

UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
CAMPUS MONTE CARMELO  
INSTITUTO DE CIÊNCIAS AGRÁRIAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRICULTURA E INFORMAÇÕES  
GEOESPACIAIS – PPGAIG

JOÃO VICTOR DO NASCIMENTO LIMA

MODELAGEM DO CRESCIMENTO E PRODUÇÃO DE FLORESTAS  
PLANTADAS USANDO INDICADORES DE CAPACIDADE FOTOSSINTÉTICA  
DERIVADOS DE SENSORIAMENTO REMOTO E APRENDIZADO DE  
MÁQUINAS

Monte Carmelo – MG

2025

UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
CAMPUS MONTE CARMELO  
INSTITUTO DE CIÊNCIAS AGRÁRIAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRICULTURA E INFORMAÇÕES  
GEOESPACIAIS – PPGAIG

JOÃO VICTOR DO NASCIMENTO LIMA

MODELAGEM DO CRESCIMENTO E PRODUÇÃO DE FLORESTAS  
PLANTADAS USANDO INDICADORES DE CAPACIDADE FOTOSSINTÉTICA  
DERIVADOS DE SENSORIAMENTO REMOTO E APRENDIZADO DE  
MÁQUINAS

Dissertação apresentada ao Programa de Pós-Graduação em Agricultura e Informações Geoespaciais da Universidade Federal de Uberlândia, Campus Monte Carmelo, como parte das exigências para obtenção do título de “Mestre” sob orientação do Prof. Drº. George Deroco Martins e coorientação Prof. Drº. Álvaro Augusto Vieira Soares.

Monte Carmelo – MG

2025

Dados Internacionais de Catalogação na Publicação (CIP)  
Sistema de Bibliotecas da UFU, MG, Brasil.

---

L732m  
2025      Lima, João Victor do Nascimento, 2001-  
Modelagem do crescimento e produção de florestas plantadas usando indicadores de capacidade fotossintética derivados de sensoriamento remoto e aprendizado de máquinas [recurso eletrônico] / João Victor do Nascimento Lima. - 2025.

Orientador: George Deroco Martins.

Coorientador: Álvaro Augusto Vieira Soares.

Dissertação (Mestrado) - Universidade Federal de Uberlândia, Programa de Pós-graduação em Agricultura e Informações Geoespaciais.

Modo de acesso: Internet.

Disponível em: <http://doi.org/10.14393/ufu.di.2026.5022>

Inclui bibliografia.

Inclui ilustrações.

1. Agricultura. I. Martins, George Deroco, 1987-, (Orient.). II. Soares, Álvaro Augusto Vieira, 1986-, (Coorient.). III. Universidade Federal de Uberlândia. Programa de Pós-graduação em Agricultura e Informações Geoespaciais. IV. Título.

---

CDU: 631

Nelson Marcos Ferreira  
Bibliotecário-Documentalista - CRB-6/3074


**UNIVERSIDADE FEDERAL DE UBERLÂNDIA**
**Coordenação do Programa de Pós-Graduação em Agricultura e  
Informações Geoespaciais**

 Rodovia LMG 746, Km 01, s/nº, Bloco 1AMC, Sala 1A202, Monte Carmelo-MG, CEP 38.500-000 Telefone:  
(34) 3810-1033 - ppgaig@iciag.ufu.br

**ATA DE DEFESA - PÓS-GRADUAÇÃO**

Programa de Pós-Graduação em:	Agricultura e Informações Geoespaciais				
Defesa de:	Dissertação de Mestrado Acadêmico				
Data:	12/12/2025	Hora de início:	14:00h	Hora de encerramento:	17h e 20 min
Matrícula do Discente:	32322AIG009				
Nome do Discente:	João Victor do Nascimento Lima				
Título do Trabalho:	Incorporando variáveis proxy de capacidade fotossintética a modelos de crescimento e produção florestal por meio de sensoriamento remoto e aprendizado de máquinas				
Área de concentração:	Informações geoespaciais e tecnologias aplicadas à produção agrícola				
Linha de pesquisa:	Desenvolvimento e Aplicações de Métodos em Informações Geoespaciais				
Projeto de Pesquisa de vinculação:	Mapeamento de doenças e variáveis agrícolas em culturas agrícolas a partir de dados de sensoriamento remoto				

Reuniu-se em sala virtual a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Agricultura e Informações Geoespaciais, composta pelo prof. Dr. George Deroco Martins (Universidade Federal de Uberlândia-UFU), orientador do candidato, prof. Dr. Carlos Alberto Araújo Júnior (Universidade Federal de Minas Gerais - UFMG - Montes Claros), e prof. Dr. Diogo Nepomuceno Cosenza (Universidade Federal de Viçosa - UFV).

Iniciados os trabalhos, o presidente da mesa, Dr. George Deroco Martins, apresentou a Comissão Examinadora e o discente, agradeceu a presença do público e concedeu-lhe a palavra para a exposição de seu trabalho. A duração da apresentação e o tempo destinado à arguição e às respostas ocorreram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a banca, em sessão secreta, atribuiu o resultado final, considerando a candidato: **Aprovado**

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.

[https://www.sei.ufu.br/sei/controlador.php?acao=procedimento\\_trabalhar&acao\\_origem=procedimento\\_controlar&acao\\_retorno=procedimento\\_co...](https://www.sei.ufu.br/sei/controlador.php?acao=procedimento_trabalhar&acao_origem=procedimento_controlar&acao_retorno=procedimento_co...)  
1/2 18/12/2025, 08:20 SEI/UFU - 6924782 - Ata de Defesa - Pós-Graduação



Documento assinado eletronicamente por **George Deroco Martins, Professor(a) do Magistério Superior**, em 15/12/2025, às 09:26, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Diogo Nepomuceno Cosenza, Usuário Externo**, em 16/12/2025, às 08:31, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Carlos Alberto Araújo Júnior, Usuário Externo**, em 16/12/2025, às 17:26, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://www.sei.ufu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **6924782** e o código CRC **93801458**.

Referência: Processo nº 23117.087370/2025-06

SEI nº 6924782

## **AGRADECIMENTOS**

Gostaria de agradecer a Deus pela minha vida, vida dos meus pais e da minha família. Agradeço ao apoio e incentivo dos meus pais, Humberto Paula Lima e Angela Maria Silva do Nascimento, pela paciência e educação que me deram. Agradeço aos meus irmãos Paulo Junior do Nascimento Lima e Ayrton Vannuti do Nascimento Lima, que me proporcionaram conversar e forças para que eu pudesse seguir em frente. Agradeço a minha namorada Ketila Silva por todo apoio e carinho nos momentos de mais necessidade.

Agradeço também ao meu orientador George Deroco Martins e coorientador Álvaro Augusto Viera Soares por todos os ensinamentos e aprendizados durante a minha jornada no mestrado. Aos membros da banca, professor Diogo Nepomuceno Cosenza e professor Carlos Alberto Araújo Júnior.

Por fim, agradeço a Universidade Federal de Uberlândia, o Programa de Pós-Graduação em Agricultura e Informações Geoespaciais por todo apoio e estrutura técnica e intelectual.

## **BIOGRAFIA**

João Victor do Nascimento Lima é graduado em Engenharia de Agrimensura e Cartográfica pela Universidade Federal de Uberlândia (UFU), curso iniciado em 2018 e concluído em 2023. Durante a graduação, integrou o Programa de Educação Tutorial (PET – Engenharia de Agrimensura e Cartográfica) entre os anos de 2020 e 2023, desenvolvendo atividades de ensino, pesquisa e extensão.

Atualmente é mestrando no Programa de Pós-Graduação em Agricultura e Informações Geoespaciais, onde dedica seus estudos à agricultura e silvicultura de precisão, com ênfase em sensoriamento remoto e aplicações de inteligência artificial. Seu interesse científico concentra-se na modelagem, análise espacial e desenvolvimento de técnicas avançadas para monitoramento e tomada de decisão em sistemas agrícolas e florestais.

