SUA IMPORTÂNCIA.....12</b>
<b>2.2</b>	<b>PLANTIOS COMERCIAIS DE <i>EUCALYPTUS</i>.....14</b>
<b>2.3</b>	<b>DENSIDADE BÁSICA DA MADEIRA.....17</b>
<b>2.3.1</b>	<b>Métodos não-destrutivos para determinação da densidade básica da madeira 20</b>
<b>2.4</b>	<b>MODELAGEM PREDITIVA DE ATRIBUTOS FLORESTAIS.....22</b>
<b>2.5</b>	<b>INTELIGÊNCIA ARTIFICIAL E APRENDIZADO DE MÁQUINA.....25</b>
<b>2.6</b>	<b>MODELOS DE EFEITO MISTO.....29</b>
<b>3</b>	<b>CONSIDERAÇÕES FINAIS.....30</b>
<b>REFERÊNCIAS.....</b>	<b>32</b>
<b>SEGUNDA PARTE - ARTIGO.....</b>	<b>39</b>
<b>1</b>	<b>INTRODUÇÃO.....42</b>
<b>2</b>	<b>MATERIAL E MÉTODOS.....44</b>
<b>2.1</b>	<b>CARACTERIZAÇÃO DA ÁREA DE ESTUDO.....44</b>
<b>2.2</b>	<b>CARACTERIZAÇÃO DA ÁREA DE ESTUDO.....45</b>
<b>2.3</b>	<b>AMOSTRAGEM DA DUREZA E DENSIDADE BÁSICA DA MADEIRA (DB).....47</b>
<b>2.4</b>	<b>SISTEMA DE MODELOS PARA GERAÇÃO DO INVENTÁRIO DE BIOMASSA (SMIB) 48</b>
<b>2.4.1</b>	<b>Seleção e construção de novas variáveis.....49</b>
<b>2.4.2</b>	<b>Modelos de efeito misto.....51</b>
<b>2.5</b>	<b>CRITÉRIOS DE AVALIAÇÃO DE SMIB.....52</b>
<b>3</b>	<b>RESULTADOS.....53</b>
<b>3.1</b>	<b>SELEÇÃO E CONSTRUÇÃO DAS EXPRESSÕES ATRAVÉS DA PG.....53</b>
<b>3.2</b>	<b>AJUSTE DOS MODELOS DE EFEITO MISTO PGM1 E PGM2.....54</b>
<b>3.3</b>	<b>PRECISÃO E INCERTEZA DOS MODELOS PGM1 E PGM2.....56</b>
<b>3.4</b>	<b>ESTIMATIVA DE BIOMASSA AO NÍVEL DE ÁRVORE POR PGM2.....57</b>
<b>4</b>	<b>DISCUSSÃO.....58</b>
<b>5</b>	<b>CONCLUSÃO.....59</b>
<b>REFERÊNCIAS.....</b>	<b>60</b>

## PRIMEIRA PARTE – REVISÃO BIBLIOGRÁFICA

### 1 INTRODUÇÃO

A densidade básica (DB) é um dos fatores mais importantes para o entendimento da qualidade da madeira de diferentes espécies florestais e, principalmente, para plantios dos gêneros *Eucalyptus* e/ou *Pinus sp.* Portanto, é de extrema relevância a quantificação dessa variável, que é utilizada para o entendimento da biomassa produzida e para nortear os cálculos de rendimentos das indústrias florestais. Embora seja uma variável de fácil obtenção pelos métodos convencionais, há um gasto significativo de tempo e recursos (tanto humanos como financeiros) para a sua quantificação em laboratório. Nessas amostragens, a escala é muitas vezes reduzida, não suprimindo a demanda operacional das empresas de celulose/papel, carvão, painéis ou até mesmo produtos sólidos de madeira. Isso afeta também pesquisas conduzidas pelos setores de melhoramento genético das empresas, que necessitam cada vez mais de agilidade na geração de informações para avaliações precoces de novos clones.

A quantificação da DB é complexa, devendo levar em consideração sua alta variabilidade e interações de fatores, como, por exemplo, idade, material genético, sítios, disponibilidade hídrica/precipitação, altitude, declividade, luminosidade, etc. Além disso, há a combinação desses fatores com a composição química, física e anatômica da madeira, dentro de um único indivíduo, quer seja ao longo do fuste (sentido longitudinal), como ainda na direção medula-casca (sentido radial). Nos últimos anos, surgiram diversos procedimentos e equipamentos para sua quantificação, como uma alternativa às caras metodologias tradicionais destrutivas. Nesse sentido, uma grande quantidade de equipamentos vem sendo utilizada, como é o caso dos Espectrômetros de Infravermelho Próximo (NIR), Resistógrafos e Pilodyns, que oferecem maior agilidade na coleta de informações, baixo custo de amostragem e facilidade de manuseio.

Apesar dos avanços pontuais na sua quantificação em campo, ainda há um problema relacionado à escala empresarial e à sua dimensionalidade operacional para um grande volume de dados. Uma possibilidade a ser empregada condiz com a estruturação de uma modelagem da própria variável, envolvendo uma rede amostral ótima, por meio de um agrupamento de características semelhantes via modelos de efeito misto. Esse caminho confere maiores flexibilidade por meio da generalização, possibilitando uma estimativa coerente da variável de interesse, mesmo empregando um conjunto reduzido de observações. A ideia é simétrica ao observado nos trabalhos de Scolforo et al. (2019); Bronisz e Mehtätalo

(2020), porém, estes adotaram a estimativa de altura total e volume individual das árvores, por meio da calibração específica de coeficientes, possibilitada pela matriz de variância/covariância da modelagem de efeito misto.

Dada a importância da quantificação da DB e sua relação com a qualidade da madeira, seja vinculada aos produtos finais do processamento (celulose, papel, carvão, painéis, etc.), ou para a avaliação precoce de materiais genéticos, tornam-se necessários estudos que tornem o processo de estimativa da variável cada vez mais rápido e confiável, visto que os gestores florestais utilizam a informação para a tomada de decisões. Diante da problemática abordada, é de suma importância a criação de ferramentas práticas que permitam fazer inferências acuradas e ágeis sobre a variável, principalmente para o *Eucalyptus sp.*, em virtude da rápida evolução do melhoramento genético e de seus altos custos.

## **2 REVISÃO DE LITERATURA**

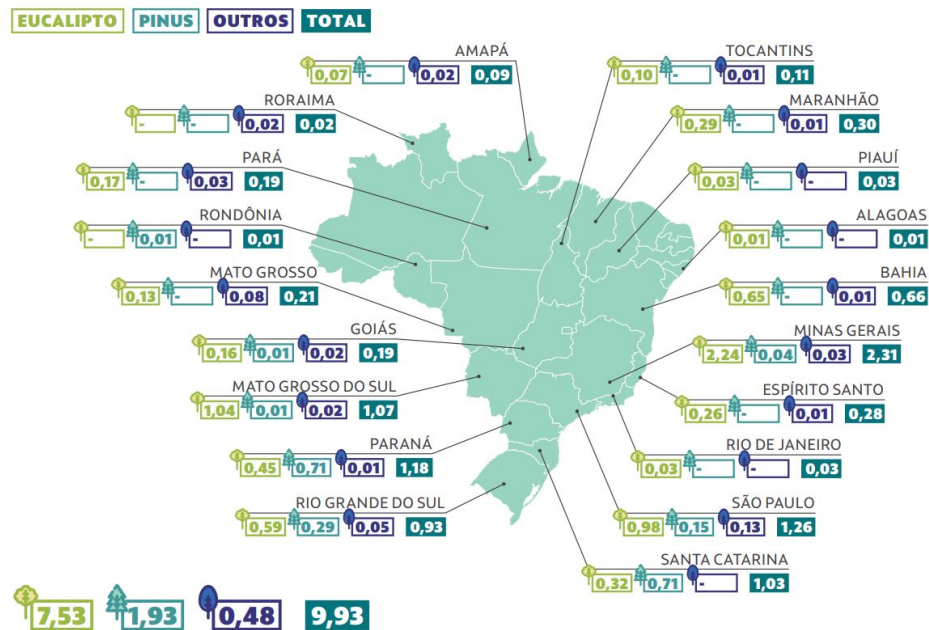
### **2.1 Setor Florestal e a sua importância**

Embora os plantios florestais representem uma pequena fração da área coberta por florestas no mundo, eles têm uma crescente importância nos dias atuais, visto as questões ambientais e políticas atreladas à busca por alternativas que visem a diminuição de emissão de gases do efeito estufa e redução da utilização de combustíveis fósseis (FAO, 2020). Dessa forma, o eucalipto, por exemplo, surge como uma alternativa viável para as mais variadas finalidades e até mesmo para o setor de geração de energia por meio da biomassa. A árvore de eucalipto é uma espécie com distribuição global e ampla destinação, sendo utilizada na produção de carvão vegetal, celulose, papel, lenha, óleos essenciais, madeira para construção, entre outras finalidades (FERRAZ FILHO, SCOLFORO e MOLA-YUDEGO, 2014; MAGNANO et al., 2016). A relevância do gênero ressalta a necessidade de métodos adequados de monitoramento e medição (RAIMUNDO et al., 2017).

A superfície terrestre é recoberta, atualmente, por 4,06 bilhões de hectares de florestas, o que representa 31% do globo, e mais da metade desse total (54%) está concentrada em apenas 5 países (Federação Russa, Brasil, Canadá, Estados Unidos da América e China). As florestas plantadas contribuem com 7% desse total (FAO, 2020). Segundo o relatório anual da Indústria Brasileira de Árvores (IBÁ, 2022), o setor de árvores plantadas no Brasil alcançou números bastante expressivos em relação a investimentos, empregos e receita bruta gerados a partir dos plantios e novos investimentos. Em 2021, a área de florestas plantadas atingiu a

marca de 9,93 milhões de hectares, comumente em terras previamente degradadas (benefício sustentável do setor), sendo que 7,53 milhões de hectares desse total referem-se ao gênero *Eucalyptus*. Com destaque para os Estados de Minas Gerais, São Paulo e Mato Grosso do Sul, que juntos somam aproximadamente 60% desse total (Figura 1).

Figura 1 - Distribuição das florestas plantadas no Brasil.



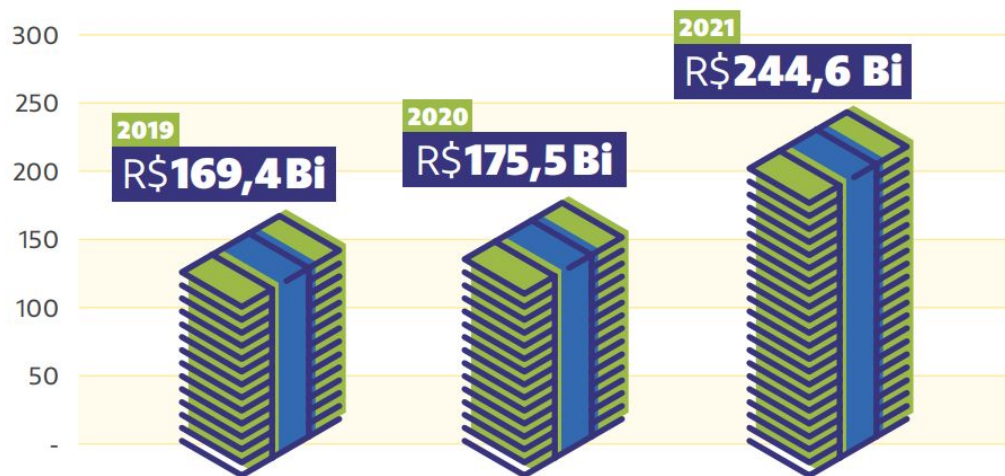
Fonte: Extraído de Relatório Anual IBÁ 2022 – FGV IBRE, Ibá & Canopy.

As florestas plantadas geraram aproximadamente mais de 2 milhões de postos de trabalho diretos e indiretos no setor, o que é de extrema importância para a evolução social de regiões distantes de grandes centros (MOREIRA, SIMIONI e OLIVEIRA, 2017). Além disso, o setor impulsiona a economia local em mais de 1.000 municípios em todo o país, muitos deles situados em regiões distantes dos grandes centros industrializados. Isso acaba impactando positivamente o Índice de Desenvolvimento Humano (IDH) desses municípios, que conseqüentemente apresentam um IDH mais elevado do que a média dos estados nos quais estão inseridos.

Outro ponto importante a ser mencionado é que os pequenos produtores também se beneficiam desse cenário por meio da prática do fomento florestal. A prática desse tipo de modalidade de subsídio fez com que o número de agricultores envolvidos no setor aumentasse para 2 milhões, permitindo a diversificação do uso de suas terras e a obtenção de renda extra com o plantio de florestas, prática que auxilia amplamente no desenvolvimento de áreas rurais (CUNHA et al., 2019).

Considerando o contexto atual do setor, permitiu às indústrias do ramo florestal que fossem instauradas as maiores produções históricas das cadeias de celulose, papel e painéis de madeira em 2021, resultando em um crescimento expressivo de 7,5% (contrastado ao mesmo período em 2020) e atingindo uma receita bruta de R\$ 244,6 bilhões (Figura 2). O panorama atual é extremamente benéfico para o desenvolvimento social, econômico e sustentável do Brasil (RIBASKI, 2018) no curto, médio e longo prazo, visto que há previsão de um grande investimento (R\$ 60,4 bilhões) destinado a florestas, pesquisa e desenvolvimento, e novas unidades industriais entre os anos de 2022 e 2028. Diante desse cenário extremamente positivo, é de suma importância que os estudos e pesquisas no âmbito de florestas plantadas continuem avançando, fornecendo informações cada vez mais precisas e em um curto período de tempo para embasar as decisões dos gestores florestais (THIERSCH et al., 2006), o que garantirá um crescimento contínuo e sustentável, promovendo benefícios tanto para o setor quanto para a sociedade como um todo.

Figura 2 - Valor da produção do setor florestal brasileiro.



Fonte: Extraído de Relatório Anual IBÁ 2022 – FGV IBRE & Ibá.

## 2.2 Plantios comerciais de *Eucalyptus*

No Brasil, as plantações de *Eucalyptus sp.* têm desempenhado um papel importante nas indústrias de base florestal, principalmente devido ao rápido crescimento/desenvolvimento das árvores e à facilidade de implantação em grandes extensões de terras, além da adaptação do gênero às diferentes condições climáticas impostas pelo cenário brasileiro (COELHO et al., 2020; IBÁ, 2017). Outro fator relevante para isso é o

constante melhoramento genético conduzido por pesquisadores, universidades e empresas do setor florestal, resultando em plantios clonais que possibilitam a produção de madeira mais homogênea, refletindo em maior qualidade no produto final (JARDIM et al., 2017).

O Brasil tem se destacado como um dos principais produtores mundiais de celulose (o 2º maior, atrás dos Estados Unidos da América) (IBÁ, 2022), e a estrutura e anatomia da madeira do Eucalipto são relevantes nesse processo (FOELKEL, 2009), além de que o país é a nação mais eficiente em termos de produção de madeira proveniente de plantios de eucalipto no mundo (MACHADO et al., 2014). As empresas de celulose e papel buscam matérias-primas com características específicas para aumentar a produção e melhorar as propriedades do produto final (Figura 3), sendo o Eucalipto uma fonte importante de fibra curta para esse fim (SILVERIO et al., 2007). Todas essas características, aliadas à ampla área plantada de *Eucalyptus*, fazem com que seja necessária uma correta mensuração e quantificação da volumetria e, também, das características relacionadas à qualidade da madeira dos plantios, visto que há uma correlação direta com o produto final produzido pelas unidades fabris (FANTUZZI NETO, 2012).

Figura 3 – Ciclo do *Eucalyptus* para indústria de celulose (Do plantio a produção de celulose).



Fonte: Do Autor.



Considerando a expansão dos plantios florestais no cenário brasileiro, sabe-se do papel fundamental que a matéria-prima madeira exerce nas indústrias, e nesse contexto, a correta mensuração dos recursos é crucial para um planejamento correto da produção de maneira eficaz e sustentável (THIERSCH et al., 2006). Nesse contexto, a correta mensuração dessas plantações revela-se crucial para um planejamento de produção eficaz e sustentável. Através de técnicas de mensuração precisas e atualizadas, como monitoramento via sensoriamento remoto e análise geoespacial, é possível obter insights valiosos sobre o crescimento, a saúde das árvores e a produtividade das áreas cultivadas. Esses dados fornecem a base necessária para decisões informadas, como a determinação do momento ideal para a colheita, a alocação eficiente de recursos e a implementação de práticas de manejo que maximizem tanto a produção quanto a conservação dos ecossistemas circundantes. Assim, a correta mensuração dos plantios de eucalipto não apenas impulsiona a eficiência operacional das indústrias, mas também contribui para a busca contínua de um equilíbrio entre o desenvolvimento econômico e a preservação ambiental, além de municiar os gestores florestais para tomada de decisões assertivas.

As medições dos parâmetros biométricos dos plantios são elementos cruciais para o planejamento e gestão das plantações, o que permite a geração de estimativas de volume de madeira, índices de sítio dos povoamentos, além de cálculos da quantidade de carbono (CO<sub>2</sub>) contido nos plantios florestais. Porém, há também a necessidade de se mensurar de alguma forma características de qualidade da madeira, como, por exemplo, a densidade básica (KWAK et al., 2007; MOHAN et al., 2019), o que permitirá um melhor planejamento das produções industriais de setores como o de celulose e papel.

Atualmente, as áreas de inventário e biometria florestal de empresas dos mais diversos ramos florestais (Biomassa, Carvão, Celulose, entre outras), têm como um de seus principais produtos a entrega da informação de estoque volumétrico dos plantios, porém uma informação que melhor subsidiaria o planejamento de produção das fábricas e também o plano estratégico dessas indústrias seria a quantificação da biomassa (THIERSCH et al., 2006; DIAS et al., 2017). Para tal fim, visando a obtenção da massa seca das árvores, há a necessidade da mensuração de uma das características químicas/físicas mais importantes da madeira, que é a densidade básica, conforme mencionado por Fundova, Funda e Wu (2018).

A densidade é tida por Dias et al. (2017) como atributo mais significativo pelo fato de ser facilmente relacionada com o desempenho das operações florestais e processos industriais. Sendo assim, a quantificação dessa variável é de vital importância para o entendimento da qualidade da madeira e para as suas aplicações práticas. Tendo em vista a relevância dessa

informação, se faz necessário uma melhor mensuração e até mesmo a estimativa da variável de interesse. Para isso, um recurso fundamental no gerenciamento de dados referentes aos plantios é o inventário florestal (CAMPOS e LEITE, 2013), que consiste em recorrentes medições em períodos de tempo pré-definidos, permitindo monitorar as florestas com o intuito de evidenciar mudanças/diferenças nas florestas conforme o passar do ciclo (SCOLFORO et al., 2017).

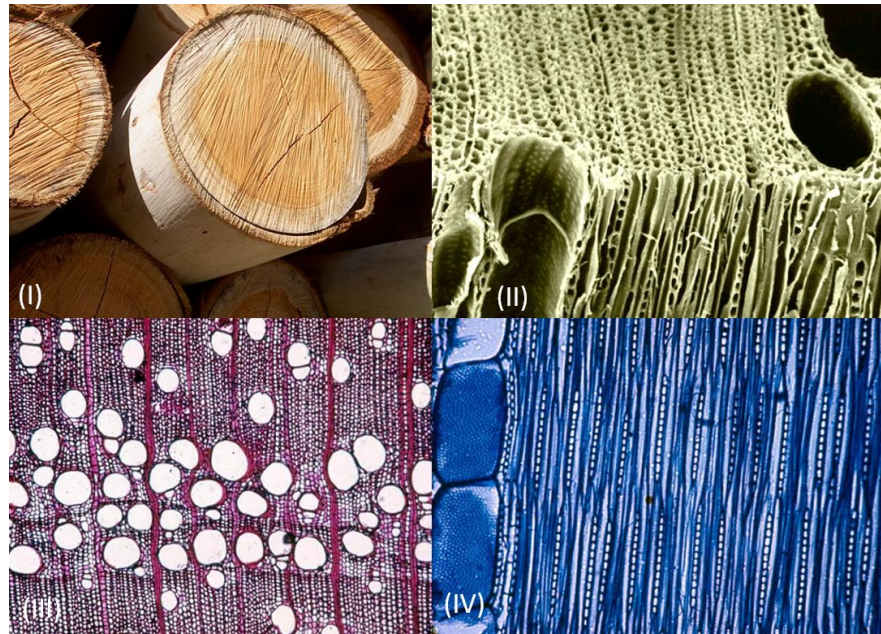
### **2.3 Densidade Básica da Madeira**

O entendimento da qualidade da madeira está diretamente vinculado à determinação da densidade básica (OLIVEIRA, HELLMEISTER e TOMAZELLO FILHO, 2005), que, por vezes, foi até relacionada como um índice de qualidade, conforme mencionado pelos autores Foelkel, Mora e Menochelli (1992), pela facilidade de determinação quando comparada às demais variáveis, pela sua ampla utilização e excelente relação com as diversas utilizações da madeira, podendo ser associada diretamente com questões técnicas e econômicas. Baseando-se na premissa de que a densidade básica da madeira é uma das variáveis mais importantes para diversas finalidades, como, por exemplo, um indicativo da quantidade de carbono que uma árvore utiliza em seu processo de formação, variando conforme idade, espécie e até longitudinalmente e radialmente em uma mesma árvore (CHAVE, 2002). É sabido que a densidade básica da madeira é uma variável de fácil obtenção, porém de difícil estimativa devido ao fato de sua variação estar relacionada aos mais diversos fatores, tanto intrínsecos à composição anatômica da madeira e genéticos (tamanho da parede celular, teores de celulose e lignina, diferentes espécies e variedades de materiais genéticos, entre outros), como também a fatores relacionados aos diferentes tipos de manejo empregados nos plantios (diferentes espaçamentos e sítios, além dos mais variados tratamentos silviculturais), conforme verificado por Lopes e Garcia (2002).

Os autores Foelkel, Brasil e Barrichelo (1971), Hansen (2000) e Lopes e Garcia (2002) definem a densidade como uma variável extremamente complexa que é o resultado de uma combinação de uma infinidade de fatores, podendo ser eles: dimensões da fibra (espessura da parede celular), comprimento e largura das células, tamanho/volume de vasos e parênquimas, proporção entre lenho inicial e tardio, além da conformação dos diferentes componentes anatômicos da madeira e dos teores de celulose, lignina e extrativos (Figura 4). Dias et al. (2017) a caracterizam como uma correlação de fatores químicos e anatômicos da madeira,

além disso, ressaltam também a sua significância nos processos da indústria de celulose de polpação e como indicador de qualidade de polpa.

Figura 4 - Tora de madeira de *Eucalyptus* (I) e sua composição (Cortes Longitudinais II e IV e Transversais II e III - Microscópio).



Fonte: Galeria de fotos Site *Eucalyptus* Online Book & Newsletter (Celso Foelkel).

A variabilidade da madeira deriva de um sistema biológico complexo formado pela interação entre sua composição química, física e mecânica. Ela pode variar entre indivíduos de espécies semelhantes ou não, e até dentro de uma mesma árvore. Essa variação ocorre, possivelmente, devido às mudanças acarretadas pelo envelhecimento e pelas diferenças de condições ambientais pelas quais a árvore foi exposta (OLIVEIRA e SILVA, 2003). A idade das florestas é um fator que exerce interferência considerável na variação da densidade básica, visto que variadas transformações são percebidas conforme os plantios ficam mais velhos, principalmente devido às variações nas composições químicas, físicas e anatômicas da madeira (CARNEIRO et al., 2014).

De acordo com Foelkel (2009 e 2015), os principais motivos pelos quais ocorrem variações na qualidade de uma madeira e, conseqüentemente, na densidade, estão relacionados à diferenciação entre espécies e materiais genéticos (clones e seminais), idade das florestas, posição da amostragem dentro de uma árvore (variação longitudinal ao longo do comprimento e radial entre a medula e a casca) e também pelas condições ambientais que abrangem os diferentes plantios. Além disso, é assegurado que a densidade está diretamente

relacionada a fatores como solo, clima, praga, doenças, declividade, déficit hídrico e luminosidade. Isso vai ao encontro de um dos pontos relatados no estudo proposto por Stape et al. (2010), que menciona que o índice de sítio influencia tanto na produtividade das florestas plantadas quanto nas propriedades da qualidade da madeira.

Uma possível finalidade que pode ser citada para a variável densidade é a sua utilização como parâmetro para a estimativa de impregnação de reagentes químicos em cavacos no processo de polpação celulósica, além de ser considerada como indicador de consumo específico da madeira (QUEIROZ et al., 2004). Os autores Gomide, Fantuzzi Neto e Regazzi (2010) identificaram uma forte correlação negativa entre a densidade básica e o consumo específico de madeira para amostras de *Eucalyptus*. No mesmo estudo mencionam que maiores valores de densidade estão atrelados a menores consumos específicos, favorecendo dessa forma o processo de produção de celulose pelas indústrias.

Uma possível finalidade que pode ser citada para a variável densidade é a sua utilização como parâmetro para a estimativa de impregnação de reagentes químicos em cavacos no processo de polpação celulósica, além de ser considerada como indicador de consumo específico da madeira (QUEIROZ et al., 2004). Os autores Gomide, Fantuzzi Neto e Regazzi (2010) identificaram uma forte correlação negativa entre a densidade básica e o consumo específico de madeira para amostras de *Eucalyptus*. No mesmo estudo, mencionam que maiores valores de densidade estão atrelados a menores consumos específicos, favorecendo assim o processo de produção de celulose pelas indústrias. Chave et al. (2006) concluem em seu estudo que um melhor entendimento/explicação da variabilidade da densidade básica da madeira é de fundamental importância para estimativas mais precisas da biomassa das florestas. Jati, Fearnside e Barbosa (2014) fazem menção ao fato de que a estimativa desta variável, relacionada à qualidade da madeira, favorece o cálculo de biomassa e carbono das espécies arbóreas amazônicas. Ambos os estudos expressam a relevância que trabalhos que procuram entender essa variação têm, pelo fato de reforçar a compreensão da dinâmica de estoques de carbono. Segundo Foelkel, Brasil e Barrichelo (1971), as empresas de celulose e papel devem considerar a densidade como um indicador pelos seguintes motivos: (a) fator importante para a transformação da variável volume em peso seco para o controle dos rendimentos industriais, (b) elemento fundamental que deve ser considerado para a quantificação da produção de celulose tanto nas fábricas como o estoque de celulose em campo – Incremento Médio Anual transformado em Incremento Médio Anual de Celulose ou IMACEL, (c) característica que facilita a verificação da uniformidade da qualidade da madeira, visando um produto final padronizado, (d) parâmetro que direciona quanto à

velocidade de impregnação do licor de cozimento na madeira, que está atrelado ao ritmo de deslignificação, (e) relação inversa direta entre tempo de refinação de celulose e densidade da madeira, considerando coníferas, e, por fim, (f) relações diretas com propriedades físico-mecânicas da celulose – resistência ao rasgo e à ruptura. Dada toda a importância desta variável, é necessário o entendimento dos métodos de determinação, sendo eles convencionais ou não-destrutivos/semi-destrutivos (GOUVÊA et al., 2011b; FOELKEL, BRASIL e BARRICHELO, 1971; DIAS et al., 2017; FUNDOVA, FUNDA e WU, 2018).

### **2.3.1 Métodos não-destrutivos para determinação da densidade básica da madeira**

Como alternativa aos métodos tradicionais de determinação da densidade, temos os métodos não-destrutivos ou semi-destrutivos, que consistem na retirada de uma porção/amostra das árvores para avaliação posterior em laboratório, sem a necessidade do abate, ou por meio da realização de uma medição indireta ao longo do fuste. Essa metodologia tem como principais motivadores a otimização de tempo e recursos que seriam dispendidos derrubando árvores e levando-as até um laboratório para a realização dos procedimentos convencionais, permitindo uma avaliação em campo do material (GOUVÊA et al., 2011b). Um dos primeiros métodos semi-destrutivos relatados surgiu no século XVIII, conforme estudo proposto por Pressler (1866, citado por GAO et al., 2017), em que há a retirada, por meio de uma broca, de uma quantidade de material lenhoso da árvore em pé. No entanto, a etapa de levar as amostras até o laboratório ainda não é eliminada, acarretando um elevado período de tempo para a obtenção da variável de interesse.

Desde então, novos equipamentos surgiram e muitos trabalhos foram conduzidos com o intuito de criar metodologias que permitissem a pesquisadores/empresas do setor florestal fazer inferências sobre a qualidade da madeira de forma mais precisa e em um curto período de tempo. Abaixo estão listados alguns dos principais equipamentos que permitem uma avaliação não-destrutiva da densidade. Os mais comuns encontrados na literatura são: o Pilodyn (durômetro de impacto) (GREAVES et al., 1996; WU et al., 2010; KIEN et al., 2008), o Resistógrafo (DIAS et al., 2017; GOUVÊA et al., 2011b; LIMA et al., 2006) e a tecnologia NIR (Espectrômetros de Infravermelho Próximo) (MAGALHÃES et al., 2006; HEIN, 2010; SCHIMLECK et al., 2018).

Há um número considerável de trabalhos onde se tem como objetivo principal a comparação entre os 3 aparelhos, visando encontrar qual tem uma melhor estimativa da densidade real da madeira (GOUVÊA et al., 2011b; FUNDOVA, FUNDA e WU, 2018), ou

estudos que visam verificar o quão relacionável a medida de um dos aparelhos é em função da variável de interesse (HEIN et al., 2009; NEVES et al., 2011). A utilização em larga escala desses equipamentos, aliada à utilização de boas técnicas de modelagem (como redes neurais artificiais, modelos de efeito misto, algoritmos, etc.), é uma prática pouco difundida em empresas florestais.

Os Espectrômetros de Infravermelho Próximo (NIR) (Figura 5 – I) são aparelhos que permitem, por meio da avaliação de amostras de serragem, a análise de propriedades físico-químicas da madeira (Figura 4). Entre elas, pode-se analisar também a densidade básica de forma rápida e sem viés de operadores, que muitas vezes influencia diretamente na qualidade da estimativa gerada. Seu princípio de funcionamento é baseado na reflectância da radiação eletromagnética (a partir de comprimento de onda conhecido, na faixa do espectro eletromagnético do infravermelho próximo – 800 a 2.500 nm) que incide nas amostras de serragem de madeira. Parte dessa radiação é absorvida pelos grupos funcionais que contêm partículas de carbono, por exemplo. O restante da radiação que não foi absorvida será refletida e quantificada pelos detectores do espectrômetro, resultando em um espectro de absorbância conforme descrito por Lazzarotto (2016), onde podem-se verificar diferenças entre as amostras de materiais genéticos distintos, porções diferentes da árvore, e uma infinidade de fatores que influenciam diretamente na densidade. Como principais motivos para a utilização dessa tecnologia, podem-se citar a agilidade, baixo custo das análises e a facilidade de replicação de resultados/testes. No entanto, há alguns pontos que devem ser considerados, como a dificuldade na calibração do aparelho para a realização das amostras, além de ser uma técnica extremamente sensível a variações sazonais (SCHIMLECK et al., 2019).

Figura 5 - Equipamentos não-destrutivos para determinação da densidade básica da madeira (I) Espectômetro Multiuso MPA II Bruker®, (II) Resistograph® e (III) Pilodyn®.



Fonte: Do Autor (2023).

Outro equipamento comumente utilizado em muitos estudos relacionados à determinação da densidade básica é o Resistógrafo (Figura 5 - II). Seu funcionamento consiste na penetração de uma agulha/broca de metal na madeira, com uma força constante, gerando assim um gráfico de perfil, onde é possível verificar a resistência à penetração, correlacionável com a densidade das árvores, conforme concluído por Gouvêa et al. (2011b). Os principais benefícios na utilização do Resistógrafo são o baixo custo de utilização em campo, aliado à rapidez na geração das medições, além de possuir a captura automatizada das leituras, o que permite um controle de interferências humanas nos resultados, como citado por Schimleck et al. (2019). Deve-se apenas ponderar que, para sua utilização, há a necessidade de utilização de softwares/aplicativos para a leitura dos dados, visando correlacioná-los com a densidade básica, e há uma certa dificuldade em encontrar balizadores para o processo.

Praticamente com o mesmo princípio do Resistógrafo, temos o Pilodyn (Figura 5 - III). Um equipamento manual que apresenta um sistema de molas, que quando acionado penetra uma agulha de metal com uma força constante na madeira. Por meio de uma escala presente no aparelho, é possível realizar a leitura, em milímetros, de quanto a agulha penetrou na árvore, correlacionando a profundidade de penetração da agulha com a densidade básica da madeira. Algumas críticas na utilização do aparelho visando a correlação com a densidade se dão pela grandeza da escala das medições (1 milímetro lido incorretamente pode ter um alto impacto nos resultados das pesquisas) e pelo aparelho proporcionar uma medida superficial do fuste. Por outro lado, é um equipamento de fácil manuseio e não precisa de mão-de-obra qualificada para a utilização, sendo possível a incorporação nas atividades usuais de

inventário florestal e proporcionar resultados satisfatórios, como encontrado por Thiersch et al. (2006).

## 2.4 Modelagem preditiva de atributos florestais

Modelos preditivos são de fundamental importância na estimativa de variáveis dendrométricas no contexto da biometria florestal, visto que, na maioria das vezes, o levantamento em campo dessas informações é dispendioso e envolve coleta de amostras de forma destrutiva, como é o caso da densidade básica da madeira (HUSCH, BEERS e KERSHAW, 2002). Para o desenvolvimento de modelos regionais, há a necessidade de obtenção de grandes quantidades de amostras de diferentes povoamentos, visando garantir a representatividade dos dados para o ajuste do modelo. Essa abordagem é crucial para assegurar que as previsões sejam precisas e confiáveis em diversas situações, porém, muitas vezes, por conta da quantidade finita de recursos, isso não é viável.

Uma das etapas mais importantes para a construção de um modelo matemático, capaz de explicar significativamente uma variável de interesse, é a seleção dos parâmetros que irão compor essa equação. Nessa etapa, há a possibilidade, por meio de variados métodos, de escolha de variáveis que irão favorecer um balanceamento adequado entre a quantidade de informações que serão aproveitadas da base de dados e a variância das estimações. Além disso, por meio dessa seleção, pode-se mitigar inconvenientes relacionados à multicolinearidade. Um dos métodos de seleção de variáveis mais comuns é o de regressão *Stepwise*, que consiste na combinação de outros dois métodos denominados *seleção Forward* e *eliminação Backward*. Este método fundamenta-se na adição e remoção de variáveis a partir de um modelo nulo, até que se encontre um modelo pertinente para a explicação da situação avaliada. O critério de inclusão e/ou exclusão das variáveis preditoras é baseado na correlação que estas têm com a variável a ser predita, por meio do teste estatístico do valor F ou pelo critério de avaliação do AIC (*Akaike Information Criterion*) (MONTGOMERY, PECK e VINING, 2012). Há uma quantidade significativa de críticas relacionadas a essa técnica de seleção, pelo fato da instabilidade na escolha de um melhor modelo para uma mesma base de dados, além de quanto maior for a quantidade de variáveis a serem testadas, mais árdua será a sua utilização, por mais que este método seja um dos menos trabalhosos. Um outro ponto negativo do método é que não é possível a avaliação de multicolinearidade das variáveis selecionadas por ele.