## SUMÁRIO

<b>RESUMO.....</b>	<b>I</b>
<b>ABSTRACT.....</b>	<b>II</b>
<b>INTRODUÇÃO.....</b>	<b>1</b>
<b>REFERENCIAL TEÓRICO .....</b>	<b>4</b>
1.1 CRESCIMENTO DE FLORESTAS PLANTADAS .....	4
1.2 MODELAGEM DO CRESCIMENTO E DA PRODUÇÃO FLORESTAL .....	7
1.3 SENSORIAMENTO REMOTO PARA PREDIÇÃO DE ESTOQUE .....	9
1.3.1 Índices de textura derivados de sensores ópticos.....	11
1.4 MODELOS DE APRENDIZADO DE MÁQUINA (AM) .....	14
1.4.1 Floresta Aleatória (RF) .....	15
1.4.2 Aumento de gradiente extremo (XGBoost) .....	17
1.4.3 Máquina de Vetores de Suporte (SVM).....	19
1.4.4 Redes Neurais Artificiais (RNA).....	21
1.4.5 K-Vizinhos próximo.....	23
1.5 APRENDIZADO DE MÁQUINA NA MODELAGEM FLORESTAL .....	24
<b>MATERIAL E MÉTODOS.....</b>	<b>26</b>
1.6 MATERIAL .....	26
1.6.1 área de estudo .....	26
1.6.2 Dados de campo .....	27
1.6.3 Variáveis de sensoriamento remoto .....	29
1.7 PROCESSAMENTO DOS DADOS.....	30
1.7.1 Ajuste do modelo de Clutter .....	34
1.7.2 Treinamento dos modelos de aprendizado de máquina .....	36
1.7.3 Avaliação dos modelos.....	38
1.8 CRITÉRIOS DE SELEÇÃO DE MODELOS .....	38
<b>RESULTADOS .....</b>	<b>39</b>
1.9 ANÁLISE EXPLORATÓRIA.....	39
1.9.1 Variáveis dendrométricas x idade.....	<i>Erro! Indicador não definido.</i>
1.9.2 Índices espectrais x Idade.....	<i>Erro! Indicador não definido.</i>
1.9.3 Seleção do modelo de índice de sítio .....	45
1.9.4 modelo de Clutter e Clutter + IE/IT.....	46
1.9.5 modelos de aprendizado de máquina.....	47
<b>DISCUSSÕES.....</b>	<b>49</b>
<b>CONCLUSÃO .....</b>	<b>52</b>
<b>REFERENCIAL TEÓRICO .....</b>	<b>52</b>
<b>APENDICE .....</b>	<b>62</b>

## RESUMO

A modelagem do crescimento e da produção florestal é fundamental para o planejamento e manejo sustentável de florestas plantadas. O modelo de Clutter, amplamente utilizado no Brasil, baseia-se exclusivamente em variáveis dendrométricas convencionais (idade, área basal e índice de sítio), não considerando fatores fisiológicos como a capacidade fotossintética das plantas. Assim este estudo teve como objetivo avaliar a incorporação de variáveis proxy de capacidade fotossintética, na modelagem do crescimento e produção florestal. A pesquisa foi conduzida em plantios de eucalipto na mesorregião do Vale do Rio Doce, Minas Gerais, utilizando dados de inventário florestal contínuo (2013-2019) fornecidos pela empresa CENIBRA. Foram extraídos valores de reflectância de imagens Landsat 8 sem cobertura de nuvens, calculando-se oito índices espectrais (NDVI, GNDVI, SAVI, SR, NBRI, RVI, DVI, EVI) e oito índices de textura baseados na Matriz de Co-ocorrência de Níveis de Cinza (Energia, Entropia, Correlação, IDM, Inércia, CS, CP, HC). Cinco algoritmos de aprendizado de máquina foram testados: Florestas Aleatórias (RF), Máquina de Vetores Suporte (SVM), Aumento de Gradiente Extremo (XGBoost), Redes Neurais Artificiais (RNA) e K-Vizinhos Próximos (KNN). A validação foi realizada através de validação cruzada com 5 folds, utilizando métricas de acurácia REMQ para selecionar o melhor modelo. Os resultados demonstraram que o modelo de Clutter apresentou elevada precisão, com RMSE% inferior a 10% para volume,  $R^2$  superior a 0,93 e  $r = 0,97$ . A incorporação de índices espectrais e texturais no modelo de Clutter promoveu melhorias modestas, reduzindo o RMSE% de 8,51% para 8,46% com os índices SR e NBRI. Entretanto, os algoritmos de aprendizado de máquina, particularmente o SVM, demonstraram melhor capacidade de incorporar essas variáveis, resultando em ganhos de acurácia de até 0,7% em relação ao modelo tradicional. Os índices SR, NBRI e HC destacaram-se como os melhores preditores. Conclui-se que a inclusão de variáveis de sensoriamento remoto pode aprimorar discretamente a capacidade preditiva dos modelos tradicionais, sendo os algoritmos de aprendizado de máquina, particularmente o SVM, demonstraram melhor capacidade de incorporar essas variáveis, indicando um caminho promissor para o aperfeiçoamento dos modelos de crescimento e produção florestal.

Palavras-chaves: Manejo florestal, sensoriamento remoto, Clutter, aprendizado de máquina

## ABSTRACT

Forest growth and yield modeling is fundamental for sustainable planning and management of planted forests. The Clutter model, widely used in Brazil, relies exclusively on conventional dendrometric variables (age, basal area, and site index), without considering physiological factors such as photosynthetic capacity. This study aimed to evaluate the incorporation of proxy variables for photosynthetic capacity, derived from remote sensing, into forest growth models, comparing the performance of the traditional Clutter model, its modified variant with spectral and textural indices, and machine learning algorithms. The research was conducted in commercial eucalyptus plantations in the Vale do Rio Doce region, Minas Gerais, using continuous forest inventory data (2013-2019) provided by CENIBRA. Reflectance values were extracted from cloud-free Landsat 8 images, calculating eight spectral indices (NDVI, GNDVI, SAVI, SR, NBRI, RVI, DVI, EVI) and eight texture indices based on the Gray Level Co-occurrence Matrix (Energy, Entropy, Correlation, IDM, Inertia, CS, CP, HC). Five machine learning algorithms were tested: Random Forest (RF), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Artificial Neural Networks (ANN), and K-Nearest Neighbors (KNN). Validation was performed through 5-fold cross-validation, using RMSE, RMSE%, Pearson correlation ( $r$ ),  $R^2$ , and BIAS% metrics. Results demonstrated that the traditional Clutter model exhibited high accuracy, with RMSE% below 10% for volume,  $R^2$  greater than 0.93, and  $r = 0.97$ . The incorporation of spectral and textural indices into the Clutter model promoted modest improvements, reducing RMSE% from 8.51% to 8.46% with SR and NBRI indices. However, machine learning algorithms, particularly SVM, demonstrated superior capability in incorporating these variables, resulting in accuracy gains of up to 0.7% compared to the traditional model. The SR, NBRI, and HC indices stood out as the best predictors. It is concluded that the inclusion of remote sensing variables can discretely enhance the predictive capacity of traditional models, with machine learning algorithms being more efficient in incorporating these variables, indicating a promising path for improving forest growth and yield models.

Keywords: Forest management, remote sensing, Clutter, machine learning

## INTRODUÇÃO

O setor florestal brasileiro ocupa uma posição de destaque no cenário mundial, com cerca de 10 milhões de hectares de florestas plantadas. Aproximadamente 75% dessa área é composta por plantações de eucalipto, enquanto os 25% restantes são distribuídos entre espécies como *Pinus spp.*, seringueira (*Hevea brasiliensis*), acácia (*Acacia mangium*), teca (*Tectona grandis*) e paricá (*Schizolobium amazonicum*) (IBÁ, 2022). A concentração geográfica dos plantios de eucalipto evidencia a importância econômica desse cultivo, com 44% da área total localizada na região Sudeste, onde o estado de Minas Gerais se destaca como responsável por 63% da produção regional. As demais regiões apresentam a seguinte distribuição: Centro-Oeste (20%), Sul (18%), Nordeste (13%) e Norte (5%) (IBÁ, 2022).