Os autores Watzlawick, Kirchner e Sanquetta (2009) utilizaram a análise de regressão múltipla de *Stepwise* para a seleção das variáveis que melhor estimassem a biomassa arbórea e o carbono orgânico em florestas de Araucárias de forma não destrutiva. As variáveis testadas foram obtidas por meio do processamento de imagens do satélite IKONOS II. No estudo, menciona-se que o método utilizado é satisfatório e que a seleção de variáveis permitiu bons ajustes de modelos para a mensuração da biomassa. Orellana e Figueiredo (2017) também utilizaram em seu estudo o método de regressão de *Stepwise* para a seleção de variáveis que melhor estimassem os parâmetros utilizados no ajuste de modelos de projeção da distribuição diamétrica em florestas nativas com a função de *Weibull*. Os mesmos concluem que o método de predição de parâmetros surge como uma boa possibilidade e se une a outras alternativas conhecidas como a Matriz de Transição e/ou a Razão de Movimento.

Outro método comumente utilizado, visando a escolha de variáveis em situações onde têm-se um elevado número de parâmetros coletados, ausência de observações e até mesmo casos de alta correlação entre as variáveis, é a regressão PLS (*Partial Least Squares – Regressão dos Mínimos Quadrados Parciais*). A técnica consiste em uma análise multivariada que capta a correlação existente entre variáveis dependentes e independentes em uma base de dados e procura reduzi-la a um pequeno número de combinações lineares, permitindo uma melhor manipulação dos dados e, conseqüentemente, a seleção facilitada das variáveis fundamentais para o modelo (ZIMMER e ANZANELLO, 2014; ANDERSEN e BRO, 2010).

Hein, Lima e Chaix (2010) simularam em seu estudo regressões de mínimos quadrados parciais baseadas em dados coletados por meio da espectroscopia de infravermelho próximo para verificar a influência da utilização desses tratamentos matemáticos, evidenciar amostras outliers e para a seleção de comprimentos de onda que melhor auxiliem na modelagem da densidade básica da madeira e do módulo de elasticidade em testes de compressão das fibras de corpos de prova de *Eucalyptus*. Os mesmos concluem que a utilização do método PLS é satisfatória para relacionar os dados de medições do NIR com as variáveis de interesse, havendo um expressivo ganho nos modelos preditivos testados.

A metodologia Lasso (*Least Absolute Shrinkage and Selection Operator*) é também uma relevante proposta para a seleção de variáveis visando o ajuste de um modelo matemático em cenários de adversidade de vastas bases de dados, nos quais os estimadores de mínimos quadrados apresentam limitações, sugerida por Tibshirani (1996), por meio de técnicas de Aprendizado de Máquina (*Machine Learning*) (MURPHY, 2012). Esse método é possibilitado por meio da minimização dos quadrados dos erros do modelo com a utilização de um parâmetro de ajustes (*Tuning parameter*), seguindo a premissa de que este deve ser

maior que o somatório dos valores absolutos dos coeficientes do modelo. Dessa maneira, os parâmetros não nulos são considerados mais significativos e ideais para a modelagem, conforme menciona Altoé (2017) em seu estudo.

Fiorentin et al. (2020) utilizam a metodologia Lasso em comparação com alguns outros métodos, visando a seleção de variáveis para a criação de um modelo para a estimativa de probabilidade de sobrevivência em plantios de *Pinus Taeda*. Os autores esperavam inicialmente que os métodos de penalização, como é o caso do *Lasso*, fossem mais apropriados para a seleção de variáveis que apresentam correlação; no entanto, notou-se que esta metodologia não teve efeito satisfatório nos modelos propostos. Verificaram também que o método de *Stepwise* foi o que apresentou melhor desempenho para a seleção de variáveis.

Em um estudo que tinha como objetivo desenvolver um modelo preditivo relacionando os efeitos das mudanças climáticas com o crescimento em altura de *Pinus Taeda* no sudeste dos Estados Unidos, Farjat et al. (2015) também utilizaram a regressão Lasso como um de seus métodos de seleção de variáveis para a composição do modelo. É mencionado que esta metodologia se destaca perante as demais pelo fato de selecionar uma menor quantidade de variáveis preditoras perante os demais métodos testados.

Também pode-se lançar mão de metodologias que se utilizam de inteligência computacional para a seleção de melhores variáveis preditoras, como, por exemplo, os algoritmos genéticos (AG), que se tornam viáveis alternativas devido à sua eficiência, versatilidade e robustez (COSTA FILHO e POPPI, 2002). Este tipo de técnica é baseado em simulações matemáticas da teoria de Darwin relacionada ao processo de evolução biológica das espécies por meio de fatores como a herdabilidade, seleção natural, recombinação e mutação. Esse tipo de meta-heurística permite a resolução de problemas de otimização por meio de uma abordagem mais generalista, entregando soluções satisfatórias para a grande parte das situações abordadas, reduzindo significativamente o tempo de processamento e esforço computacional. Alguns outros métodos de inteligência computacional que possuem um funcionamento semelhante e apresentam resultados satisfatórios são as Redes Neurais Artificiais (RNA) (ASILTÜRK e ÇUNKAŞ, 2011) e o *Random Forest* (Breiman, 2001).

Miranda et al. (2022) propõem em seu trabalho uma abordagem híbrida utilizando o algoritmo genético para a seleção de variáveis preditoras (ambientais e dendrométricas), conciliado com a metodologia de aprendizado de máquina (*Random Forest*) para modelar a relação hipsométrica de árvores individuais. Os autores concluem que o método híbrido proposto apresenta resultados satisfatórios comparáveis a modelos de estimativa de altura

tradicionais, permitindo a adição de variáveis ambientais que auxiliaram na explicação da variável resposta.

## **2.5 Inteligência artificial e aprendizado de máquina**

As técnicas de inteligência artificial (IA) vêm tendo uma grande relevância em aplicações na engenharia florestal, ocasionada, principalmente, pelas inúmeras vantagens quando comparada aos modelos de regressão tradicionais. Haykin (2001) apontou, em seus estudos, um desempenho superior da IA e ressaltou a capacidade de resolução de problemas complexos com facilidade. Como uma das principais vantagens proporcionadas pela utilização da inteligência artificial, destaca-se a necessidade de uma quantidade diminuta de amostras, quando comparada aos métodos tradicionais de regressão, para um mesmo nível de precisão das estimativas geradas (BINOTI, BINOTI e LEITE, 2013). Além disso, a IA não depende das suposições e premissas desejadas no processo de regressão, o que dá maior flexibilidade no tratamento de dados complexos e não-lineares, conforme descrito pelos autores Cordeiro et al., 2015. Isso torna esse tipo de aplicação mais eficiente em termos de recursos financeiros e de tempo, sendo assim, é uma alternativa para empresas do setor florestal pela constante busca na redução de orçamentos e aumento na precisão das estimativas. Conclui-se dessa forma que a IA tem um grande potencial nas mais variadas linhas de pesquisa do setor florestal, visto que concilia precisão das estimativas, baixa amostragem e capacidade para lidar com dados altamente complexos e suas relações, o que torna a abordagem promissora para as áreas de manejo e mensuração de recursos florestais.

Em função do aumento recente e expressivo do número de observações, complexidade das problemáticas e tamanho de bases de dados nos dias atuais dos mais variados campos de pesquisa, a necessidade de criação de metodologias e ferramentas capazes de entender os problemas, as relações entre variáveis independentes e convergir para melhores resultados e estimativas ficou mais evidente. Baseado nesse contexto, o aprendizado de máquina surge como uma robusta ferramenta, que a princípio era utilizada para identificação de padrões em bases de dados, tornando possível realizar previsões futuras por meio de tendências observadas previamente (LIDBERG, NILSSON e ÅGREN, 2019).

As técnicas de aprendizado de máquina são, na maioria dos casos, inspiradas na teoria evolutiva, leis da física, comportamento de animais, inteligência humana e também aspectos biológicos, podendo ser classificadas em 5 áreas. Sendo elas: Redes Neurais Artificiais, Computação Evolutiva, Inteligência de Enxames, Sistemas Imunológicos Artificiais e

Sistemas *Fuzzy* (KONAR, 2005). Com o auxílio do avanço das tecnologias e capacidade de processamento das máquinas, a utilização das técnicas de aprendizado de máquina tem sido cada vez mais utilizada nas mais diversas áreas de conhecimento e tem se mostrado muito eficazes na resolução de problemas extremamente complexos, como por exemplo, otimização de sistemas de energia (RAHMAN E MOHAMAD-SALEH, 2018) e aplicação financeira (NGAI et al., 2009). Dessa forma, pode-se ressaltar também o potencial desse tipo de ferramenta também no contexto florestal, gerando insights a respeito das informações dos povoamentos e auxiliando na tomada de decisão.

Estimativas geradas a partir de métodos metaheurísticos tendem a ser menos confiáveis em comparação com os resultados obtidos pelas técnicas de modelagem tradicionais. Isso se deve ao fato de que esses algoritmos geram soluções diferentes dependendo da semente aleatória usada no início do processo, o que requer um procedimento iterativo intenso (CHUMNEY e SIMPSON, 2006). A principal lógica, na qual as metaheurísticas se embasam, é a de ajustar as estimativas iniciais dos parâmetros até que a métrica de erro mínimo (PAYANDEH, 1983). Em outras palavras, esse tipo de metodologia lança mão de um processo de busca iterativa para encontrar os valores dos parâmetros que melhor se ajustam aos dados e minimizam o erro em relação à solução ideal, considerando que essa solução ideal é desconhecida e não pode ser obtida de forma direta. Esse processo iterativo permite que as metaheurísticas explorem diferentes regiões do espaço de busca para encontrar soluções que se aproximem do resultado desejado.

Entre as diversas técnicas metaheurísticas aplicadas, destacam-se o algoritmo genético (GA) e a Programação Genética (PG), que são amplamente utilizados em problemáticas de otimização (CHATTERJEE, LAUDATO e LYNCH, 1996; JIN et al., 2016). Inspirado no processo genético observado em organismos biológicos, estas metaheurísticas empregam princípios de seleção natural e sobrevivência para explorar o espaço de busca do problema e gerar várias soluções a cada iteração. Essa abordagem permite que o algoritmo busque diferentes soluções e evolua ao longo das gerações, aumentando as chances de encontrar uma solução próxima ao ótimo global para o problema em questão (LI, LUCASIUS e KATEMAN, 1992).

O algoritmo genético (AG) foi concebido na década de 90 por John Holland e, desde então, tem sido objeto de estudo e aplicação em diversas áreas da engenharia (Garg, 2016). A principal característica que o torna uma ferramenta poderosa de otimização é sua habilidade de buscar soluções ótimas globais, especialmente em situações com múltiplos modos e objetivos. Essa capacidade de explorar diferentes soluções e encontrar resultados superiores

faz do AG uma abordagem valiosa para resolver problemas complexos e desafiadores em diversas áreas de aplicação.

A codificação é um item fundamental nas estruturas do AG e da PG, pois fazem com que estes sejam adequados em suas diversas finalidades. Os cromossomos são responsáveis por serem os conjuntos de soluções das problemáticas abordadas e são codificados de diversas formas, como cadeias binárias, valores reais ou outras representações (Linden, 2012). Já a população inicial é construída por agrupamento de cromossomos aleatórios, sendo que o tamanho desta impacta diretamente no desempenho da metaheurística (Michalewicz, 1996).

A *fitness function* (ou função de avaliação) tem um papel de seleção dos cromossomos mais adequados para resolver a problemática proposta, por meio da atribuição de resultados para cada um dos cromossomos gerados, dessa forma, mensura a qualidade da solução obtida (GHAMISI e BENEDIKTSSON, 2015) e seleciona os indivíduos considerados como mais adequados, visando que os mesmos sejam os pais da próxima geração, de acordo com a premissa da seleção natural. O operador de *crossover* (ou cruzamento) desempenha o papel de gerar novos indivíduos ao combinar os cromossomos dos pais. Essa etapa é descrita por Chuang, Chen e Hwang (2015) como uma das mais importantes, visto que ela pode evitar ótimos locais e aumentar a diversidade da população. Ghamisi e Benediktsson (2015) descrevem que outra etapa fundamental para o bom funcionamento da metaheurística abordada é a mutação, que consiste na modificação, de forma aleatória, em indivíduos e que visa garantir a diversidade genética das populações nas diferentes gerações ao longo do tempo.

Com o passar das gerações, é possível encontrar uma boa solução ao se combinar diferentes possibilidades, buscando constantemente melhorar os resultados obtidos (Balieiro et al., 2014). Para finalizar o algoritmo de forma adequada, é essencial estabelecer um critério de parada, conforme sinalizado por Gomide (2009). Diversas abordagens podem ser utilizadas, tais como definir um número máximo de iterações, um tempo máximo de processamento ou aguardar a estabilização da função fitness (RODRIGUES et al., 2004).

Pode-se concluir que a PG compartilha princípios semelhantes ao AG, em estrutura e fundamentação, porém se diferencia pela abordagem de modelagem. Enquanto o AG requer um modelo pré-definido, a PG, também conhecida pela alcunha de regressão simbólica, é capaz de gerar a estrutura do modelo, selecionando variáveis e coeficientes adequados com base no problema em questão, conforme descrito por Koza (1990). Em resumo, algoritmo genético/programação genética podem ser destacadas como estratégias poderosas de busca global, que simulam a evolução biológica para encontrar soluções ótimas para problemas de

otimização lineares e não lineares. Sua utilização permite a exploração de diferentes soluções e a busca por resultados melhores, tornando-o uma ferramenta valiosa em diversas áreas da engenharia.

A técnica de *Random Forest* (RF) pode ser descrita como um algoritmo de aprendizado de máquina baseado em árvores de decisão que visa sanar problemáticas de classificação e regressões, desenvolvido por Breiman (2001). A proposição tem uma abordagem flexível, baseada na geração de modelos de interação com uma vasta quantidade de dados (árvores de regressão) e também na obtenção de valores médios de estimativas dos modelos (WAGER e ATHEY, 2018). Para ajustar o modelo, o algoritmo depende de três parâmetros importantes: *n*tree (número de árvores na floresta), *nodesize* (tamanho do nó terminal) e *m*try (número de recursos aleatórios usados para dividir um nó da árvore) (O'BRIEN; ISHWARAN, 2019). Na etapa de treinamento do modelo de RF, são construídas uma infinidade de árvores de decisão, onde cada uma é uma amostra aleatória da base de dados com reposição (mais conhecido como “ensacamento” ou *bootstrap*) (BELGIU e DRĂGU, 2016). Além disso, durante a construção de cada árvore, apenas um subconjunto aleatório de características (atributos) é considerado para fazer as divisões nos nós da árvore. Esse processo de randomização torna o RF mais robusto e ajuda a evitar o *overfitting* do modelo aos dados de treinamento.

Após construir todas as árvores, o *Random Forest* realiza uma combinação das previsões de cada árvore para chegar a uma previsão final. Para problemas de classificação, é feita uma votação entre as árvores para determinar a classe mais frequente, e para problemas de regressão, é realizada uma média das previsões das árvores. Embora o RF seja uma técnica muito difundida em vários ramos da tecnologia, é importante mencionar que o algoritmo, por vezes, pode apresentar pouca precisão em conjuntos de dados complexos, que envolvem interações variáveis complexas ou muitas características distintas (KUMAR e SHAIKH, 2017; SPEISER et al., 2019). No entanto, pode ser utilizado em conjunto com outras técnicas de inteligência computacional para aumentar sua precisão e acurácia, conforme proposto pelos autores de Miranda et al. (2022).

## 2.6 Modelos de efeito misto

Na modelagem estatística, a análise de regressão é uma das técnicas mais comumente empregadas para estimar as relações entre uma variável dependente e uma ou mais variáveis independentes. Esse tipo de metodologia é útil quando a medição da variável dependente

demanda um esforço considerável de tempo e/ou recursos (FREEDMAN, 2009). A forma mais comum de análise de regressão é a linear, que busca encontrar a linha que melhor se ajusta aos dados observados, utilizando critérios estatísticos específicos, como o método dos mínimos quadrados. Essa técnica busca minimizar a soma das diferenças quadradas entre os dados observados e os valores preditos, garantindo uma previsão mais precisa e confiável (MONTGOMERY, PECK e VINING, 2012).

Em resumo, a análise de regressão desempenha um papel essencial na biometria florestal ao relacionar uma variável dendrométrica de difícil obtenção com variáveis que podem ser facilmente obtidas. Dessa forma, é possível alcançar uma gestão mais eficiente e sustentável dos recursos florestais, além de facilitar o monitoramento e a tomada de decisões com base em informações precisas e confiáveis (CARIELO et al., 2016). Considerando que possam ocorrer correlações entre as observações da variável de interesse (densidade básica da madeira), como, por exemplo, dependência espacial (pelo fato dos indivíduos que serão mensurados pelas unidades amostrais estarem relativamente próximos), dependência biológica (clones terem estruturas genéticas/fenotípicas semelhantes) e dependência temporal (mensurações de um mesmo indivíduo em períodos de tempo distintos), a modelagem clássica não é a mais indicada, pelo fato da violação da hipótese dos erros independentes, acarretando em estimativas não confiáveis (DE OLIVEIRA et al., 2021). Visando evitar esse tipo de percalço na modelagem, tem-se a possibilidade da utilização de modelos de efeito misto, que contêm em sua estrutura a inserção de efeitos fixos e aleatórios (que permitem incorporar as dependências entre as observações), conferindo maior robustez na estimativa da variável de interesse e diluindo o erro (VONESH, 2006).