A relevância socioeconômica do setor é incontestável. Em 2023, o Brasil registrou saldo positivo de 1.483.598 empregos formais, enquanto a silvicultura gerou 33 mil novos postos de trabalho, acumulando 2,69 milhões de empregos diretos e indiretos, sendo 690 mil empregos diretos. Além disso, o setor contribuiu significativamente para a economia nacional, gerando R\$ 24,3 bilhões em tributos federais e estaduais no mesmo ano (IBÁ, 2022). Para além do aspecto econômico, as florestas plantadas desempenham múltiplas funções ecossistêmicas, incluindo sequestro de carbono, proteção do solo, conservação da biodiversidade e regulação do ciclo hidrológico (Sanquetta et al., 2018).

O eucalipto, denominação atribuída ao conjunto de espécies e híbridos dos gêneros *Eucalyptus*, *Angophora* e *Corymbia*, destaca-se no setor florestal brasileiro em razão de suas elevadas taxas de crescimento, ampla versatilidade de uso e notável capacidade de adaptação a distintas condições edafoclimáticas. Dentre estes três gêneros, são cultivados em larga escala pelo setor florestal o gênero *Eucalyptus* e *Corymbia*. Dada relevância econômica e ambiental, plantio de eucalipto têm sido intensamente estudados em diversos aspectos como, por exemplo, sobre ciclagem de nutrientes, potencial energético, crescimento e produção florestal (Fiorentin et al., 2015), balanço de carbono (Baesso et al., 2010), dinâmica da evapotranspiração e segurança hídrica (Brasil, 2024), dentre outros.

Diante desse contexto, o manejo eficiente das florestas plantadas depende de estimativas confiáveis sobre o estoque futuro de madeira produzida por estes povoamentos. Essas estimativas fornecem subsídios essenciais para análises econômicas e tomadas de decisão em diferentes escalas temporais e espaciais (Figueiredo et al., 2014; Melo et al., 2017). Como

destacam Campos & Leite (2006), o sucesso do manejo florestal está intrinsecamente associado à capacidade de projetar cenários futuros de produção com elevado grau de confiabilidade.

Nesse cenário, os modelos matemáticos de crescimento e produção florestal são ferramentas indispensáveis para a gestão florestal. Entre os diversos modelos desenvolvidos ao longo das últimas décadas, o modelo de Clutter (1963) destaca-se como um dos mais amplamente utilizados para a prognose da produtividade madeireira de floresta plantadas no Brasil. Em sua formulação original, esse modelo utiliza exclusivamente variáveis obtidas a partir do inventário florestal, como idade, área basal e índice de sítio.

O modelo de Clutter tem demonstrado desempenho consistente e satisfatório em diversos estudos aplicados a florestas de eucalipto no Brasil. Campos & Leite (2006) reportam que o modelo apresenta coeficientes de determinação ( $R^2$ ) superiores a 0,90 para projeção de área basal e 0,85 para volume, com erros percentuais geralmente inferiores a 10% em plantios comerciais bem manejados. Scolforo et al. (2008) avaliaram o modelo de Clutter em povoamentos de *Eucalyptus grandis* em Minas Gerais, obtendo  $R^2$  de 0,94 para área basal e 0,91 para volume, com erro padrão da estimativa de 8,7% e 11,3%, respectivamente. Dias et al. (2005) constataram que o modelo apresentou boa precisão na estimativa de volume para *Eucalyptus* spp., com erro padrão residual de 9,2% e distribuição homogênea dos resíduos.

Apesar de seu ótimo desempenho, pesquisadores têm buscado alternativas para aprimorar os modelos de crescimento e produção florestal. Leite et al. (2011) desenvolveram uma modificação incorporando a variável área basal remanescente após desbaste, ampliando a aplicabilidade do modelo para povoamentos submetidos a diferentes intensidades de desbaste. Silva et al. (2016) propuseram a inclusão de variáveis climáticas (precipitação e temperatura) no modelo de Clutter para eucalipto, obtendo redução de 12% no erro padrão das estimativas em comparação com o modelo original. Téó et al. (2011) sugeriram modificações para incluir o efeito de geadas em plantios de *Pinus taeda*, através da inserção de variáveis indicadoras de ocorrência de danos. Parte deste aprimoramento tem sido no sentido de incorporar variáveis provenientes de sensoriamento remoto. Pereira et al. (2016) demonstraram que a inserção de uma variável espacial no modelo de Clutter pode melhorar o desempenho da predição de volume em até 6,57% quando incorporada à equação da área basal.

Uma limitação fundamental dos modelos tradicionais é não considerar fatores fisiológicos que impactam o crescimento florestal. Por exemplo, a incorporação de variáveis que reflitam o vigor da copa e sua consequente capacidade fotossintética pode aprimorar as

estimativas. Como o modelo de Clutter é um modelo de projeção, que utiliza o estado atual do povoamento para estimar uma condição futura, variáveis que representem a capacidade fotossintética das plantas no momento atual podem, em princípio, atuar como bons preditores adicionais. O crescimento das árvores está intrinsecamente ligado à capacidade fotossintética, sendo a folha a estrutura responsável pela captação da radiação solar e sua conversão em fotoassimilados que constituem a biomassa (Binkley et al., 2003; Coops, 2015). Esta variável fisiológica crucial não é contemplada na formulação tradicional do modelo de Clutter, representando uma lacuna significativa na modelagem do crescimento florestal.

De forma complementar, índices espectrais derivados do sensoriamento remoto, como o NDVI (*Normalized Difference Vegetation Index*), têm se mostrado promissores por sua correlação com parâmetros biofísicos das florestas, incluindo biomassa, área foliar e vigor vegetativo (Viana et al., 2016). A utilização de índices de vegetação como a Razão Simples (SR) e o Soil-Adjusted Vegetation Index (SAVI) tem demonstrado eficácia nas estimativas de volume de madeira e biomassa, com erros-padrão em torno de 12,34% em áreas de cerrado, conforme relatado por Miguel et al. (2015). Os avanços dos sistemas orbitais de observação da Terra têm possibilitado a aquisição de informações relacionadas ao vigor vegetativo em grandes áreas, com alta frequência temporal e relativamente baixo custo.

Além dos índices espectrais tradicionais, uma nova dimensão de informação vem ganhando destaque na comunidade científica: os índices de textura derivados de sensores ópticos. Estes índices capturam a variabilidade espacial dos valores de pixels em uma imagem, fornecendo informações complementares sobre a estrutura do dossel florestal que não são capturadas pelos índices espectrais convencionais. A textura da imagem pode revelar aspectos da heterogeneidade estrutural das florestas, relacionando-se diretamente com características como densidade do povoamento, estrutura do dossel e, conseqüentemente, com a biomassa florestal. Estudos têm demonstrado que modelos incorporando métricas de textura apresentam correlações superiores a 0,86 com biomassa observada, superando significativamente abordagens baseadas apenas em variáveis espectrais (Kelsey & Neff, 2014).

Além das propostas de modificação dos modelos tradicionais fundamentadas em regressões lineares e não lineares, observa-se uma evolução significativa nas abordagens analíticas disponíveis. Os modelos de aprendizado de máquina (AM) surgem como alternativas promissoras por sua capacidade de identificar padrões complexos entre variáveis, sem a necessidade de assumir relações funcionais previamente definidas (Brink et al., 2016). Esta

característica é muito pertinente para a incorporação de variáveis de sensoriamento remoto a modelos de crescimento e produção. A relação entre variáveis espectrais e atributos dendrométricos nem sempre é linear, demandando abordagens mais flexíveis (Santos et al., 2023). Portanto, a integração dessas técnicas avançadas com variáveis de sensoriamento remoto representa uma fronteira de inovação na modelagem florestal.

Estudos recentes têm demonstrado que a incorporação de variáveis espectrais em modelos de crescimento tradicionais pode resultar em melhorias significativas na acurácia das estimativas. Da mesma forma, a aplicação de algoritmos como Florestas Aleatórias, Máquina de Vetores Suporte e Redes Neurais Artificiais, combinados com parâmetros do modelo de Clutter e variáveis espectrais, tem mostrado resultados promissores na melhoria da acurácia das estimativas de produção florestal (Araujo et al., 2023).