Os efeitos aleatórios incorporados aos modelos mistos estão diretamente ligados a agrupamentos específicos (como, por exemplo, grupos clonais). Desta forma, confere-se maior flexibilidade à estrutura de covariância neste tipo de modelagem, retificando possíveis tendências errôneas (PINHEIRO e BATES, 2006). Há uma grande quantidade de estudos conduzidos sobre a temática, onde bons resultados foram evidenciados. Calegário et al. (2005) concluem que houve uma redução significativa do erro padrão residual nos modelos com a inclusão de efeitos fixos e aleatórios para a estimativa de crescimento de povoamentos de *Eucalyptus sp.*, recomendando este tipo de modelagem para as diversas aplicações no âmbito florestal. Dos Santos, Garcia e Geraldi (2004) fazem uso de modelos de efeito misto para a avaliação da posição de toras em árvores *Eucalyptus Grandis* para verificação posterior da qualidade da madeira serrada, encontrando resultados satisfatórios em sua pesquisa. Mendonça, Carvalho e Calegário (2015) propõem o uso de modelos mistos generalizados para

a estimativa de alturas em florestas de *Eucalyptus sp.*, e concluem que este tipo de técnica tem um grande potencial, visto que há um grande aumento na precisão dos modelos avaliados. Guilley, Hervé e Nepveu (2004) demonstram em seu estudo que há a possibilidade da utilização da modelagem de efeito misto na estimativa da densidade básica da madeira de *Quercus petraea Liebl.*, encontrando resultados satisfatórios quando avaliam a influência de diferentes sítios, regiões e fatores silviculturais. De Oliveira et al. (2021), utilizando modelos com variáveis de efeito aleatório para a estimativa de densidade básica em plantios de *Eucalyptus sp.*, concluem que a matriz de correlação deste tipo de modelo foi a mais adequada para a heterogeneidade da densidade, indicando esse tipo de modelagem para embasar decisões relacionadas à qualidade da madeira.

Embora, seja necessário esclarecer que a indicação desse tipo de modelagem não substitui a necessidade de uma seleção criteriosa de variáveis e uma rede amostral otimizada, que são fundamentais para capturar a alta variabilidade da característica de interesse.

### 3 CONSIDERAÇÕES FINAIS

A metodologia de Programação Genética (PG) para a seleção de variáveis e geração de um modelo apresenta um grande potencial ainda pouco explorado em estudos visando a estimativa da densidade básica da madeira de forma precisa. O método permite uma varredura em um amplo espaço de busca que pode abranger uma elevada quantidade de combinações das variáveis independentes, o que não seria possível com a utilização de métodos tradicionais e/ou manuais consagrados. Com o avanço computacional e do poder de processamento na atualidade, esse tipo de metodologia desponta como uma alternativa viável para auxiliar na seleção de variáveis e geração de modelos. Além disso, proporciona insights sobre as variáveis selecionadas e como a variação destas influencia na densidade básica da madeira para diferentes clones em ambientes contrastantes.

A utilização de métodos não destrutivos, como o Pilodyn, para o auxílio da modelagem da característica de interesse para diferentes clones em diferentes locais, conciliada a técnicas de modelos de efeitos mistos, pode ser uma alternativa viável, visto que na literatura tem apresentado bons resultados. Isso tornaria possível a proposta de operacionalização no processo de inventários florestais e biometria florestal das empresas do setor, suprimindo a necessidade de geração de grandes números de amostras em curtos períodos de tempo e com recursos reduzidos, além da avaliação prévia para novos clones introduzidos na base por parte dos setores de melhoramento florestal.



Dada a deficiência de estudos utilizando uma metodologia híbrida que visa a utilização de métodos não destrutivos para o auxílio da modelagem básica da densidade básica da madeira, em conjunto com a Programação Genética visando a seleção de variáveis e geração de um modelo linear inicial e, por fim, a utilização de técnicas de modelagem de efeito misto para a melhoria na acurácia das estimativas, tornam-se necessários trabalhos que demonstrem o potencial dessas ferramentas. Além disso, atrelar essas tecnologias ao contexto industrial pode contribuir para a melhoria dos processos nos quais a indústria de celulose está inserida.

## REFERÊNCIAS

- ALTOÉ, T. F. Modelagem e espacialização do estoque de carbono em fragmentos nativos de Minas Gerais em função de fatores edafoclimáticos. **Tese (Doutorado) – Universidade Federal de Lavras**, 2017.
- ANDERSEN, C. M.; BRO, R. Variable selection in regression - a tutorial. **Journal of chemometrics**, v. 24, n. 11-12, p. 728-737, 2010.
- ASILTÜRK, I.; ÇUNKAŞ, M. Modeling and prediction of surface roughness in turning operations using artificial neural network and multiple regression method. **Expert Systems with Applications**, v. 38, n. 5, p. 5826–5832, 2011.
- BALIEIRO, A. et al. A multi-objective genetic optimization for spectrum sensing in cognitive radio. **Expert Systems with Applications**, Oxford, v. 41, n. 8, p. 3640–3650, 2014.
- BELGIU, M.; DRĂGU, L. Random forest in remote sensing: A review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 114, p. 24–31, 2016.
- BINOTI, M. L. M. da S.; BINOTI, D. H. B.; LEITE, H. G. Aplicação de redes neurais artificiais para estimação da altura de povoamentos equiâneos de eucalipto. **Revista Árvore**, Viçosa, v. 37, n. 4, p. 639-645, 2013.
- BREIMAN, L. Random Forests. **Machine Learning**, v. 45, p. 5–32, 2001.
- BRONISZ, K.; MEHTÄTALO, L. Mixed-effects generalized height–diameter model for young silver birch stands on post-agricultural lands. **Forest Ecology and Management**, Amsterdam, v. 460, p. 117901, 2020.
- CALEGARIO, N. et al. Estimativa do crescimento de povoamentos de *Eucalyptus* baseada na teoria dos modelos não lineares em multinível de efeito misto. **Ciência Florestal**, Santa Maria, v. 15, n. 3, p. 285-292, 2005.
- CAMPOS, J. C. C.; LEITE, H. G., Mensuração florestal: Perguntas e respostas. **Livro - Universidade Federal de Viçosa**, Viçosa, 605p. 2013.
- CARNEIRO, A. C. O. et al. Potencial energético da madeira de *Eucalyptus sp.* em função da idade e de diferentes materiais genéticos. **Revista Árvore**, Viçosa, v. 38, p. 375-381, 2014.
- CARIELO, P. et al. Estimativa do DAP em função do diâmetro do toco para plantios de *Eucalyptus urograndis* implantados em diferentes espaçamentos. **Colloquium Agrariae**. p. 100-105. 2016.
- CHATTERJEE, S.; LAUDATO, M.; LYNCH, L. A. Genetic algorithms and their statistical applications: an introduction. **Computational Statistics & Data Analysis**, v. 22, n. 6, p. 633-651, 1996.
- CHAVE, J. Medição da densidade da madeira em árvores tropicais, manual de campo. **Manual de Campo, Projeto PAN AMAZONIA**, Toulouse, p. 7, 2002.

- CHAVE, J. et al. Regional and phylogenetic variation of wood density across 2456 neotropical tree species. **Ecological applications**, Washington, v. 16, n. 6, p. 2356-2367, 2006.
- CHUANG, Y. C.; CHEN, C. T.; HWANG, C. A real-coded genetic algorithm with a direction-based crossover operator. **Information Sciences**, Londres, v. 305, n. 1, p. 320–348, 2015.
- CHUMNEY, E. C. G.; SIMPSON, K. N. **Methods and designs for outcomes research**. ASHP, 2006.
- CORDEIRO, M. A. et al. Estimativa do volume de *Acacia mangium* utilizando técnicas de redes neurais artificiais e máquinas vetor de suporte. **Pesquisa Florestal Brasileira**, Colombo, v. 35, n. 83, p. 255-261, set. 2015.
- COSTA FILHO, P. A.; POPPI, R. J. Aplicação de algoritmos genéticos na seleção de variáveis em espectroscopia no infravermelho médio: determinação simultânea de glicose, maltose e frutose. **Química Nova**, São Paulo, v. 25, p. 46-52, 2002.
- COELHO, J. C. F. et al. Wood grain angles variations in *Eucalyptus* and their relationships to physical-mechanical properties. **Holzforschung**, v. 74, n. 12, p. 1089-1097, 2020.
- CUNHA, G. T. et al. Panorama do setor florestal brasileiro com ênfase no estado de Minas Gerais. **Enciclopedia Biosfera - Centro Científico Conhecer**, Goiânia, v. 16, n. 29, p. 1582-1602, 2019.
- DE OLIVEIRA, R. A. et al. Aplicação de modelos lineares de efeitos mistos para avaliar a densidade básica da madeira de duas espécies e um híbrido de *Eucalyptus*. **Scientia Florestalis**, Piracicaba, v.49, n.129, 2021.
- DIAS, D. D. C. et al. Uso da técnica de resistografia e de variáveis dendrométricas na modelagem da densidade básica de povoamentos clonais de *Eucalyptus*. **Ciência Florestal**, Santa Maria, 27, 609-619. 2017.
- DOS SANTOS, P. E. T.; GARCIA, J. N.; GERALDI, I. O. Posição da tora na árvore e sua relação com a qualidade da madeira serrada de *Eucalyptus grandis*. **Scientia Florestalis**, Piracicaba, n. 66, p. 142 - 151, 2004.
- FANTUZZI NETO, H. Qualidade da madeira de eucalipto para produção de celulose Kraft. 2012. 105f. **Tese (Doutorado em Ciência Florestal) – Universidade Federal de Viçosa**, Viçosa, MG, 2012.
- FARJAT, A. E. et al. Modeling climate change effects on the height growth of *loblolly pine*. **Forest Science**, v. 61, n. 4, p. 703-715, 2015.
- FAO - Food and Agriculture Organization of the United Nations. **Global Forest Resources Assessment - 2020 - Key Findings**. Disponível em: <<https://www.fao.org/3/ca9825en/ca9825en.pdf>>. (Acesso em 18 de março de 2023).
- FERRAZ FILHO, A. C.; SCOLFORO, J. R. S.; MOLA-YUDEGO, B. The coppice-with-standards silvicultural system as applied to *Eucalyptus* plantations - a review. **Journal of Forestry Research**, v. 25, n. 2, p. 237-248, 2014.

FIorentin, L. D. et al. Generalized linear models for tree survival in *Loblolly Pine* plantations. **Cerne**, Lavras, v. 25, p. 347-356, 2020.

FOELKEL, C. E. B.; BRASIL, M. A. M.; BARRICHELO, L. E. G. Métodos para determinação da densidade básica de cavacos para coníferas e folhosas. **IPEF**, Piracicaba, v. 2, n. 3, p. 65-74, 1971.

FOELKEL, C. E. B.; MORA, E.; MENOCELLI, S. Densidade básica: sua verdadeira utilidade como índice de qualidade na madeira de eucalipto para produção de celulose. **O papel**, Campos do Jordão, v. 53, n. 5, p. 35 - 40, 1992.

FOELKEL, C. E. B. ***Eucalyptus* online book** – Capítulo 17: Diferenciando polpas de mercado e papéis de eucalipto através da gestão dos finos celulósicos da polpa. 2009. Disponível em: <[https://www.eucalyptus.com.br/eucaliptos/PT17\\_Finoscelulosicos.pdf](https://www.eucalyptus.com.br/eucaliptos/PT17_Finoscelulosicos.pdf)>. (Acesso em: 20 de outubro de 2021).

FOELKEL, C. E. B. ***Eucalyptus* online book** – Capítulo 42: Qualidade da madeira do eucalipto – acerca dos acertos e erros na utilização da densidade básica como indicador de qualidade de madeiras. 2015. Disponível em: <[https://www.eucalyptus.com.br/eucaliptos/PT42\\_Densidade\\_Basica\\_Acertos&Erros.pdf](https://www.eucalyptus.com.br/eucaliptos/PT42_Densidade_Basica_Acertos&Erros.pdf)>. (Acesso em: 20 out. 2021).

FREEDMAN, D. A. **Statistical models: theory and practice**. Cambridge University Press, 2009.

FUNDOVA, I.; FUNDA, T.; WU, H.X. Non-destructive wood density assessment of Scots pine (*Pinus sylvestris* L.) using Resistograph and Pilodyn. **PLoS ONE**, São Francisco, 13, v.13, n. 9 p. e0204518, 2018.

GHAMISI, P.; BENEDIKTSSON, J. A. Feature selection based on hybridization of genetic algorithm and particle swarm optimization. **IEEE Geoscience and Remote Sensing Letters**, Nova York, v. 12, n. 2, p. 309–313, 2015.

GAO, S. et al. A critical analysis of methods for rapid and nondestructive determination of wood density in standing trees. **Annals of Forest Science**, v. 74, n. 2, p. 27, 2017.

GARG, H. A hybrid GSA-GA algorithm for constrained optimization problems. **Information Sciences**, New York, v. 478, p. 292–305, 2016.

GOMIDE, J. L.; FANTUZZI NETO, H.; REGAZZI, A. J. Análise de critérios de qualidade da madeira de eucalipto para produção de celulose kraft. **Revista Árvore**, Viçosa, v. 34, p. 339-344, 2010.

GOMIDE, L. R. **Planejamento florestal espacial**. 2009. 256p. Tese (Doutorado em Engenharia Florestal) - Universidade Federal do Paraná, 2009.

GOUVÊA, A. F. G. et al. Determinação da densidade básica da madeira de *Eucalyptus* por diferentes métodos não destrutivos. **Revista Árvore**, Viçosa, v. 3, n. 2, p. 349-358, fev. 2011b.

GREAVES, B. L. et al. Use of a Pilodyn for the indirect selection of basic density in *Eucalyptus nitens*. **Canadian Journal of Forest Research**, v. 26, n. 9, p. 1643-1650, 1996.

- GUILLEY, E.; HERVÉ, J.; NEPVEU, G. The influence of site quality, silviculture and region on wood density mixed model in *Quercus petraea* Liebl. **Forest Ecology and Management**, Amsterdam, v. 189, n. 1-3, p. 111-121, 2004.
- HANSEN, C. P. Application of the Pilodyn in Forest Tree Improvement. **Nota Técnica, Danida Forest Seed Centre**, Humlebaek - Dinamarca, v. 2, 2000.
- HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001. 900p
- HEIN, P. R. G. et al. Near infrared spectroscopy for estimating wood basic density in *Eucalyptus urophylla* and *Eucalyptus grandis*. **Cerne**, Lavras, v.15, n.2, p.133-141, 2009.
- HEIN, P. R. G. Multivariate regression methods for estimating basic density in *Eucalyptus* wood from near infrared spectroscopic data. **Cerne**, Lavras, v. 16, p. 90-96, 2010.
- HEIN, P. R. G.; LIMA, J. T.; CHAIX, G. Otimização de calibrações baseadas em espectroscopia no infravermelho próximo para estimativa de propriedades da madeira de *Eucalyptus*. **Floresta**, Curitiba, v. 40, n. 3, 2010.
- HOLLAND, J. Genetic Algorithms. **Scientific american**, v. 267, n. 1, p. 44-50, 1992.
- HUSCH, B.; BEERS, T. W.; KERSHAW, J. A. **Forest Mensuration 4th edition** Wiley. 2002.
- IBÁ - Indústria Brasileira de Árvores. **Relatório Anual Ibá 2017**, ano base 2016. Disponível em: <https://iba.org/datafiles/publicacoes/pdf/iba-relatorioanual2017.pdf>. (Acesso em 20 de outubro de 2021).
- IBÁ - Indústria Brasileira de Árvores. **Relatório Anual Ibá 2022**, ano base 2021 Disponível em: <https://www.iba.org/datafiles/publicacoes/relatorios/relatorio-anual-iba2022-compactado.pdf> . (Acesso em 16 de março de 2023).
- JARDIM, J. M. et al. Avaliação da qualidade e desempenho de clones de eucalipto na produção de celulose. **O papel**, v. 78, n. 11, p. 122-129, 2017.
- JATI, S. R.; FEARNSIDE, P. M.; BARBOSA, R. I. Densidade da madeira de árvores em savanas do norte da Amazônia brasileira. **Acta Amazonica**, Manaus, v. 44, p. 79-86, 2014.
- JIN, Y. et al. Selection of sand models and identification of parameters using an enhanced genetic algorithm. **International Journal for Numerical and Analytical Methods in Geomechanics**, v. 40, n. 8, p. 1219-1240, 2016.
- KIEN, N. D. et al. Genetic variation in wood basic density and pilodyn penetration and their relationships with growth, stem straightness, and branch size for *Eucalyptus urophylla* in Northern Vietnam. **New Zealand Journal of Forestry Science**, v. 38, n. 1, p. 160-175, 2008.
- KONAR, A. **Computational Intelligence: Principles, Techniques and Applications**. New York: Springer-Verlag Berlin Heidelberg, 2005.
- KOZA, J. R. **Genetic programming: A paradigm for genetically breeding populations of computer programs to solve problems**. Stanford University, Department of Computer Science, Stanford, 1990.