Diante desse cenário, este estudo parte da hipótese de que a incorporação de variáveis espectrais, representativas da condição fisiológica e da área foliar das árvores, pode aprimorar significativamente a acurácia da prognose volumétrica em relação aos modelos convencionais e aos baseados exclusivamente em aprendizado de máquina. Dois são os objetivos principais deste trabalho: a) avaliar a incorporação de variáveis espectrais na estrutura tradicional do modelo de Clutter e b) desenvolver e avaliar um modelo que integre as variáveis preditoras do modelo de Clutter com informações derivadas de sensoriamento remoto utilizando técnicas de aprendizado de máquina para capturar relações não-lineares entre as variáveis.

## **REFERENCIAL TEÓRICO**

### **1.1 Crescimento de florestas plantadas**

O crescimento florestal pode ser definido como o incremento progressivo de uma árvore ou povoamento em dimensões como altura, diâmetro, área basal e volume ao longo do tempo (Pretzsch, 2009). Trata-se de um processo dinâmico que resulta da interação complexa entre fatores genéticos e ambientais, mediados por processos fisiológicos fundamentais, como fotossíntese, respiração, absorção de água e nutrientes, e alocação de carbono (Landsberg & Sands, 2011).

A base do crescimento vegetal está na fotossíntese, processo pelo qual as plantas utilizam energia solar para converter dióxido de carbono ( $\text{CO}_2$ ) e água em carboidratos, liberando oxigênio como subproduto. A produtividade primária bruta (PPB) representa a taxa

total de assimilação de carbono pela vegetação, enquanto a produtividade primária líquida (PPL) reflete a quantidade de carbono efetivamente incorporada na biomassa após descontar os gastos com respiração (Larcher, 2003). No contexto florestal, a PPL determina o potencial de acúmulo de biomassa lenhosa, que é o principal objetivo econômico das florestas plantadas para produção de celulose.

O crescimento florestal é influenciado por uma ampla gama de fatores bióticos e abióticos. Entre os fatores abióticos, destacam-se a disponibilidade de luz, água e nutrientes, além das condições de temperatura e características do solo. A disponibilidade de luz afeta diretamente a taxa fotossintética, sendo particularmente relevante nas fases iniciais de desenvolvimento das árvores e na dinâmica de competição em povoamentos florestais (Silva et al., 2012). A água constitui um fator frequentemente limitante, influenciando não apenas a atividade fotossintética, mas também a absorção de nutrientes e a expansão celular. Segundo Stape et al. (2010), a disponibilidade hídrica pode explicar até 75% da variação na produtividade de plantações de eucalipto em diferentes regiões do Brasil. A nutrição mineral representa outro fator crítico para o crescimento florestal. Macronutrientes como nitrogênio, fósforo e potássio desempenham papéis fundamentais em processos metabólicos, estruturais e regulatórios nas plantas. Gonçalves et al. (2013) demonstraram que práticas adequadas de fertilização podem aumentar a produtividade de povoamentos de eucalipto em até 30%, dependendo das condições edáficas e climáticas locais.

Entre os fatores bióticos que afetam o crescimento florestal, destacam-se a presença de pragas, doenças e plantas invasoras, além da competição intraespecífica. Em plantios comerciais de eucalipto no Brasil, pragas como o percevejo-bronzeado (*Thaumastocoris peregrinus*) e doenças como a ferrugem (*Puccinia psidii*) podem causar reduções significativas na produtividade se não adequadamente manejadas (Alfenas et al., 2004).

O acúmulo de biomassa ao longo do tempo em um povoamento florestal segue um padrão característico, geralmente representado por uma curva sigmóide. Nas fases iniciais, o crescimento é lento devido à pequena área foliar disponível para interceptação da radiação solar. Com o desenvolvimento da copa e expansão do sistema radicular, o crescimento se acelera, atingindo seu ponto máximo quando o povoamento estabelece cobertura completa do solo. Posteriormente, observa-se uma desaceleração gradual do crescimento à medida que aumentam as demandas respiratórias e a competição por recursos (Burkhart & Tomé, 2012).

No contexto das florestas plantadas, o crescimento é comumente quantificado por meio de três índices principais: Incremento Corrente Anual (ICA), que representa o crescimento ocorrido apenas no último ano; Incremento Periódico Anual (IPA), que corresponde ao crescimento médio anual durante um período específico; e Incremento Médio Anual (IMA), que representa o crescimento médio anual desde o plantio até a idade atual (Campos & Leite, 2006). A culminação do IMA, ponto em que este se iguala ao ICA, é frequentemente utilizada como referência para determinar a idade ótima de desbaste em povoamentos equiâneos.

A produção florestal, por sua vez, pode ser entendida como o resultado acumulado do crescimento ao longo do tempo, subtraídas as perdas por mortalidade. Conforme explica (Campos & Leite, 2006), a produção em um povoamento florestal compreende cinco componentes principais: produção total, que representa o volume acumulado de todas as árvores (vivas e mortas) até determinada idade; produção corrente, correspondente ao volume das árvores vivas na idade atual; mortalidade, que representa o número ou volume de árvores que não sobreviveram ao longo do tempo; ingresso, referente às árvores que passaram a ser mensuráveis em inventários sucessivos; e colheita, que corresponde ao volume efetivamente removido durante operações de desbaste ou corte final.

O eucalipto, espécie dominante nas florestas plantadas brasileiras, se caracteriza por um crescimento excepcionalmente rápido em comparação com outras espécies arbóreas. Em condições favoráveis, plantios clonais de eucalipto no Brasil podem atingir incrementos médios anuais superiores a 50 m<sup>3</sup>/ha/ano, posicionando o país entre os líderes mundiais em produtividade florestal (IBÁ, 2022). Esse desempenho resulta de décadas de melhoramento genético, associado ao desenvolvimento de técnicas silviculturais apropriadas às condições tropicais.

As mudanças climáticas representam um desafio adicional para a compreensão e previsão do crescimento florestal. Segundo Laclau et al. (2013), o eucalipto pode beneficiar-se da elevada concentração atmosférica de CO<sub>2</sub>; porém, é vulnerável a extremos climáticos como secas prolongadas e ondas de calor. Os impactos das alterações climáticas sobre essas florestas variam conforme a região e dependem do equilíbrio entre efeitos potencialmente positivos (maior concentração de CO<sub>2</sub>) e negativos (aumento das temperaturas e alterações no regime hídrico).

Estudos conduzidos por Sanquetta et al. (2018) analisaram a dinâmica do volume, biomassa, área de superfície e carbono das florestas plantadas no Brasil ao longo de 26 anos

(1990–2016). Os resultados evidenciaram um aumento significativo na área de florestas plantadas no país, acompanhado por um crescimento expressivo nos estoques de biomassa e carbono, com um acréscimo de aproximadamente 170%. Esse aumento representa a remoção de cerca de 5 bilhões de toneladas de CO<sub>2</sub> da atmosfera, demonstrando a relevância das florestas plantadas para a mitigação das mudanças climáticas.

## **1.2 Modelagem do crescimento e da produção florestal**

A modelagem do crescimento e da produção florestal constitui uma disciplina fundamental para o planejamento e manejo sustentável de recursos florestais. Burkhart & Tomé (2012) definem modelos florestais como abstrações matemáticas que descrevem sistemas florestais, permitindo simular seu comportamento sob diferentes condições e cenários de manejo. Esses modelos evoluíram significativamente ao longo do tempo, refletindo tanto o avanço do conhecimento científico quanto o desenvolvimento de ferramentas computacionais mais sofisticadas.

Historicamente, a modelagem florestal tem suas raízes nas tabelas de produção desenvolvidas na Europa Central durante os séculos XVIII e XIX. Essas tabelas, baseadas em observações empíricas, forneciam estimativas de produção para povoamentos homogêneos sob regimes de manejo específicos. Com o avanço da estatística e da computação no século XX, surgiram modelos mais flexíveis e dinâmicos, capazes de incorporar variabilidade espacial e temporal (Pretzsch, 2009).

Os modelos de crescimento e produção florestal podem ser classificados em três categorias principais, de acordo com o nível de detalhe em que o sistema florestal é representado, conforme estabelecem Campos & Leite (2006): modelos de povoamento total, modelos de distribuição diamétrica e modelos de árvores individuais. Os modelos de povoamento total, como o próprio nome sugere, tratam o povoamento como unidade básica de modelagem, utilizando variáveis agregadas como idade, área basal, volume por hectare e índice de sítio. Esses modelos são geralmente mais simples e requerem menos dados para calibração, sendo amplamente utilizados no planejamento estratégico e tático do manejo florestal. Os modelos de distribuição diamétrica, por sua vez, incorporam informações sobre a estrutura do povoamento, representando a frequência de árvores em classes de diâmetro. Essa abordagem permite estimativas mais detalhadas da produção por classe de tamanho, informação essencial para a valoração de produtos florestais com diferentes especificações dimensionais. Já os

modelos de árvores individuais representam o nível mais detalhado de modelagem, simulando o crescimento de cada árvore em função de suas características individuais e de suas interações com árvores vizinhas e com o ambiente. Esses modelos são particularmente úteis para povoamentos heterogêneos e sistemas florestais complexos, embora demandem maior quantidade de dados e maior capacidade computacional.

Dentro da categoria de modelos de povoamento total, destacam-se os modelos de densidade variável, que consideram explicitamente a influência da densidade do povoamento sobre o crescimento. O modelo de Clutter (Clutter, 1963) representa um dos exemplos mais influentes dessa abordagem, sendo amplamente adotado em empresas florestais no Brasil e em outras partes do mundo. Este modelo é um sistema de equações simultâneas que projeta a área basal e o volume do povoamento para idades futuras, com base em informações sobre a área basal atual, a idade e a qualidade do sítio. Matematicamente, o modelo pode ser expresso da seguinte forma:

$$\ln G_2 = \ln G_1 \left( \frac{I_1}{I_2} \right) + \alpha_1 \left( 1 - \frac{I_1}{I_2} \right) + \alpha_2 \left( 1 - \frac{I_1}{I_2} \right) * S$$

$$\ln V_2 = \beta_0 + \beta_1 \left( \frac{I_1}{I_2} \right) + \beta_2 S + \beta_3 \ln G_2$$

Onde:  $\ln G_1$  é a área basal de partida ou atual,  $I_1$  é a idade em meses de partida ou atual,  $I_2$  é a idade em meses, de interesse para prognose,  $S$  é o índice de sítio,  $\ln G_2$  é a área basal na idade de interesse para prognose,  $\ln V_2$  é o volume do talhão na idade de interesse para prognose,  $\alpha_1$ ,  $\alpha_2$ ,  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$  são coeficientes a serem estimados.

O modelo de Clutter destaca-se por sua versatilidade e consistência matemática. A versatilidade deve-se ao fato de que a variável resposta (volume,  $\text{m}^3 \text{ ha}^{-1}$ ) é independente da distribuição de tamanhos das árvores, permitindo estimativas para diversos cenários de manejo. A consistência matemática, por sua vez, manifesta-se na capacidade do modelo de realizar projeções anuais sucessivas ou estimativas diretas para qualquer período futuro, mantendo a compatibilidade entre as estimativas (Campos & Leite, 2006). Entretanto, modelo tem limitações quanto a inferência sobre efeito de atributos ecofisiológicos, como a fotossíntese, alocação de carbono e resposta a fatores ambientais. Essa limitação torna-se particularmente relevante em um contexto de mudanças climáticas, onde as condições ambientais podem desviar-se significativamente daquelas sob as quais o modelo foi calibrado (Landsberg & Sands, 2011). Outra limitação importante refere-se à homogeneidade espacial assumida pelo modelo. Em condições reais,

fatores como topografia, características do solo e regimes microclimáticos podem variar consideravelmente dentro de um mesmo talhão, influenciando o crescimento de maneira heterogênea. Pereira et al. (2016) demonstraram que a inclusão de um componente espacial no modelo de Clutter pode melhorar significativamente sua capacidade preditiva, reduzindo os erros de estimativa em até 6,57%.

Diante dessas limitações, pesquisadores têm buscado alternativas para aprimorar o modelo de Clutter e outros modelos empíricos tradicionais. Uma abordagem promissora envolve a incorporação de informações derivadas de sensoriamento remoto, que podem fornecer indicadores da condição fisiológica e do vigor vegetativo dos povoamentos florestais.

### **1.3 Sensoriamento remoto para predição de estoque**

O sensoriamento remoto tem revolucionado a forma como monitoramos e gerenciamos recursos florestais, oferecendo uma visão sinóptica e temporal que seria impossível de obter exclusivamente por métodos tradicionais de campo. White et al. (2016) definem sensoriamento remoto como a ciência e a arte de obter informações sobre objetos através da análise de dados adquiridos por dispositivos que não estão em contato direto com esses objetos. No contexto florestal, essa tecnologia tem sido aplicada com sucesso em diversas áreas, incluindo mapeamento de cobertura vegetal, detecção de mudanças, monitoramento de distúrbios e, mais recentemente, na estimativa de parâmetros biofísicos e estruturais dos povoamentos.

O princípio fundamental do sensoriamento remoto aplicado à vegetação baseia-se nas interações da radiação eletromagnética com os componentes do dossel florestal. As folhas, por exemplo, absorvem fortemente a radiação nas regiões do azul (400 nm a 500 nm) e do vermelho (620 nm a 750 nm) do espectro visível (usadas na fotossíntese), e refletem moderadamente na região do verde (500 nm a 570 nm) e apresentam alta reflectância no infravermelho próximo (700 nm a 1400 nm). Esse comportamento espectral característico, conhecido como "assinatura espectral", fornece a base para o desenvolvimento de índices de vegetação e algoritmos de estimação de parâmetros biofísicos (Ponzoni et al., 2015).

Entre os diversos sensores orbitais disponíveis para aplicações florestais, as missões Landsat e Sentinel destacam-se por seu equilíbrio entre resolução espacial, temporal e espectral,

além da disponibilidade histórica de dados e acesso gratuito. O programa Landsat, em operação contínua desde 1972, representa a mais longa série temporal de observações da Terra a partir do espaço, fornecendo um arquivo inestimável para estudos de mudanças florestais. O Landsat 8, lançado em 2013, oferece imagens multiespectrais com resolução espacial de 30 metros e revisita a mesma área aproximadamente a cada 16 dias (Roy et al., 2014).

Por sua vez, a missão Sentinel-2, parte do programa Copernicus da Agência Espacial Europeia, introduziu melhorias significativas em termos de resolução espacial (10-20 metros), espectral (13 bandas) e temporal (5 dias com dois satélites), ampliando as possibilidades para o monitoramento florestal de alta frequência (Drusch et al., 2012). Além desses sensores ópticos passivos, tecnologias como o LiDAR (*Light Detection and Ranging*) e radar (*Radio Detection and Ranging*) têm emergido como ferramentas complementares, especialmente valiosas em regiões com cobertura frequente de nuvens.

A aplicação do sensoriamento remoto na predição de estoque florestal pode seguir diferentes abordagens metodológicas. De acordo com Lu et al. (2016), essas abordagens podem ser categorizadas em: métodos baseados em correlação espectral, métodos baseados em índices de vegetação, métodos de mistura espectral, métodos baseados em textura e métodos que integram múltiplas fontes de dados.

Os índices de vegetação (IV) representam uma das abordagens mais amplamente utilizadas para a estimação de parâmetros florestais a partir de dados de sensoriamento remoto. Esses índices são transformações matemáticas das reflectâncias em diferentes bandas espectrais, projetadas para realçar o sinal da vegetação e minimizar influências indesejadas do solo, atmosfera e geometria de iluminação/observação. Entre os diversos IVs desenvolvidos nas últimas décadas, o Índice de Vegetação por Diferença Normalizada (NDVI), proposto por Rouse Jr et al. (1973), permanece como um dos mais utilizados devido à sua simplicidade e eficácia em capturar variações na biomassa verde, área foliar e atividade fotossintética.

A relação entre índices espectrais e parâmetros biofísicos florestais tem sido extensivamente investigada na literatura científica. Brandão et al. (2020) demonstraram correlações significativas entre o NDVI derivado de imagens Landsat e a biomassa aérea em plantações de eucalipto em diferentes estágios de crescimento. De forma similar, Viana et al. (2016) encontraram correlações moderadas a fortes entre diversos índices de vegetação e variáveis dendrométricas como altura, diâmetro e volume em povoamentos de *Eucalyptus urophylla*.