- KUMAR, S. S.; SHAIKH, T. Empirical Evaluation of the Performance of Feature Selection Approaches on Random Forest. **International Conference on Computer and Applications (ICCA)**. Doha, p. 227-231. 2017.
- KWAK, Doo-Ahn et al. Detection of individual trees and estimation of tree height using LiDAR data. **Journal of Forest Research**, v. 12, n. 6, p. 425-434, 2007.
- LAZZAROTTO, M. et al. Espectroscopia no infravermelho próximo para estimativa da densidade básica de madeiras de *Pinus*. **Revista Ciência da Madeira (Brazilian Journal of Wood Science)**, Pelotas, v.7, n.3, 2016.
- LI, T.; LUCASIU, C. B.; KATEMAN, G. Optimization of calibration data with the dynamic genetic algorithm. **Analytica Chimica Acta**, v. 268, n. 1, p. 123-134, 1992.
- LIDBERG, W.; NILSSON, M.; ÅGREN, A. Using machine learning to generate high-resolution wet area maps for planning forest management: A study in a boreal forest landscape. **Ambio**, Stockholm, p. 1-12, 2019.
- LIMA, J. T. et al. Adequação do Resistograph para a estimativa da densidade básica da madeira de *Eucalyptus*. **ENCONTRO BRASILEIRO EM MADEIRAS E EM ESTRUTURAS DE MADEIRA–EBRAMEM**, v. 10, 2006.
- LINDEN, R. **Algoritmos Genéticos**. 3a Edição, Rio de Janeiro: Ciência Moderna, 2012.
- LOPES, G. A.; GARCIA, J. N. Densidade básica e umidade natural da madeira de *Eucalyptus saligna Smith*, de Itatinga, associadas aos padrões de casca apresentados pela população, **Scientia Forestalis**, Piracicaba, n. 62, p. 13-23, 2002.
- MACHADO, C. M. et al. **Eucaliptocultura no Brasil, silvicultura, manejo e ambiência**. Viçosa: SIF, p. 284-302, 2014.
- MAGALHÃES, W. L. E. et al. Uso da espectroscopia no infravermelho próximo (NIR) para predição não-destrutiva de densidade básica da madeira de *Pinus taeda*. **Embrapa Florestas- Comunicado Técnico**, 2006.
- MAGNAGO; L. M.; ARANTES, M. C. A.; VIDAURRE, G. B.; MOULIN, J. C.; TRUGILHO, P. F. Energy estimate and carbon stock in short-rotation Eucalyptus stands. **Revista Cerne**, Lavras v. 22, n. 4, p. 527-534, 2016
- MENDONÇA, A. R.; CARVALHO, S. P. C.; CALEGARIO, N. Modelos hipsométricos generalizados mistos na predição da altura de *Eucalyptus sp.* **Cerne**, Lavras, v. 21, p. 107-115, 2015.
- MICHALEWICZ, Z. **Genetic Algorithms + Data Structures = Evolution Programs**. 3. ed. New York: Springer-Verlag Berlin Heidelberg, 1996.
- MIRANDA, E. N. et al. Variable selection for estimating individual tree height using genetic algorithm and random forest. **Forest Ecology and Management**, Amsterdam, v. 504, p. 119828, 2022.
- MOHAN, M. et al. Optimizing individual tree detection accuracy and measuring forest uniformity in coconut (*Cocos nucifera* L.) plantations using airborne laser scanning. **Ecological Modelling**, v. 409, p. 108736, 2019.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. Introduction to linear regression analysis. **Livro**, Nova Jersey: Wiley, ed. 5, 2012.

MOREIRA, J. M. M. Á. P., SIMIONI F. J., OLIVEIRA E. D. Importância e desempenho das florestas plantadas no contexto do agronegócio brasileiro. **Floresta**, Curitiba v. 47, n. 1, p. 85-94, 2017.

MURPHY, K. P. Machine learning: a probabilistic perspective. **MIT press**, 2012.

NEVES, T. A. et al. Avaliação de clones de *Eucalyptus* em diferentes locais visando à produção de carvão vegetal. **Pesquisa Florestal Brasileira**, Colombo, v. 31, n. 68, p. 319-330, 2011.

NGAI, E. W. T.; XIU, L.; CHAU, D. C. K. Application of data mining techniques in customer relationship management: A literature review and classification. **Expert Systems with Applications**, Oxford, v. 36, n. 2 PART 2, p. 2592–2602, 2009.

O'BRIEN, R.; ISHWARAN, H. A random forests quantile classifier for class imbalanced data. **Pattern Recognition**, Oxford, v. 90, p. 232–249, 2019.

OLIVEIRA, J. T. S.; SILVA, J. C. Variação radial da retratibilidade e densidade básica da madeira de *Eucalyptus saligna* Sm. **Revista Árvore**, Viçosa, v. 27, p. 381-385, 2003.

OLIVEIRA, J. T. S.; HELLMEISTER, J. C.; TOMAZELLO FILHO, M. Variação do teor de umidade e da densidade básica na madeira de sete espécies de eucalipto. **Revista Árvore**, Viçosa v. 29, p. 115-127, 2005.

ORELLANA, E.; FIGUEIREDO, A. Uso do método da predição de parâmetros para projetar a distribuição diamétrica em florestas nativas com a função Weibull. **Ciência Florestal**, Santa Maria, v. 27, p. 981-991, 2017.

PAYANDEH, B. Some applications of nonlinear regression models in forestry research. **The Forestry Chronicle**, v. 59, n. 5, p. 244-248, 1983.

PINHEIRO, J. C., BATES, D. M. Mixed-Effects Models in Sand S-PLUS. **Springer Science & Business Media**, Nova Yorque, 2006.

PRESSLER, M. R. Der forstliche Zuwachsbohrer neuester Construction und dessen praktische Bedeutung und Anwendung für die forstliche Forschungs-, Taxations, Pflege-und Nutzungs-Technik? **Woldemar Türk**, 1866.

QUEIROZ, S. C. S. et al. Influência da densidade básica da madeira na qualidade da polpa kraft de clones híbridos de *Eucalyptus grandis* W. Hill ex Maiden x *Eucalyptus urophylla* ST Blake. **Revista Árvore**, Viçosa, v. 28, p. 901-909, 2004.

RAHMAN, I.; MOHAMAD-SALEH, J. Hybrid bio-Inspired computational intelligence techniques for solving power system optimization problems: A comprehensive survey. **Applied Soft Computing Journal**, Londres, v. 69, p. 72–130, 2018.

RAIMUNDO, M. R.; SCOLFORO, H. F.; MELLO, J. M.; SCOLFORO, J. R. S.; MCTAGUE, J. P.; REIS, A. A. Geostatistics applied to growth estimates in continuous forest inventories. **Forest Science**, v. 63, n. 1, p. 29–38, 2017

- RIBASKI, N. G. Conhecendo o setor florestal e perspectivas para o futuro. **Brazilian Journal of Animal and Environmental Research**, Curitiba v. 1, n. 1, p. 44-58, 2018.
- RODRIGUES, F. L. et al. Metaheurística algoritmo genético para solução de problemas de planejamento florestal com restrições de integridade. **Revista Árvore**, Viçosa, v. 28, n. 2, p. 233–245, 2004.
- SCHIMLECK, L. R. et al. Comparison of methods for estimating mechanical properties of wood by NIR spectroscopy. **Journal of Spectroscopy**, v. 2018, 2018.
- SCHIMLECK, L. R. et al. Non-destructive evaluation techniques and what they tell us about wood property variation. **Forests**, v. 10, n. 9, p. 728, 2019.
- SCOLFORO, H. F. et al. Incorporating rainfall data to better plan *Eucalyptus* clones deployment in eastern Brazil. **Forest Ecology and Management**, Amsterdam, v. 391, p. 145-153, 2017.
- SCOLFORO, H. F. et al. *Eucalyptus* growth and yield system: Linking individual-tree and stand-level growth models in clonal Eucalypt plantations in Brazil. **Forest Ecology and Management**, Amsterdam, v. 432, p. 1–16, 2019.
- SILVERIO, F. O. et al. Characterization of lipophilic wood extractives from clones of *Eucalyptus urograndis* cultivate in Brazil. **BioResources**, v. 2, n. 2, p. 157-168, 2007.
- SPEISER, J. L. et al. A comparison of random forest variable selection methods for classification prediction modeling. **Expert Systems with Applications**, Oxford, v. 134, p. 93–101, 2019.
- 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**, Amsterdam, v. 259, n. 9, p. 1684-1694, 2010.
- TIBSHIRANI, R. Regression shrinkage and selection via the lasso. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 58, n. 1, p. 267-288, 1996.
- THIERSCH, C. R. et al. O uso de modelos matemáticos na estimativa da densidade básica da madeira em plantios de clones de *Eucalyptus sp.* **Cerne**, Lavras v. 12, n. 3, p. 264-278, 2006.
- VONESH, E. F. Mixed models: Theory and applications. **Journal of the American Statistical Association**, 2006.
- WAGER, S.; ATHEY, S. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. **Journal of the American Statistical Association**, New York, v. 113, n. 523, p. 1228–1242, 2018.
- WATZLAWICK, L. F.; KIRCHNER, F. F.; SANQUETTA, C. R. Estimativa de biomassa e carbono em floresta com araucária utilizando imagens do satélite IKONOS II. **Ciência florestal**, Santa Maria, v. 19, n. 2, p. 169-181, 2009.
- WU, S. et al. Use of the pilodyn for assessing wood properties in standing trees of *Eucalyptus* clones. **Journal of Forestry Research**, v. 21, n. 1, p. 68-72, 2010.
- ZIMMER, J.; ANZANELLO, M. J. Um novo método para seleção de variáveis preditivas com base em índices de importância. **Production**, v. 24, p. 84-93, 2014.





**SEGUNDA PARTE - ARTIGO****TÍTULO**

**SMIB: Sistema de Modelos para Inventário de Biomassa em plantações comerciais de eucalipto no Brasil**

**TITLE**

**SMIB: Model System for Biomass Inventory in Commercial eucalypt stands in Brazil**

Artigo formatado conforme a NBR 6022 (ABNT, 2003) e adaptado as exigências do Manual de Normalização de Trabalhos Acadêmicos da UFLA.

## RESUMO

Tradicionalmente, os inventários buscam estimar de maneira precisa o volume dos povoamentos florestais. Porém, todo o planejamento da indústria de celulose é feito em biomassa, o que, por vezes, demonstra um descasamento entre informações. Considerando o contexto atual, sabe-se que há a necessidade de buscar um inventário florestal que seja tão preciso na estimativa de biomassa dos povoamentos, conforme já é nas estimativas de volume. Assim, o presente estudo teve como objetivo apresentar uma nova abordagem desenvolvida para a geração de um sistema de modelos de inventário de biomassa (SMIB) em plantios de eucalipto. Essa abordagem, definida como híbrida, utiliza métodos não destrutivos para a obtenção de densidade básica da madeira (DB), programação genética (PG) para seleção de variáveis e modelagem de efeito misto. A base de dados coletada entre 2020-2023 é proveniente de plantios de eucalipto no estado de São Paulo, onde foram inventariadas 55 parcelas para 3 clones distintos. Nestas parcelas, foram selecionadas 165 árvores para amostragem destrutiva e posterior quantificação da DB em laboratório. Em cada parcela, além das medições biométricas, coletou-se a dureza (mm) a partir do uso do Pilodyn para as mesmas árvores que foram para quantificação de DB em laboratório. A abordagem híbrida foi separada em 3 passos, onde: primeiro aplicou-se a PG para melhor detectar dentro de um espaço de busca as combinações de variáveis (expressões) que maximizassem a explicação da DB por variáveis independentes; num segundo momento ajustou-se um modelo de efeito misto (PGM1) com o uso da variável independente dureza (coletada pelo Pilodyn), visando gerar uma estimativa precisa ao ponto de servir como proxy da DB obtida em laboratório; e, por fim, gerou-se um modelo de extrapolação de DB (PGM2) a partir dos resultados do modelo do PGM1. De maneira geral, os resultados da PG indicaram variáveis que, combinadas, explicam parte significativa da variação de DB entre árvores e clones. Os resultados demonstraram que o modelo PGM1 pode ser utilizada como uma proxy da DB laboratorial. Por fim, os resultados mostraram que PGM2 gera estimativas precisas de DB e biomassa. Conclui-se que o SMIB se apresentou eficaz, viabilizando a inserção da informação DB e biomassa em escala operacional na rotina de inventário florestal.

**Palavras-chave:** Densidade básica da dadeira. Programação genética. Modelo linear de efeito misto.

## ABSTRACT

Traditionally, inventories aim to accurately estimate the volume of forest stands. However, all planning in the pulp industry is realized based on biomass. Which sometimes leads to a mismatch of information. Considering the current context, it's known that there is a need to seek a forest inventory that is as precise in estimating the biomass of stands as it is in estimating volume. Thus, the present study aimed to present a new approach developed for generating the biomass inventory in eucalyptus plantations. This approach, defined as a hybrid, uses non-destructive methods to obtain Basic Wood Density (BWD), genetic programming (PG) for variable selection, and mixed-effect modelling. The database collected between 2020-2023 comes from eucalyptus plantations in São Paulo - Brazil, where 55 plots were inventories for three different clones. For these plots, 165 trees were selected for destructive sampling and subsequent quantification of BWD in the laboratory. In each plot, in addition to biometric measurements, hardness (mm) was collected using a Pilodyn for the same trees that were used for BWD quantification in the laboratory. The hybrid approach was divided into three steps where: in Step 1, PG was applied to better detect within a search space the combinations of variables that maximized the explanation of BWD by independent variables; in Step 2, a mixed-effect model was fitted using the independent variable hardness (collected by Pilodyn), aiming to generate a precise estimate that could serve as a proxy for the BWD obtained in the laboratory; and finally, in Step 3, a BWD extrapolation model was generate based on the results of the Step 2 model. Overall, the results of Step 1 indicated variables that, when combined, explain a significant part of the BWD variation between trees and clones, while the results in Step 2 showed that the proposed approach using Pilodyn could be used as a proxy for laboratory BWD. Finally, the results of Step 3 demonstrated that the obtained estimates allow precise biomass estimations. It was concluded that the hybrid approach proved effective, thus enabling the implementation of biomass inventory on an operational scale.

**Keywords:** Basic wood density. Genetic programming. Linear mixed effects model.

## 1 INTRODUÇÃO

A partir da crescente expansão da eucaliptocultura no Brasil, os setores de base florestal, em especial o da celulose, têm se beneficiado. Tudo isso se reflete em números, onde o Brasil ocupa o posto de 2º maior produtor de celulose do mundo, com 22,5 milhões de toneladas produzidas, além de ser o maior exportador importador de celulose (US\$ 6,7 bilhões). A celulose é a principal fonte de matéria-prima resultante dos plantios comerciais de eucalipto no país, e assim, todo o planejamento das fábricas de celulose precisa ser realizado na visão da biomassa a ser recebida anualmente. Por outro lado, o planejamento florestal de curto, médio e longo prazo ainda é baseado exclusivamente em informações volumétricas (m<sup>3</sup>). Assim, há uma inconsistência de valores planejados e reais utilizados nas empresas florestais (florestal e industrial), ocasionando erros que afetam diretamente na tomada de decisões por parte dos gestores florestais e industriais (THIERSCH et al., 2006). Em momentos de crise e escassez de madeira, essa diferença pode exercer um fator negativo no suprimento da fábrica.

Ao longo das décadas, o volume vem sendo a única informação de referência para o balizamento das decisões. Essa variável é estimada através de métodos já consagrados na literatura científica, sendo baseados em estudos da forma das árvores ou equações simples de regressão para diferentes tipos de florestas (KOZAK, MUNRO e SMITH, 1969; CLUTTER, 1980; PARRESOL, HOTVEDT e CAO, 1987; FIGUEIREDO-FILHO, BORDERS e HITCH, 1996; CHICHORRO, RESENDE e LEITE, 2003; AKINDELE e LEMAY, 2006). Apesar dos significativos esforços em sua melhoria, outras variáveis deveriam apresentar o mesmo aporte científico. Há um gap de informações para a biomassa de plantios comerciais de eucalipto, no que tange modelos preditivos. E por isso, os inventários florestais de biomassa ainda são restritos no país.

A biomassa de uma árvore é a razão entre seu peso ou massa e a densidade básica (DB), sendo influenciada pela origem do material genético e seus cruzamentos (clones), taxa de crescimento e propriedades físico-químicas da madeira. Além disso, há uma infinidade de fatores extra indivíduos que podem influenciar a característica de interesse, como por exemplo, sítio, precipitação, dentre outros (ROCHA et al., 2020). Uma forma já conhecida para acelerar a obtenção de dados de DB, e que permite conhecer a sua amplitude, é a utilização de métodos não-destrutivos e a técnica de

modelagem mista. Entre os métodos não-destrutivos, destaca-se a possibilidade de uso do Resistógrafo, Pilodyn e/ou NIR, sendo que estes permitem inferir sobre a DB com precisão e eficiência (THIERSCH et al., 2006; HEIN, 2009; GOUVÊA et al., 2011b), podendo tornar-se uma proxy da DB obtida em laboratório (SCHIMLECK et al., 2018; WU et al., 2018). Dentre esses métodos, destaca-se, em especial, o uso do Pilodyn, como sendo um equipamento portátil e de fácil operacionalidade. Assim, sua aplicação no campo é viável e possibilita uma ampla amostragem em diversas localidades/clones/sítios/idades com rapidez e economia (FUNDOVA, FUNDA e WU, 2018). Por outro lado, o uso de modelos de efeitos mistos vem sendo amplamente aplicado, demonstrando eficácia nas estimativas mesmo com um número reduzido de amostras (PINHEIRO e BATES, 2006). Há uma grande quantidade de estudos desenvolvidos que utilizam essa técnica de modelagem para as mais variadas finalidades, como por exemplo, avaliação da disposição de toras ao longo de uma árvore visando o melhor aproveitamento e qualidade da madeira serrada (DOS SANTOS, GARCIA E GERALDI, 2004), estimativa da altura de povoamentos de *Eucalyptus* (MENDONÇA, CARVALHO e CALEGÁRIO, 2015) e, também, para a estimativa da densidade básica da madeira (GUILLEY, HERVÉ e NEPVEU, 2004; DE OLIVEIRA et al., 2021), e todos apresentando resultados satisfatórios e eficazes. Apesar das contribuições, a construção e identificação de variáveis ainda são um gargalo do processo, mesmo tendo métodos robustos de modelagem.

Uma das principais dificuldades na investigação científica é a seleção de variáveis, que devido a condições de escala, podem ter origem climáticas, espectrais, cadastrais e até mesmo as clássicas a nível árvore. Uma opção válida para se explorar e obter respostas é o uso da programação genética. A técnica em questão permite a realização de uma busca exaustiva pelo infinito campo das soluções, demonstrando quais as combinações entre variáveis são mais viáveis para explicar/predizer a característica de interesse (LEARDI, 1996, HUSSEIN, KHARMA, WARD, 2001, LEI, 2012). A Programação genética tem sido amplamente utilizada visto à sua capacidade de solucionar problemas complexos das mais variadas naturezas, como por exemplo, no campo da construção civil/elétrica por meio da realização de previsões relacionadas ao consumo energético para construções integradas (NAZIR et al., 2023), no campo da engenharia de tráfego por meio de previsão em tempo real risco de acidentes no trânsito em autoestradas (MA et al., 2023), no campo medicina por meio da identificação de padrões de emoções/comportamento humano para o tratamento da saúde mental de

pacientes (SAKALLE et al., 2022) entre outros. No entanto, ainda há poucos estudos que abordam a programação genética na área florestal visando a seleção de variáveis para modelagem ou até mesmo a estimação das variáveis de interesse.

Diante da necessidade de geração de estimativas precisas de biomassa nos inventários florestais, há a necessidade de estudos que criem novas abordagens para geração de estimativas precisas e ágeis para a DB. Sendo assim, este estudo propõe uma abordagem híbrida para tornar realidade o inventário de biomassa no Brasil. A abordagem proposta se baseia em (1) aplicação da Programação Genética na seleção de variáveis combinadas que melhor expliquem a variação do comportamento da DB; (2) modelagem da DB de diferentes clones através do uso do Pilodyn pela técnica de modelos de efeitos mistos reconhecendo que esse modelo terá capacidade de ser uma proxy da DB laboratorial; (3) extrapolação da DB ao nível de árvore individual em cada clone, sendo a partir de então gerada a informação de biomassa para cada árvore.

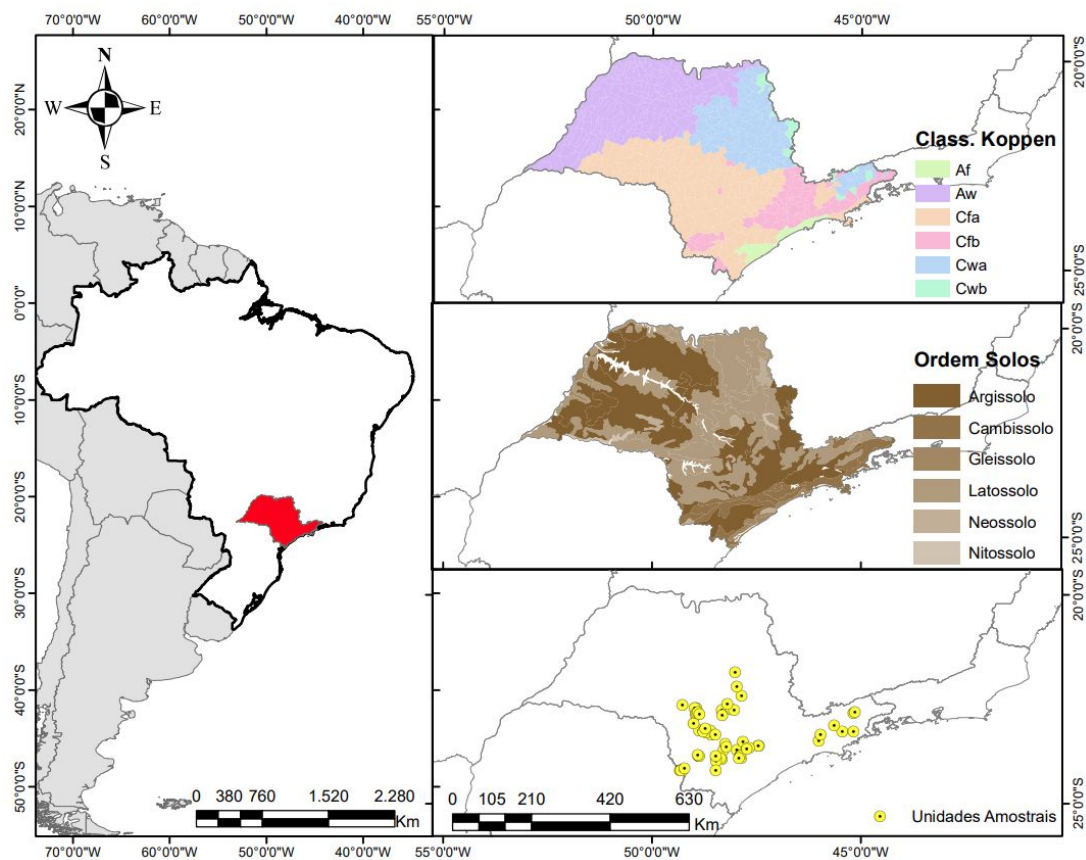
## **2 MATERIAL E MÉTODOS**

### **2.1 Caracterização da área de estudo**

A área de estudo abrange regiões do estado de São Paulo (SP) com variações climáticas que correspondem a Cfa (Clima subtropical úmido, com verão quente), Cfb (Clima subtropical oceânico temperado, com verão ameno), Cwa (Clima subtropical úmido de inverno seco) e Cwb (Clima subtropical de altitude de inverno seco e verão ameno), de acordo com o critério de *Köppen* (Alvares et al., 2013). Há ainda uma considerável amplitude de solos (45% de Latossolos, 27% de Neossolos, 18% de Argissolos e 10% de Cambissolos), conforme o sistema de classificação de solos da Embrapa (2013). Por outro lado, a altitude varia de 500 m até 1.100 m acima do nível do mar. Por fim, a área de estudo cobre extensas áreas de plantios comerciais de eucalipto, sendo referência no estudo de diversas zonas de crescimento do eucalipto em SP (Figura 1).



Figura 1 - Distribuição espacial das unidades amostrais mensuradas para o desenvolvimento dos modelos e suas interações ambientais.



Fonte: Do autor (2023).

## 2.2 Caracterização da área de estudo

Os dados são provenientes de plantios comerciais clonais de eucalipto em SP, e as coletas de campo foram realizadas entre 2020 e 2023. Os talhões amostrados foram selecionados de maneira aleatória dentro dos diferentes estratos climáticos (4 zonas), genéticos (3 clones), produtivos (3 classes de sítio) e de idade (5 classes de idade). De maneira geral, buscou-se uma amostragem que garantisse representar toda a amplitude de crescimento do eucalipto no estado de SP (Tabela 1).

Tabela 1 - Resumo do volume total com casca (Vol.,  $\text{m}^3.\text{ha}^{-1}$ ) e densidade básica da madeira (DB,  $\text{kg}.\text{m}^{-3}$ ) para os diferentes clones, zonas climáticas, classes de idade e sítio em SP.

Estrato	Total de parcelas	Volume total com casca ( $\text{m}^3.\text{ha}^{-1}$ )			Densidade básica ( $\text{kg}.\text{m}^{-3}$ )		
		Mínimo	Médio	Máximo	Mínima	Média	Máxima
<b>Clones (CL)</b>							
CL 1	21	16,3	246,7	468,8	379,7	471,5	605,8
CL 2	19	64,6	196,4	390,4	354,2	434,5	541,4
CL 3	15	34,7	230,0	334,2	350,5	436,9	513,5
<b>Zonas Climáticas (ZC)</b>							
ZC 1	26	67,5	241,0	468,8	350,5	449,0	535,7
ZC 2	12	16,3	214,1	334,2	384,8	431,8	513,5
ZC 3	5	91,5	188,3	305,6	398,4	451,7	496,0
ZC 4	12	74,4	211,8	375,3	355,9	467,7	605,8
<b>Classes de Idade (ID)</b>							
ID <= 3 Anos	13	16,3	85,8	177,9	354,2	429,0	488,5
ID ~ 4 anos	10	73,5	182,2	266,5	350,5	447,4	509,4
ID ~ 5 anos	10	185,9	248,4	327,8	358,3	451,0	541,4
ID ~ 6 anos	14	246,3	312,1	411,2	369,7	454,5	531,8
ID >= 7 Anos	8	261,8	334,1	468,8	398,4	481,8	605,8
<b>Classes de Sítio (S)</b>							
S III	19	16,3	181,7	375,3	354,2	448,4	605,8
S II	18	74,4	242,7	373,1	350,5	455,7	605,4
S I	18	64,6	258,4	468,8	355,9	445,2	530,5

Fonte: Do autor (2023).

Um total de 55 talhões foi selecionado, sendo instalada uma parcela aleatória e circular ( $400 \text{ m}^2$ ) dentro de cada talhão. Na parcela, as árvores contidas tiveram seus diâmetros a 1,30 metros acima do solo (DAP, cm) medidos, e a altura (Ht, m) foi medida para uma fração de 10 árvores das fileiras centrais. A altura das árvores dominantes (Hd, m) foi medida, conforme o conceito de Assmann (1970). Para a estimativa da altura das demais árvores, foram adotadas equações hipsométricas, bem como para o volume total com casca (Scolforo et al., 2019). Além disso, informações ao nível de povoamento também foram obtidas, como o número de fustes por hectare (Nf), a área basal (G,  $\text{m}^2.\text{ha}^{-1}$ ), o volume individual total com casca ( $\text{m}^3$ ), o volume total com casca (Vol.,  $\text{m}^3.\text{ha}^{-1}$ ), o incremento médio anual (Imatcc,  $\text{m}^3.\text{ha}^{-1}$ ), Pv50 (Hakamada et al., 2015) e o índice de sítio aos 7 anos (Is, m).

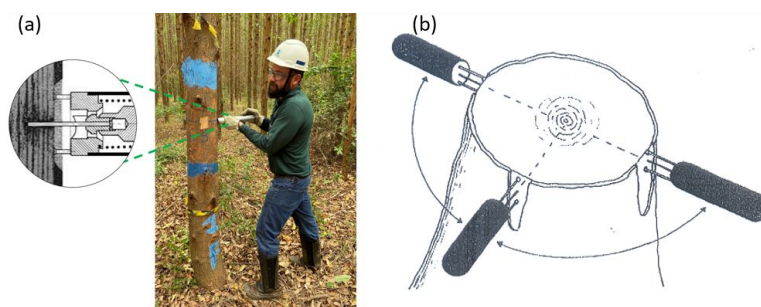
### 2.3 Amostragem da dureza e densidade básica da madeira (DB)

Em conjunto com as medições de inventário, um total de 3 árvores por parcela (165 árvores) foi selecionado para medição de sua dureza (mm) com uso do Pilodyn, bem como sua quantificação real de DB. As 3 árvores selecionadas seguiram a distribuição de DAP da parcela: uma árvore com  $DAP \leq$  percentil 25, uma árvore com DAP em torno do percentil 50 e uma árvore com  $DAP \geq$  percentil 75.

Individualmente, as árvores foram perfuradas em três pontos ortogonais na altura do DAP (Figura 2) pelo Pilodyn. Essas medidas foram convertidas na média da dureza/resistência (mm) à penetração (característica correlacionada à densidade básica da madeira).

Posteriormente, essas mesmas árvores foram abatidas e seccionadas (0%, 25%, 50%, 75% e 100%) em relação às suas alturas comerciais. Toretos sem casca (1 m) dessas posições foram retirados para posterior secagem (redução da umidade). Essas amostras foram saturadas em água para cálculo do volume por meio do método da balança hidrostática. Na sequência, as amostras foram secas em estufa de circulação de ar forçada a uma temperatura constante de 105 graus Celsius até que seu peso seco constante fosse atingido. Finalmente, calculou-se a DB de cada árvore amostrada, que pode ser definida como a divisão da massa seca em estufa e o volume da madeira acima do ponto de saturação das fibras. Adotaram-se as diretrizes da norma NBR 11941 da Associação Brasileira de Normas Técnicas (ABNT, 2003) durante todo o processo.

Figura 2 - Coleta da dureza da árvore com uso do Pilodyn nas 3 posições de coleta.



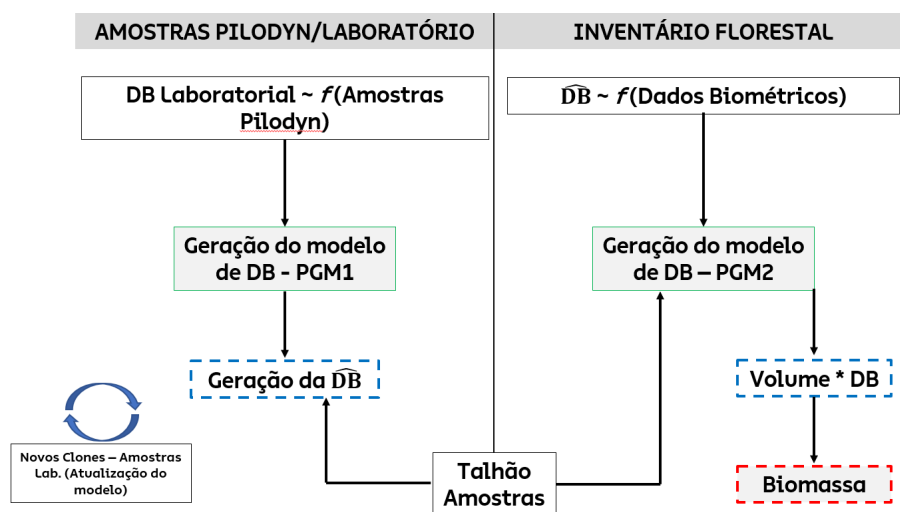
Fonte: Do autor (2023).

## 2.4 Sistema de modelos para geração do Inventário de Biomassa (SMIB)

A abordagem proposta visa à integração de modelos para reduzir o esforço amostral e temporal na determinação da DB, ao mesmo tempo permitindo sua implementação no inventário florestal de forma precisa (Figura 3). Nesse sentido, a abordagem buscou integrar os seguintes modelos:

- PGM1: modelo de efeito misto para prever a DB, cujo objetivo é demonstrar a capacidade de generalização e redução da intensidade amostral de maneira que coletas futuras com Pilodyn possam funcionar como uma proxy da DB laboratorial. Modelo PGM1 foi ajustado através da variável dureza/Pilodyn e das variáveis construídas e selecionadas pela programação genética. A programação genética considerou variáveis biométricas [Dap, Ht, idade, Nf, vitcc, Vtcc, Imatcc, Hd, Is, G, Pv50, área ocupada por cada árvore (Área Planta)].
- PGM2: modelo de efeito misto para prever a DB ao nível de árvore individual no inventário florestal, sendo que a DB utilizada como variável dependente é proveniente do PGM1 (estimativa assumida como proxy da DB laboratorial). O PGM2 foi ajustado com uso das variáveis construídas e selecionadas pela programação genética (mesmas variáveis anteriores).

Figura 3 - Fluxograma conectando o passo a passo do trabalho.



Fonte: Do Autor (2023).

O modelo PGM1 foi ajustado a partir dos resultados de 110 árvores (ano base: 2020/2021), enquanto 55 árvores (ano base: 2022/2023) foram utilizadas no ajuste do PGM2. Essa separação na base de dados visou garantir a independência entre as amostras, e, assim, os resultados preditos por PGM1 e PGM2 sempre foram comparados com a DB observada em laboratório.

Para a conclusão do SMIB e a obtenção da biomassa, utilizou-se um modelo volumétrico (Scolforo et al., 2019), que, multiplicado pela estimativa da DB, permite estimar a biomassa de cada árvore. Logo, a validação realizada permite avaliar se todo o sistema SMIB para a geração do Inventário de Biomassa é robusto o suficiente para a rotina corporativa de inventário e biometria florestal no Brasil. Vale ressaltar que novas coletas de DB passariam a ser realizadas para modelagem de novos clones, assim como para a validação contínua da qualidade do sistema SMIB.

Por fim, todo o processamento foi realizado no software R (R Core Team, 2023), sendo utilizado o pacote gramEvol (Noorian et al., 2016) para a programação genética, nlme (Pinheiro et al., 2019) no ajuste dos modelos de efeito misto e ggplot2 (Wickham, 2016) para as análises gráficas.

#### **2.4.1 Seleção e construção de novas variáveis**

A programação genética (PG) é uma técnica de aprendizado de máquina e de otimização inspirada na evolução biológica. É uma variação do Algoritmo Genético (AG), desenvolvida por Koza (1990). A PG é usual em problemas de otimização, e permite encontrar soluções em espaços de busca complexos. A estrutura dos programas é vinculada a árvores, sendo cada ramo de decisão formado por uma expressão matemática (funções). No presente estudo adotou-se o inverso ( $\frac{1}{x}$ ), logaritmo ( $\log x$ ) e exponencial ( $a^x$ ), já os operadores aritméticos que compõem as arestas e conectam estas funções foram a multiplicação, divisão, soma e subtração.