Miguel et al. (2015) avaliaram o potencial de índices como SR (Simple Ratio) e SAVI (Soil-Adjusted Vegetation Index) para a estimativa de volume de madeira e biomassa em áreas de Cerrado, obtendo erros-padrão de aproximadamente 12,34%. Os autores concluíram que a integração de múltiplos índices pode melhorar significativamente a precisão das estimativas em comparação com o uso de índices isolados.

Outra variável biofísica de grande relevância para a produtividade florestal é o Índice de Área Foliar (IAF), que representa a área total de folhas por unidade de área de terreno. O IAF está diretamente relacionado à interceptação da radiação fotossinteticamente ativa (PAR) e, conseqüentemente, à fotossíntese do dossel e à produção primária (Gower et al., 1999). Xu et al. (2020) identificaram quatro abordagens principais para a estimação do IAF por sensoriamento remoto: modelos baseados em reflectância, modelos baseados em índices de vegetação, modelos derivados de dados de LiDAR e modelos de aprendizado de máquina.

Embora os índices de vegetação tradicionais apresentem correlações significativas com parâmetros florestais, essas relações frequentemente não são lineares e podem ser afetadas por fatores como saturação do sensor em áreas de alta biomassa, influência do substrato e efeitos atmosféricos. Além disso, a resposta espectral da vegetação é influenciada por múltiplos fatores, incluindo estrutura do dossel, propriedades ópticas das folhas, geometria de iluminação e observação, e condições atmosféricas (Ponzoni et al., 2015).

Nesse sentido, embora os índices de vegetação tradicionais forneçam informações relevantes sobre a condição da vegetação, suas limitações, como saturação do sensor em áreas de alta biomassa e sensibilidade a fatores externos, reforçam a necessidade de incorporar métricas complementares. Nesse contexto, os índices de textura derivados de sensores ópticos surgem como uma alternativa promissora, pois permitem captar padrões espaciais de variação na reflectância que não são identificados pelos índices espectrais convencionais. Dessa forma, a integração de informações espectrais e texturais possibilita uma caracterização mais robusta da estrutura florestal e de seus atributos, contribuindo para análises mais precisas em estudos de crescimento e produção.

### ***1.3.1 ÍNDICES DE TEXTURA DERIVADOS DE SENSORES ÓPTICOS***

Os índices de textura representam uma dimensão complementar e fundamental na análise de imagens de sensoriamento remoto para aplicações florestais, capturando informações sobre a variabilidade espacial dos valores de pixels que não são detectadas pelos índices

espectrais tradicionais. A textura de uma imagem refere-se à variação sistemática dos valores de brilho entre pixels vizinhos, fornecendo informações valiosas sobre a estrutura espacial dos objetos na superfície terrestre. No contexto florestal, estes índices podem revelar características importantes da estrutura do dossel, densidade do povoamento, heterogeneidade da vegetação e, conseqüentemente, relacionar-se com a biomassa e volume de madeira (Haralick & Shanmugam, 1973).

A matriz de co-ocorrência de níveis de cinza (Gray Level Co-occurrence Matrix - GLCM), desenvolvida por Haralick & Shanmugam (1973), constitui o método mais amplamente utilizado para extração de métricas de textura em sensoriamento remoto. Esta matriz descreve a frequência de ocorrência de pares de valores de pixels separados por uma distância específica em uma direção particular, permitindo a quantificação de propriedades texturais da imagem. A partir da GLCM, podem ser derivadas diversas métricas, sendo as mais utilizadas em estudos florestais: média (mean), variância (variance), homogeneidade (homogeneity), contraste (contrast), dissimilaridade (dissimilarity), entropia (entropy), segundo momento angular (angular second moment) e correlação (correlation) (Kelsey & Neff, 2014).

Estudos têm demonstrado a superioridade dos índices de textura em relação aos índices espectrais tradicionais para estimativa de biomassa florestal. Kelsey & Neff (2014) investigaram oito métricas GLCM baseadas no Landsat TM em quatro tamanhos de janela ( $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$ ) para estimativa de biomassa em florestas temperadas do Colorado, Estados Unidos. Os resultados indicaram que modelos incluindo variáveis de textura apresentaram correlação de  $r=0,86$  com a biomassa observada, superando significativamente modelos baseados apenas em variáveis físicas e espectrais. As métricas de entropia, média e correlação calculadas da banda 2 do Landsat 7 em janela  $3 \times 3$  foram as mais eficazes, demonstrando que a textura é particularmente sensível a mudanças na biomassa após distúrbios florestais.

Estudos recentes têm evidenciado que a combinação de índices espectrais e de textura pode melhorar significativamente a acurácia das estimativas de volume. Velasco Pereira et al. (2023) utilizaram modelos de sinergia de sensores ALOS PALSAR-Sentinel 1-Landsat 8 para mudanças temporais na biomassa de florestas mediterrâneas de pinheiro, calculando variáveis de textura com GLCM em janela  $5 \times 5$  pixels.

Pesquisas nacionais recentes têm incorporado índices de textura em estudos de monitoramento ambiental. Rosa (2021) utilizou índices de textura derivados de imagens Landsat para análise temporal da cobertura florestal na Mata Atlântica, integrando informações

temporais, sazonais e de textura para classificação de uso e cobertura da terra. Almeida (2022) aplicou índices de textura baseados na GLCM para estimativa de biomassa em florestas do Amapá, concluindo que os três índices de textura principais forneceram previsões superiores aos métodos tradicionais.

As métricas derivadas da GLCM fornecem informações complementares sobre diferentes aspectos da textura da imagem, cada uma com interpretações específicas no contexto florestal, conforme a Tabela 1. A compreensão destas métricas é fundamental para seleção apropriada de variáveis de textura em modelos de estimativa de biomassa florestal.

*Tabela 1 – Relação do Índices de Textura com o eucalipto*

<b>Índice de Textura</b>	<b>Significado principal</b>	<b>Valor alto – Interpretação</b>	<b>Valor baixo - Interpretação</b>
Energia	Mede a uniformidade ou repetição dos tons de cinza.	Superfície homogênea; povoamento denso e uniforme; dossel fechado.	Alta variação tonal; clareiras, bordas, talhões jovens ou com falhas.
Entropia	Mede a desordem ou complexidade da textura.	Alta heterogeneidade; mistura de copas, variação de altura, floresta natural.	Estrutura simples; dossel contínuo e regular; povoamento homogêneo.
Correlação	Mede o grau de correlação entre tons vizinhos.	Textura contínua; copas regulares; vigor uniforme.	Variação abrupta; descontinuidade no dossel; falhas ou reboleiras.
Momento de Diferença Inversa (IDM)/Homogeneidade	Mede a homogeneidade local.	Textura suave; tons semelhantes; floresta uniforme e bem desenvolvida.	Textura rugosa; grandes diferenças entre pixels; áreas em regeneração.
Inércia/Contraste	Mede o contraste local entre tons de cinza.	Forte variação estrutural; topografia acidentada, dossel irregular, clareiras.	Superfície lisa e regular; estrutura uniforme do dossel.
Sombra do Cluster (CS)	Mede a assimetria tonal (predomínio de claro ou escuro).	Presença de áreas sombreadas intensas ou superfícies muito iluminadas; desequilíbrio tonal.	Distribuição tonal equilibrada; iluminação homogênea.

Proeminência do Cluster (CP)	Mede a curtose da distribuição tonal (acentuação do contraste).	Presença de bordas, estradas, clareiras, transições abruptas.	Textura estável, sem variações bruscas de tons.
Correlação de Haralick (HC)	Versão robusta da correlação entre tons vizinhos.	Textura suave e contínua, com relação linear estável.	Padrões irregulares e desordenados, ruído tonal ou mistura espectral.

Fonte: (Li et al., 2012)

A incorporação de índices espectrais e texturais amplia significativamente o potencial do sensoriamento remoto para estimativas florestais, mas a complexidade das relações entre essas variáveis e os parâmetros dendrométricos demanda técnicas analíticas mais robustas. Nesse cenário, os modelos de aprendizado de máquina (AM) têm se destacado como ferramentas capazes de lidar com grandes volumes de dados multivariados e de capturar relações não lineares entre variáveis preditoras e resposta. Dessa forma, a integração entre métricas derivadas do sensoriamento remoto e algoritmos de aprendizado de máquina representa uma abordagem promissora para aprimorar a acurácia e a generalização das estimativas de biomassa, volume e outros atributos florestais, constituindo-se em uma tendência crescente nas pesquisas da área.