De maneira geral, o funcionamento da PG pode ser descrito da seguinte maneira:

- 1) Codificação do cromossomo (indivíduo), o qual representa uma solução potencial, é caracterizada por meio de árvores ou grados.
- 2) População Inicial, na qual é gerada, de forma aleatória, uma população de indivíduos de soluções candidatas (utilizou-se uma população de 100 indivíduos).

3) Função de avaliação (*Fitness*), na qual é determinada uma função para avaliar a solução candidata quanto à sua adequação ao problema. Para o problema em questão, buscou-se minimizar o erro quadrático médio (RMSE) derivado de um modelo de regressão linear múltiplo. Para mitigar a possível multicolinearidade, a PG foi ajustada até que o modelo encontrado apresentasse um valor de VIF (*Variation Inflation Factor*) menor que 10, além de sinais nos coeficientes que respeitassem as correlações entre DB e variáveis independentes.

4) Seleção, na qual, com base nos valores de avaliação, são selecionados indivíduos dentro da população para reprodução. Soluções com maior aptidão apresentam maior probabilidade de serem selecionadas, mas também há a chance de soluções menos aptas serem escolhidas, permitindo diversidade na população (uma característica importante da evolução biológica).

5) Cruzamento (*crossover*), que é o principal operador da PG. Consiste na combinação de pais selecionados da população para formar novas soluções (filhos). Ou seja, a partir do cruzamento de indivíduos de características diferentes, permite a criação de novas soluções potencialmente melhores. O operador de *crossover* determinado foi um ponto de corte com ajuste de amplitude;

6) Mutação, na qual outro operador permite diversidade na população. Seu uso evita a convergência prematura para uma solução subótima (solução ótima local). Este operador tem como fundamento fazer pequenas alterações aleatórias em um operador ou terminal. Para a implementação da PG, optou-se pelo uso da mutação uniforme com ajuste de amplitude, sendo a probabilidade de mutação expressa pela equação [1]. À medida que a PG evolui (processo de iteração até atingir um critério de parada – no método de aprendizado em questão utilizou-se 200), por meio dos operadores, será encontrada uma solução.

$$\text{Chance de mutação} = \frac{1}{(\text{Comprimento do Genoma} + 1)} \quad [1]$$

Apesar de a PG ser um método estocástico, a sua estratégia de iteração permite percorrer com maior eficiência o espaço de busca das variáveis e suas combinações. A combinação de expressões geradas pelo algoritmo permite entregar um modelo de acordo com as regras inseridas no método. Ao final da convergência, tem-se uma combinação de bons ótimos locais de expressões do tipo linear múltiplo [2]:

$$DB_i = \beta_0 + \beta_1 X_1 + \dots + \beta_j X_n + e_i \quad [2]$$

Em que: DB<sub>i</sub> é a densidade básica da madeira (kg.m<sup>-3</sup>) da árvore i; β<sub>j</sub> é o coeficiente a ser estimado; X<sub>n</sub> é a expressão matemática com a seleção e combinação de variáveis geradas pela PGn.

Foram disponibilizados para PG as seguintes variáveis: DAP, Ht, idade, Nf, vitcc, Vol., Imatcc, Hd, Is, G, Pv50, área ocupada por cada planta (Área Planta) e dureza (Pilodyn), sendo que esta última foi fixada com obrigatória para construção e seleção das expressões a serem utilizadas no modelo PGM1.

#### 2.4.2 Modelos de efeito misto

Ao final do processo, a PG expressa uma combinação ótima em um modelo geral linear múltiplo; no entanto, este ainda pode ser de difícil interpretação e sem a devida possibilidade de replicação para diferentes clones.

A combinação de variáveis (considerando que a variável dureza (Pilodyn) foi fixada como obrigatória para a construção das expressões) obtida no processo anterior (PG) foi utilizada em um modelo linear de efeito misto (PGM1) [3]:

$$DB_i = f(X_i, \beta_i, \mu_i) + \varepsilon_i \quad [3]$$

Em que: DB: densidade básica da madeira (kg.m<sup>-3</sup>) da árvore i; X<sub>i</sub>: variáveis independentes selecionadas pela PG; β<sub>i</sub>: coeficiente fixos do modelo; μ<sub>i</sub>: vetor de parâmetros aleatórios associado ao i-ésimo grupo clonal; ε<sub>i</sub>: erro aleatório da DB i.

Cabe ressaltar que, para a construção do PGM1, foram utilizadas 110 amostras (árvores) coletadas entre 2020/2021. A técnica de modelagem de efeito misto foi empregada para distinguir os diferentes clones; ao mesmo tempo, essa técnica também possibilitou a utilização total da base, maximizando a precisão na estimativa de cada clone.

Após o processo mencionado acima, os dados coletados de 55 amostras (árvores) entre 2022/2023 foram utilizados para o ajuste do PGM2. O processo diferiu para o PGM2 nos seguintes pontos: (1) a variável dureza (Pilodyn) não foi utilizada na construção e seleção das expressões pela PG; e (2) aplicou-se o modelo ajustado PGM1

(Eq. 3) para cada uma das 55 amostras, assumindo que essas estimativas funcionam como uma proxy da DB laboratorial ( $\widehat{DB}_i$ ) por se aproveitar da coleta da dureza das árvores.

Para o ajuste do modelo PGM2 [4]:

$$\widehat{DB}_i = f(X_i, \beta_i, \mu_i) + \varepsilon_i \quad [4]$$

Em que: variáveis foram definidas previamente.

## 2.5 Critérios de Avaliação de SMIB

Os modelos de efeito misto gerados foram avaliados segundo as estatísticas de avaliação R<sup>2</sup> [5], RMSE [6] e Erro médio [7]. Além disso, avaliou-se os gráficos de resíduos com o intuito de evidenciar possíveis tendências nas estimativas dos modelos.

$$R^2 = 1 - \frac{\sum_{i=1}^n (vo - ve)^2}{\sum_{i=1}^n (vo - vmo)^2} \quad [5]$$

$$RMSE(kg/m^3) = \sqrt{\frac{\sum_{i=1}^n (vo - ve)^2}{n}} \quad [6]$$

$$Erro\ médio = \frac{\sum_{i=1}^n (vo - ve)}{n} \quad [7]$$

Em que: vo é o valor observado da variável de interesse; ve é o valor estimado da variável de interesse; vmo é o valor médio da variável de interesse.

Por fim, aplicou-se uma reamostragem por Bootstrap com reposição para avaliar a adequação de cada modelo a novas bases de dados simuladas, permitindo avaliar diferentes cenários e condições. Para o estudo em questão optou-se pela geração de 5.000 repetições e após a geração das estatísticas, foram gerados histogramas de resíduos para verificação da distribuição dos erros.

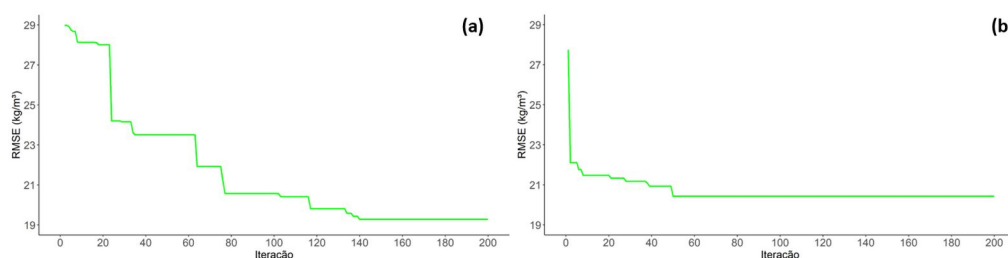


### 3 RESULTADOS

#### 3.1 Seleção e construção das expressões através da PG

À medida que a PG evolui, as expressões que melhor explicam a DB tendem a permanecer no modelo, o que é evidenciado pela diminuição gradual no valor de RMSE ( $\text{kg}\cdot\text{m}^{-3}$ ), gerando, por fim, um modelo com menor erro médio de estimativa (Figura 4).

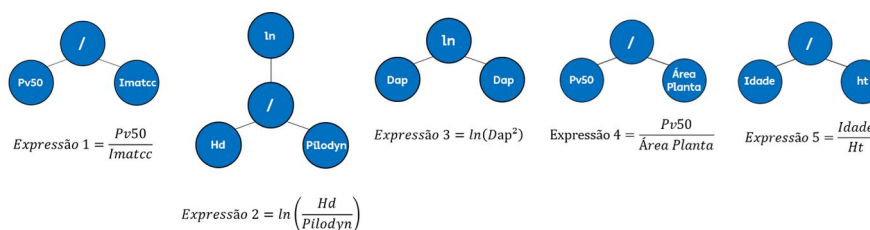
Figura 4 – Relação Iteração x RMSE na seleção e construção de expressões a serem utilizadas posteriormente para ajuste dos modelos PGM1(a) e PGM2(b).



Fonte: Do Autor (2023).

A seleção e construção das expressões através da PG para posterior ajuste de PGM1 resultou em um conjunto de variáveis que teoricamente seria difícil de ser encontrado por métodos convencionais de seleção (Figura 5). As combinações de variáveis (expressões) de 1 a 5 apresentaram as seguintes correlações com a DB: 35%, 53%, 38%, -61% e 47%, respectivamente.

Figura 5 - Expressões geradas pela PG a serem utilizadas no ajuste do modelo PGM1.

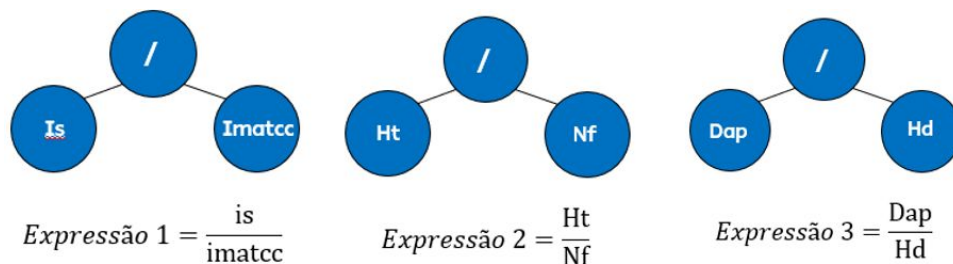


Fonte: Do Autor (2023).

Considerando a expressão 1, tem-se que, para florestas de mesma produtividade ( $I_{matcc}$ ), áreas mais heterogêneas tendem a ter uma menor DB do que florestas mais homogêneas. Por outro lado, ao avaliar duas florestas homogêneas, mas com produtividades distintas, observa-se que as florestas mais produtivas tendem a ter menores valores da variável de interesse se comparadas às áreas menos produtivas. Um entendimento semelhante pode ser extraído das expressões 2-5.

As expressões resultantes da PG para o posterior ajuste do modelo PGM2 podem ser observadas na Figura 6. A interpretação das expressões segue a mesma lógica discutida anteriormente. As expressões de 1 a 3 apresentaram as seguintes correlações com a DB: 19%, 72% e 46%, respectivamente.

Figura 6 - Expressões geradas pela PG a serem utilizadas no ajuste do modelo PGM2.



Fonte: Do Autor (2023).

### 3.2 Ajuste dos modelos de efeito misto PGM1 e PGM2

A modelagem de efeito misto para PGM1 considerou os diferentes clones como efeito aleatório do modelo. As variáveis independentes (descritas anteriormente) foram obtidas pela PG (Figura 5), resultando na Eq. 8, com parâmetros e métricas de avaliação apresentados na tabela 2:

$$DB = (\beta_0 + \mu_0) + \beta_1 \frac{Pv50}{I_{matcc}} + \beta_2 \ln\left(\frac{H_d}{Pilodyn}\right) + \beta_3 \ln(Dap^2) + \beta_4 \frac{Pv50}{Área\ planta} + (\beta_5 + \mu_5) \quad [8]$$

Em que: variáveis foram definidas anteriormente.

Tabela 2 - Parâmetros fixos e aleatórios do modelo ajustado PGM1.

<b>Efeito Fixo</b>				
<b>Parâmetros</b>	<b>Estimado</b>	<b>Erro Padrão</b>	<b>t-value</b>	<b>p-value</b>
Intercepto	416,214	14,95	27,83	0
Pv50/Imatcc	88,602	9,31	9,52	0
ln(Hd/Pilodyn)	123,106	13,46	9,15	0
ln(Dap <sup>2</sup> )	0,028	0,03	1,02	0,04
Pv50/Área planta	-10,867	2,15	-5,05	0
Idade/Ht	-258,311	48,17	-5,36	0
<b>Efeito Aleatório</b>				
<b>Parâmetros</b>	<b>Erro Padrão</b>	<b>Correlação</b>		
Intercepto	6,96	Intercepto	Hd/	
ln(Hd/Pilodyn)	6,67	-0,994		
Idade/Ht	48,24	-0,997	0,997	
Resíduos	18,96			

Fonte: Do Autor (2023).

De maneira geral, o modelo ajustado PGM1 (Eq. 8) apresentou um alto valor de R<sup>2</sup> (81%), erro médio tendendo a zero (4,58 x 10<sup>-13</sup> kg.m<sup>-3</sup>), além de um RMSE de 18,2 kg.m<sup>-3</sup>.

Posteriormente, com base nas expressões resultantes da segunda PG (conforme Figura 6), ajustou-se o modelo PGM2 (Eq. 9). Vale ressaltar que a DB para o modelo PGM2 é proveniente do modelo PGM1. Por fim, a tabela 3 apresenta os parâmetros e métricas de avaliação do modelo.

$$\widehat{DB} = (\beta_0 + \mu_0) + (\beta_1 + \mu_1) \frac{Is}{Imatcc} + (\beta_2 + \mu_2) \frac{Ht}{Nf} + (\beta_3 + \mu_3) \frac{Dap}{Hd} \quad [9]$$

Em que: variáveis foram definidas anteriormente.

Tabela 3 - Parâmetros fixos e aleatórios do PGM2.

<b>Efeito Fixo</b>				
<b>Parâmetros</b>	<b>Estimado</b>	<b>Erro Padrão</b>	<b>t-value</b>	<b>p-value</b>
Intercepto	276,272	26,80	10,31	0,00
Is/Imatcc	16,285	29,66	0,55	0,05
Ht/Nf	8159,767	1751,97	4,66	0,00
Dap/Hd	60,592	34,30	1,77	0,0833
<b>Efeito Aleatório</b>				
<b>Parâmetros</b>	<b>Erro Padrão</b>	<b>Correlação</b>		
Intercepto	42,12	Intercepto	isr/	ht/
Is/Imatcc	47,89	-0,998		
Ht/Nf	2874,32	0,388	-0,344	
Dap/Hd	54,82	-0,366	0,321	-1,00
Resíduos	14,20			

Fonte: Do Autor (2023).

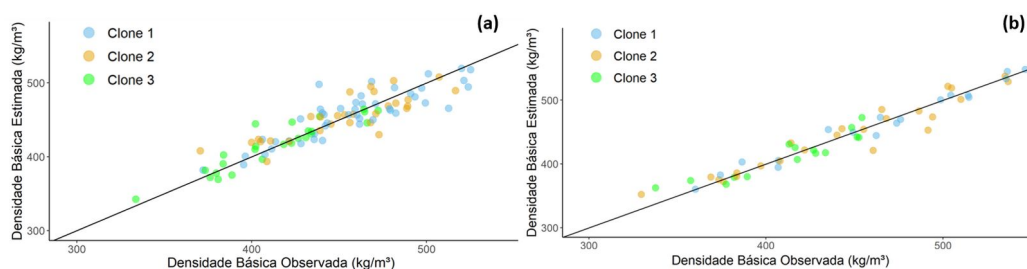
O modelo ajustado PGM2 apresentou qualidade dentro dos limites aceitáveis, com valor de  $R^2$  de 94%, erro médio tendendo a zero ( $-1,19 \times 10^{-13} \text{ kg.m}^{-3}$ ) e RMSE de  $13,2 \text{ kg.m}^{-3}$ .

### 3.3 Precisão e incerteza dos modelos PGM1 e PGM2

A Figura 7 evidencia a alta precisão das estimativas de DB geradas pelos modelos PGM1 (Eq. 8) e PGM2 (Eq. 9). De maneira geral, percebe-se que as estimativas do PGM1 têm, de fato, precisão para serem utilizadas como proxy da DB laboratorial.

Por fim, percebe-se que o modelo PGM2, mesmo sendo ajustado com a DB proveniente da estimativa resultante do PGM1, oferece alta precisão na quantificação da DB, o que valida a estratégia de modelagem proposta no SMIB.

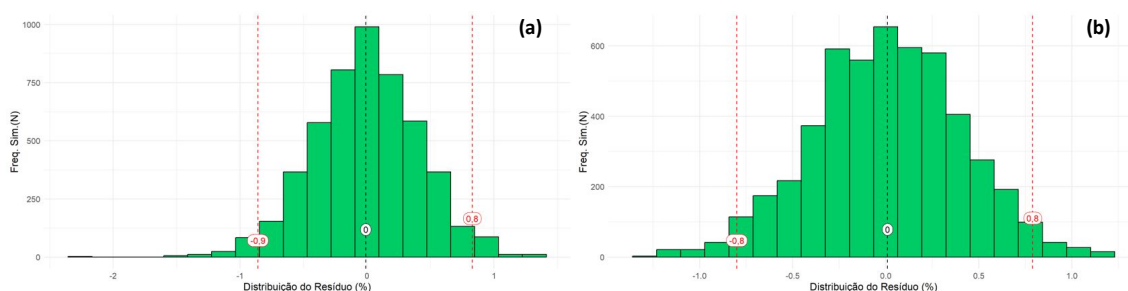
Figura 7 - DB observada em laboratório x DB estimada pelo PGM1 (a); DB observada em laboratório x DB estimada pelo PGM2(b).