#### 1.4 Modelos de Aprendizado de Máquina (AM)

O aprendizado de máquina (AM) representa um ramo da inteligência artificial que se concentra no desenvolvimento de algoritmos capazes de aprender padrões a partir de dados e realizar previsões ou decisões sem serem explicitamente programados para executar tarefas específicas. Diferentemente dos modelos estatísticos tradicionais, que geralmente requerem pressupostos sobre a distribuição dos dados e a forma funcional das relações entre variáveis, os algoritmos de AM podem adaptar-se automaticamente à estrutura inerente aos dados, identificando relações complexas e não-lineares (Hastie et al., 2009).

No contexto da modelagem florestal, a AM oferecem vantagens significativas em comparação com abordagens convencionais baseadas em regressão. Enquanto os modelos de regressão tradicionais, como o modelo de Clutter, requerem uma forma funcional predefinida (geralmente baseada em conhecimento teórico ou empírico), a AM podem estabelecer relações complexas entre variáveis preditoras e respostas, potencialmente capturando interações e não-linearidades que seriam difíceis de especificar a priori (Joibary, 2013). Outra vantagem

importante da AM é sua capacidade de lidar com grandes volumes de dados multidimensionais, característica particularmente relevante em aplicações que integram dados de sensoriamento remoto com medições convencionais de campo. Além disso, muitos algoritmos de AM são robustos a problemas como multicolinearidade e são capazes de lidar com dados ausentes ou ruidosos, situações comuns em conjuntos de dados florestais (Breiman, 2001).

Entre os diversos algoritmos de AM aplicados à modelagem florestal, destacam-se o Florestas Aleatórias (RF), Máquina de Vetores Suporte (SVM) e Aumento De Gradiente Extremo (XGBoost). Cada um desses algoritmos possui características distintas que os tornam adequados para diferentes tipos de problemas e conjuntos de dados.

#### ***1.4.1 FLORESTA ALEATÓRIA (RF)***

Floresta Aleatória (*Random Forest*) é um método introduzido por Breiman (2001) que combina as previsões de múltiplas árvores de decisão para produzir um resultado final mais preciso e robusto. O algoritmo baseia-se em dois princípios fundamentais: *bagging* (*bootstrap aggregating*) e seleção aleatória de características. No processo de *bagging*, cada árvore é treinada com uma amostra *bootstrap* do conjunto de dados original, ou seja, uma amostra aleatória com reposição que contém aproximadamente 63% dos dados originais. Os dados não selecionados para treinar uma árvore específica, conhecidos como "*out-of-bag*" (OOB), são utilizados para avaliar o desempenho dessa árvore, proporcionando uma validação interna durante o treinamento. O RF introduz aleatoriedade adicional na construção de cada árvore através da seleção aleatória de características. Em cada nó da árvore, apenas um subconjunto aleatório das variáveis preditoras está disponível para divisão, forçando diversidade entre as árvores do conjunto (Figura 1). Essa característica é particularmente valiosa quando existem fortes preditores no conjunto de dados, pois evita que todas as árvores sejam dominadas por esses preditores (Breiman, 2001).

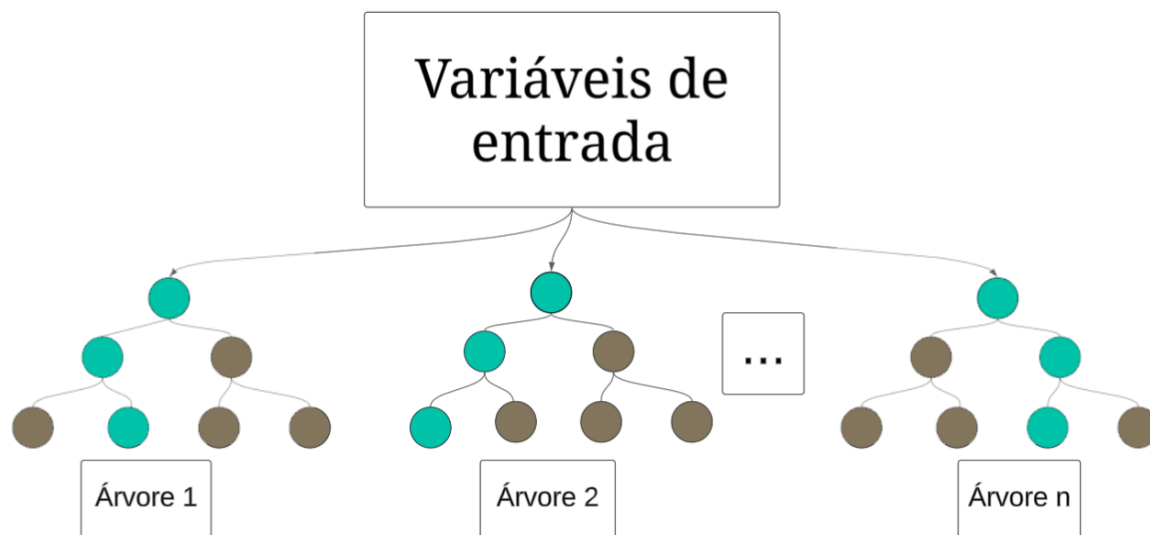


Figura 1 – Exemplo de uma floresta aleatória

Fonte: Adaptado de Araujo et al. (2023)

Para problemas de regressão, como a predição de volume ou biomassa florestal, o resultado final do RF é obtido pela média das previsões individuais de todas as árvores no conjunto. Matematicamente, a previsão para uma nova observação  $x$  pode ser expressa como:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n f_i(x)$$

Onde  $n$  é o número total de árvores no conjunto e  $f_i(x)$  é a previsão da  $i$ -ésima árvore para a observação  $x$ .

O RF oferece diversas vantagens para aplicações florestais. Primeiro, é relativamente robusto ao *overfitting*, especialmente quando o número de árvores é grande. Segundo, fornece medidas internas de importância das variáveis, permitindo identificar quais preditores contribuem mais significativamente para o modelo. Terceiro, pode lidar eficientemente com interações complexas entre variáveis e relações não-lineares, características comuns em dados ecológicos. Por fim, requer relativamente pouca preparação de dados e ajuste de hiperparâmetros em comparação com outros algoritmos avançados (Cosenza et al., 2021, 2025; Cutler et al., 2007).