Fonte: Do Autor (2023).

Por fim, a simulação bootstrap evidencia que a incerteza média tanto do modelo PGM1 (Eq. 8) quanto do modelo PGM2 (Eq. 9) é mínima (Figura 8). Os resultados permitem verificar que mesmo para combinações improváveis de bases de dados, ambos os modelos tem incerteza média tendendo a 0.

Figura 8 - Incerteza média através da simulação bootstrap de 5.000 bases de dados para os modelos PGM1 (Eq. 8) (a); e PGM2 (Eq. 9) (b).



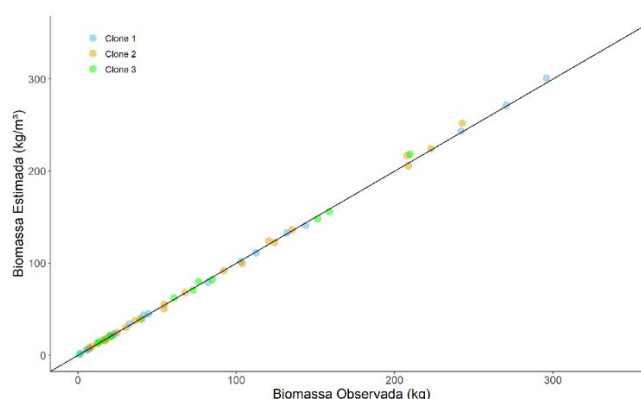
Fonte: Do Autor (2023).

### 3.4 Estimativa de biomassa ao nível de árvore por PGM2

A biomassa individual das árvores amostradas entre 2022/2023 foi calculada através da multiplicação do volume total com casca de cada árvore pela DB estimada através do modelo PGM2 (Eq. 9).

De maneira clara, a Figura 9 evidencia a alta aderência da biomassa estimada à biomassa observada, independentemente do clone observado e da classe de biomassa das árvores. Logo, valida-se o potencial uso do SMIB para inserção da informação de biomassa na rotina dos inventários florestais.

Figura 9 - Biomassa observada x biomassa estimada (DB estimada a partir da Eq. 9).



Fonte: Do Autor (2023).

## 4 DISCUSSÃO

O inventário florestal, convencionalmente, entrega ao final do processo de medição as informações em volume ( $m^3$ ). A área de planejamento de produção necessita, entretanto, de estimativas confiáveis também de biomassa, visando melhores tomadas de decisão (Thiersch et al., 2006; Anupam et al., 2015). Sendo assim, este estudo apresentou um sistema de modelos para geração do inventário de biomassa (SMIB). Os resultados reportados validam a abordagem proposta e tornam possível este avanço na obtenção de uma nova informação nos inventários florestais.

O inventário de biomassa sempre foi condicionado à necessidade de se obter de maneira rápida e precisa informações de DB dos diferentes clones, sendo que a grande amplitude genética presente nos plantios do Brasil sempre foi desafiadora nesse aspecto (Palermo et al., 2012).

O modelo PGM1 visa gerar uma estimativa de DB que funcione como uma proxy da DB laboratorial. Justifica-se assim o uso da técnica de modelos de efeito misto, uma vez que novos clones poderão ser absorvidos no modelo desenvolvido a partir da agregação de novos dados e reparametrização do mesmo (Scolforo et al., 2018), e da coleta da variável dureza (que apresenta alta correlação com a DB).

Pelo uso do PGM1, pode-se selecionar diversos locais e, através do uso do Pilodyn e informações biométricas, torna-se possível a geração de informações precisas da DB para os diferentes clones (Thiersch et al., 2006). Sendo assim, permite-se a geração de um grande banco de dados de maneira ágil e que, por fim, irá proporcionar o conhecimento da amplitude da DB dos diferentes clones (Gao et al., 2017; Schimleck et al., 2018; Schimleck et al., 2019). Thiersch et al. (2006) discutiram que modelos nesses

moldes seriam suficientes para a estimativa da DB. Por fim, PGM1 permite diminuir consideravelmente o número de amostras laboratoriais, sendo que a partir de agora as mesmas passam a ser utilizadas para validação contínua do modelo, assim como para obtenção de DB dos novos materiais genéticos.

Já o modelo PGM2 se beneficia das informações geradas pelo PGM1, e assim sendo, pode-se gerar um modelo de extrapolação com coeficientes controlados ao nível clonal e somente dependente de variáveis biométricas. Embora o sistema de modelos proposto (SMIB) quebre premissas estatísticas, onde não se pode utilizar variáveis estimadas por um modelo como input para outro (condição de acúmulo de vieses), a metodologia proposta segue o mesmo conceito de modelos de crescimento e produção.

Por fim, a PG se mostrou uma técnica robusta e capaz de auxiliar na detecção de uma combinação ótima de expressões em busca da maximização da precisão das estimativas dos modelos ajustados (Muni et al., 2006; Aryadous, 2015). O processo heurístico permitiu uma busca exaustiva dentro da bacia de soluções e na detecção de regiões ótimas que auxiliem na explicação da variável dependente (Lu et al., 2016; Kammerer, 2020). A PG também abre um novo caminho para a avaliação de expressões selecionadas que impactam a variável de interesse (Zhang e Wong, 2008; Niekum et al., 2010; Taghizadeh-Mehrjardi, 2016).

Ainda há lacunas a serem preenchidas, como por exemplo, avaliar se o SMIB mantém a precisão das estimativas para períodos de mudança no comportamento do regime hídrico no curto prazo. Além disso, é importante avaliar como as demais características da madeira interferem na produção de celulose. Isso permitirá que os inventários possam, cada vez mais, reportar informações mais próximas às aquelas esperadas pelos setores industriais de processamento da madeira. Finalmente, cabe pontuar que este estudo apresenta um sistema inovador para a eucaliptocultura do Brasil, o qual certamente permitirá evoluções na linha do planejamento da produção anual.

## 5 CONCLUSÃO

O SMIB tem amplo potencial de operacionalização na rotina dos inventários florestais. O estudo demonstra que o SMIB ainda necessita manter, em menor escala, amostras de DB laboratorial tanto para contínua validação do sistema de modelos quanto para absorção no sistema dos novos clones desenvolvidos pelo melhoramento

florestal. Por fim, SMIB também permite gerar diversos insights a respeito de variáveis extrínsecas e como estas explicam o comportamento da DB de diferentes clones e ao longo do gradiente ambiental.



## REFERÊNCIAS

- ABNT - Associação Brasileira de Normas Técnicas. **NBR 11941: Madeira-Determinação da densidade básica**, 2003.
- AKINDELE, S. O.; LEMAY, V. M. Development of tree volume equations for common timber species in the tropical rain forest area of Nigeria. **Forest Ecology and Management**, Amsterdam, v. 226, n. 1-3, p. 41-48, 2006.
- ALVARES, C. A. et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, Berlin, v. 22, n. 6, p. 711-728, 2013.
- ANUPAM, K. et al. Sustainable raw material selection for pulp and paper using SAW multiple criteria decision making design. **IPPTA J**, v. 27, n. 1, p. 67-76, 2015.
- ARYADOUS, V. Application of evolutionary algorithm-based symbolic regression to language assessment: Toward nonlinear modeling. **Psychological Test and Assessment Modeling**. v. 57, p. 301-337, 2015.
- ASSMANN, E. The principles of forest yield study. 1970.
- CHICHORRO, J. F.; RESENDE, J. L. P.; LEITE, H. G. Equações de volume e de taper para quantificar multiprodutos da madeira em floresta atlântica. **Revista Árvore**, Viçosa v. 27, p. 799-809, 2003.
- CLUTTER, J. L. Development of taper functions from variable-top merchantable volume equations. **Forest Science**, v. 26, n. 1, p. 117-120, 1980.
- DE ALMEIDA, M. N. F. et al. Heartwood variation of *Eucalyptus urophylla* is influenced by climatic conditions. *Forest Ecology and Management*, v. 458, p. 117743, 2020.
- DE OLIVEIRA, R. A. et al. Aplicação de modelos lineares de efeitos mistos para avaliar a densidade básica da madeira de duas espécies e um híbrido de *Eucalyptus*. **Scientia Florestalis**, Piracicaba, v.49, n.129, 2021.
- DOS SANTOS, P. E. T.; GARCIA, J. N.; GERALDI, I. O. Posição da tora na árvore e sua relação com a qualidade da madeira serrada de *Eucalyptus grandis*. **Scientia Florestalis**, Piracicaba, n. 66, p. 142 - 151, 2004.
- EMBRAPA, et al. Sistema brasileiro de classificação de solos. **Centro Nacional de Pesquisa de Solos - Empresa Brasileira de Pesquisa Agropecuária**, Rio de Janeiro, v. 3, 2013.
- FERREIRA, C. A.; FREITAS, M.; FERREIRA, M.. Densidade básica da madeira de plantações comerciais de eucaliptos, na região de Mogi-Guaçu (SP). **IPEF**, Piracicaba, v. 18, p. 106-117, 1979.
- FIGUEIREDO-FILHO, A.; BORDERS, B. E.; HITCH, K. L. Taper equations for *Pinus taeda* plantations in Southern Brazil. **Forest Ecology and Management**, Amsterdam, v. 83, n. 1-2, p. 39-46, 1996.

FOELKEL, C. E. B.; MORA, E.; MENOCELLI, S. Densidade básica: sua verdadeira utilidade como índice de qualidade na madeira de eucalipto para produção de celulose. **O papel**, Campos do Jordão, v. 53, n. 5, p. 35 - 40, 1992.

FOX, J. et al. The car package. **R Foundation for Statistical Computing**, v. 1109, p. 1431, 2007.

FUNDOVA, I.; FUNDA, T.; WU, H.X. Non-destructive wood density assessment of Scots pine (*Pinus sylvestris* L.) using Resistograph and Pilodyn. **PLoS ONE**, São Francisco, 13, v.13, n. 9 p. e0204518, 2018.

GALLO, R. et al. Growth and wood quality traits in the genetic selection of potential *Eucalyptus dunnii* Maiden clones for pulp production. **Industrial Crops and Products**, v. 123, p. 434-441, 2018.

GAO, S. et al. A critical analysis of methods for rapid and nondestructive determination of wood density in standing trees. **Annals of Forest Science**, v. 74, n. 2, p. 27, 2017.

GOUVÊA, A. F. G. et al. Determinação da densidade básica da madeira de *Eucalyptus* por diferentes métodos não destrutivos. **Revista Árvore**, Viçosa, v. 3, n. 2, p. 349-358, fev. 2011b.

GUILLEY, E.; HERVÉ, J.; NEPVEU, G. The influence of site quality, silviculture and region on wood density mixed model in *Quercus petraea* Liebl. **Forest Ecology and Management**, Amsterdam, v. 189, n. 1-3, p. 111-121, 2004.

HAKAMADA, R. E. et al. Uniformidade entre árvores durante uma rotação e sua relação com a produtividade em *Eucalyptus* clonais. **Revista Cerne**, Lavras, v. 21, p. 465-472, 2015.

HEIN, P. R. G. et al. Near infrared spectroscopy for estimating wood basic density in *Eucalyptus urophylla* and *Eucalyptus grandis*. **Cerne**, Lavras, v.15, n.2, p.133-141, 2009.

HUSSEIN, F.; KHARMA, N.; WARD, R. Genetic algorithms for feature selection and weighting, a review and study. In: Proceedings of sixth international conference on document analysis and recognition. **IEEE**, p. 1240-1244. 2001.

KAMMERER, L. et al. Symbolic regression by exhaustive search: Reducing the search space using syntactical constraints and efficient semantic structure deduplication. **Genetic programming theory and practice XVII**, p. 79-99, 2020.

KOZA, J. R. Genetic programming: A paradigm for genetically breeding populations of computer programs to solve problems. Stanford University, Department of Computer Science, Stanford, 1990.

KOZAK, A.; MUNRO, D. D.; SMITH, J. H. G. Taper functions and their application in forest inventory. **The Forestry Chronicle**, v. 45, n. 4, p. 278-283, 1969.

LEARDI, R. Genetic algorithms in feature selection. In: Genetic algorithms in molecular modeling. **Academic Press**, p. 67-86. 1996.

- LEI, S. A feature selection method based on information gain and genetic algorithm. **IEEE**, p. 355-358. 2012.
- LU, Q.; REN, J.; WANG, Z. Using genetic programming with prior formula knowledge to solve symbolic regression problem. **Computational intelligence and neuroscience**, v. 2016, p. 1-1, 2016.
- MA, X. et al. A genetic programming approach for real-time crash prediction to solve trade-off between interpretability and accuracy. **Journal of Transportation Safety & Security**, v. 15, n. 4, p. 421-443, 2023.
- MENDONÇA, A. R.; CARVALHO, S. P. C.; CALEGARIO, N. Modelos hipsométricos generalizados mistos na predição da altura de *Eucalyptus sp.* **Cerne**, Lavras, v. 21, p. 107-115, 2015.
- MUNI, D. P.; PAL, N. R.; DAS, J. Genetic programming for simultaneous feature selection and classifier design. **IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)**, v. 36, n. 1, p. 106-117, 2006.
- NAZIR, K. et al. Energy consumption predictions by genetic programming methods for PCM integrated building in the tropical savanna climate zone. **Journal of Building Engineering**, v. 68, p. 106115, 2023.
- NIEKUM, S.; BARTO, A. G.; SPECTOR, L. Genetic programming for reward function search. **IEEE Transactions on Autonomous Mental Development**, v. 2, n. 2, p. 83-90, 2010.
- NOORIAN, F.; DE SILVA, A. M.; LEONG, P. H. W. gramEvol: Grammatical evolution in R. **Journal of Statistical Software**, v. 71, p. 1-26, 2016.
- PALERMO, G. P. M. et al. Determinação da densidade da madeira de *Pinus elliottii* Engelm, através de atenuação de radiação gama comparada a métodos tradicionais. **Floresta e Ambiente**, v. 11, n. 1, p. 1-6, 2012.
- PARRESOL, B. R.; HOTVEDT, J. E.; CAO, Q. V. A volume and taper prediction system for bald cypress. **Canadian Journal of Forest Research**, v. 17, n. 3, p. 250-259, 1987.
- PINHEIRO, J. C. et al. nlme: Linear and nonlinear mixed effects models. **R Package version 3.1-141** Disponível em: <<https://CRAN.R-project.org/package=nlme>> (Acesso em: 25 de maio de 2023), 2019.
- PINHEIRO, J. C., BATES, D. M. Mixed-Effects Models in Sand S-PLUS. **Springer Science & Business Media**, Nova Yorque, 2006.
- R CORE TEAM. R: A Language and Environment for Statistical Computing. **R Foundation for Statistical Computing**, Vienna. Disponível em: <<http://www.R-project.org>> (Acesso em: 10 de outubro de 2021), 2023.
- ROCHA, S. M. G. et al. Influence of climatic variations on production, biomass and density of wood in eucalyptus clones of different species. **Forest Ecology and Management**, Amsterdam v. 473, p. 118290, 2020.

- SAKALLE, A. et al. Genetic programming-based feature selection for emotion classification using EEG signal. **Journal of Healthcare Engineering**, v. 2022, 2022.
- SCHIMLECK, L. R. et al. Comparison of methods for estimating mechanical properties of wood by NIR spectroscopy. **Journal of Spectroscopy**, v. 2018, 2018.
- SCHIMLECK, L. R. et al. Non-destructive evaluation techniques and what they tell us about wood property variation. **Forests**, v. 10, n. 9, p. 728, 2019.
- SCOLFORO, H. F. et al. Generalized stem taper and tree volume equations applied to eucalyptus of varying genetics in Brazil. **Canadian Journal of Forest Research**, v. 49, n. 5, p. 447-462, 2019.
- SCOLFORO, H. F. et al. *Eucalyptus* growth and yield system: Linking individual-tree and stand-level growth models in clonal Eucalypt plantations in Brazil. **Forest Ecology and Management**, Amsterdam, v. 432, p. 1–16, 2019.
- TAGHIZADEH-MEHRJARDI, R.; NABIOLLAHI, K.; KERRY, R. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. **Geoderma**. v. 266, p. 98-110, 2016.
- THIERSCH, C. R. et al. O uso de modelos matemáticos na estimativa da densidade básica da madeira em plantios de clones de *Eucalyptus sp.* **Cerne**, Lavras v. 12, n. 3, p. 264-278, 2006.
- XAVIER, A. C.; KING, C. W.; SCANLON, B. R. Daily gridded meteorological variables in Brazil (1980–2013). **International Journal of Climatology**, v. 36, n. 6, p. 2644-2659, 2016.
- WICKHAM, H., ggplot2: Elegant Graphics for Data Analysis. **Springer**, Nova Yorque. Disponível em: <<https://cran.r-project.org/web/packages/ggplot2/index.html>> (Acesso em: 25 de maio de 2023). 2016.
- WIEMANN, M. C.; WILLIAMSON, G. B. Biomass determination using wood specific gravity from increment cores. United States Department of Agriculture, Forest Service, **Forest Products Laboratory**, 2013.
- WU, S. et al. Use of the pilodyn for assessing wood properties in standing trees of *Eucalyptus* clones. **Journal of Forestry Research**, v. 21, n. 1, p. 68-72, 2010.
- ZHANG, M.; WONG, P. Genetic programming for medical classification: a program simplification approach. **Genetic Programming and Evolvable Machines**, v. 9, p. 229-255, 2008.