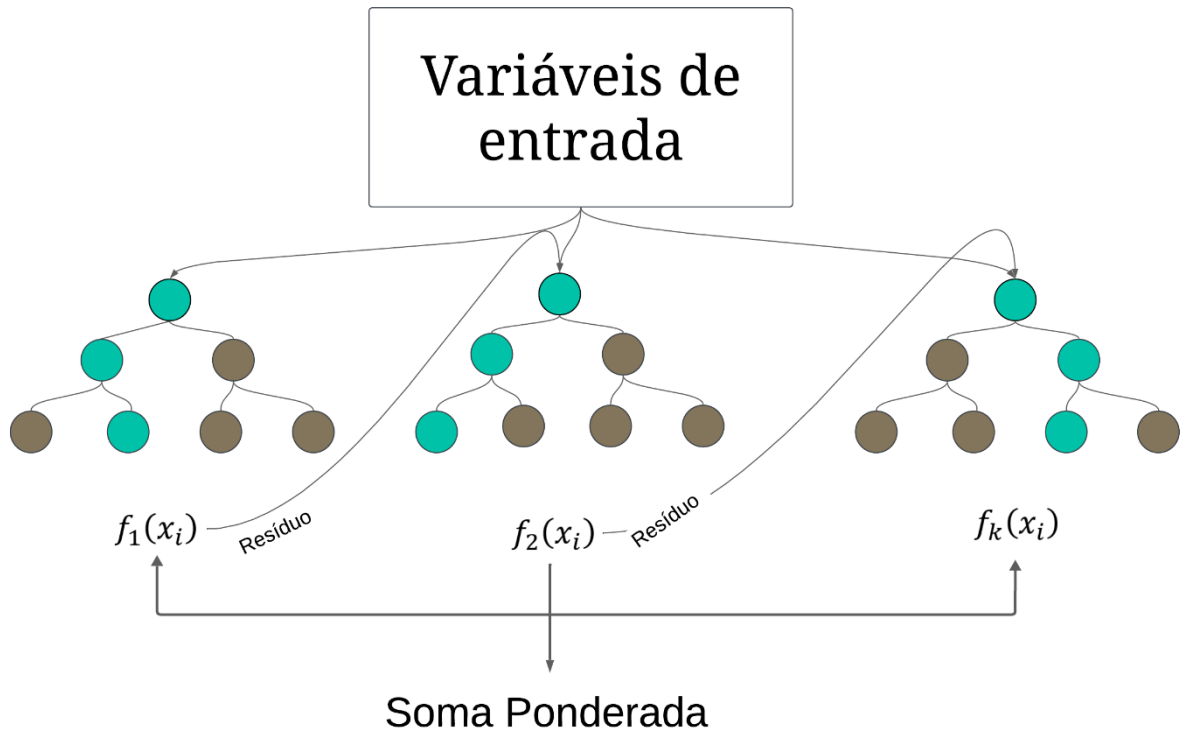
No campo da modelagem florestal, o RF tem sido amplamente aplicado para estimativas de volume, biomassa, carbono e outros parâmetros estruturais. Lima et al. (2022) utilizaram RF para estimar o volume de madeira em plantações de eucalipto a partir de variáveis dendrométricas e índices espectrais derivados de imagens Landsat, obtendo resultados

superiores aos modelos de regressão convencionais. De forma similar, (Gómez et al., 2012) aplicaram RF para mapear biomassa florestal usando dados LiDAR e multiespectrais, destacando a capacidade do algoritmo em capturar relações complexas entre a estrutura tridimensional do dossel e a biomassa aérea.

#### ***1.4.2 AUMENTO DE GRADIENTE EXTREMO (XGBOOST)***

O Aumento de gradiente extremo (*Extreme Gradient Boosting*) é um algoritmo de aprendizado de máquina baseado em árvores de decisão que implementa o framework de *gradient boosting* com várias otimizações para eficiência computacional e desempenho preditivo. Desenvolvido por (Chen & Guestrin, 2016), o XGBoost rapidamente se tornou uma das ferramentas mais populares em competições de ciência de dados e aplicações práticas devido à sua eficiência computacional e precisão.

O princípio fundamental do *gradient boosting* é construir um modelo forte através da combinação sequencial de modelos fracos (geralmente árvores de decisão rasas), onde cada novo modelo é treinado para corrigir os erros dos modelos anteriores (Figura 3). Diferentemente do *Random Forest*, onde as árvores são treinadas independentemente e combinadas por média simples, no *boosting* cada árvore é adicionada sequencialmente para minimizar uma função de perda.



*Figura 3 – Exemplo de um XGBoost*

Fonte: Adaptado de (Zou et al., 2022a)

Matematicamente, o modelo XGBoost pode ser representado como:

$$\hat{y}_i = \sum_{j=1}^n f_j(x_i)$$

Onde  $\hat{y}_i$  é a previsão para a observação  $i$ ,  $n$  é o número total de árvores, e  $f_j$  representa a  $j$ -ésima árvore de decisão.

O objetivo de treinamento é minimizar a seguinte função regularizada:

$$L = \sum_i l(y_i, \hat{y}_i) + \sum_i \Omega(f_j)$$

Onde  $l$  é a função de perda que mede a diferença entre a previsão  $\hat{y}_i$  e o valor real  $y_i$ , e  $\Omega(f_j)$  é um termo de regularização que penaliza a complexidade do modelo para evitar overfitting (Chen & Guestrin, 2016).

O XGBoost introduz várias inovações em relação ao *gradient boosting* tradicional. Primeiro, utiliza uma formulação regularizada para controlar a complexidade do modelo, incluindo termos de penalidade para o número de folhas nas árvores e para os pesos das folhas. Segundo, implementa um algoritmo aproximado de divisão de árvores que permite processamento eficiente de grandes conjuntos de dados. Terceiro, incorpora técnicas para lidar eficientemente com dados esparsos e valores ausentes. Por fim, utiliza várias otimizações computacionais, como paralelização, *cache-aware access* e computação *out-of-core* para conjuntos de dados que não cabem na memória (Chen & Guestrin, 2016).

No contexto da modelagem florestal, o XGBoost tem demonstrado excelente desempenho em diversas aplicações. Chen et al. (2019) utilizaram XGBoost para estimar o volume de madeira em florestas mistas a partir de variáveis derivadas de LiDAR e imagens multiespectrais, obtendo resultados superiores a métodos tradicionais de regressão. Zou et al. (2022) aplicaram XGBoost para prever a produtividade florestal considerando variáveis climáticas e edáficas, destacando a capacidade do algoritmo em capturar interações complexas entre fatores ambientais e crescimento.

As principais vantagens do XGBoost para aplicações florestais incluem seu desempenho preditivo geralmente superior em comparação com outros algoritmos, capacidade de lidar com diferentes tipos de variáveis (numéricas, categóricas), robustez a outliers através de funções de perda personalizáveis, e interpretabilidade através de métricas de importância de características. Além disso, o XGBoost oferece mecanismos eficientes para evitar *overfitting*, como regularização, amostragem de características e dados, e parada antecipada baseada em um conjunto de validação (Chen & Guestrin, 2016).

#### ***1.4.3 MÁQUINA DE VETORES DE SUPORTE (SVM)***

Máquina de Vetores de Suporte (*Support Vector Machine*) é um algoritmo de aprendizado supervisionado desenvolvido inicialmente para classificação por Vapnik (2013) e posteriormente estendido para problemas de regressão. O SVM baseia-se em princípios da teoria de aprendizado estatístico e busca encontrar um hiperplano ótimo que maximiza a margem entre diferentes classes (no caso de classificação) ou que melhor se ajusta aos dados dentro de uma margem de tolerância (no caso de regressão) conforme demonstrado na Figura 2.

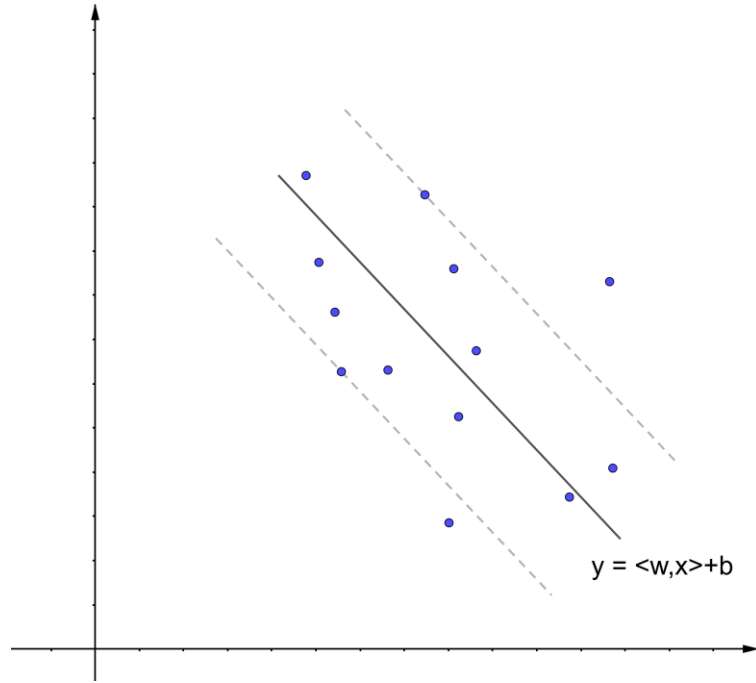


Figura 2 – Exemplo de um SVM

Fonte: Adaptado de Araujo et al. (2023)

Para problemas de regressão, a variante conhecida como *Support Vector Regression* (SVR) busca ajustar uma função que tenha no máximo  $\varepsilon$  de desvio dos valores alvo observados, enquanto mantém a função tão plana quanto possível. Matematicamente, para um conjunto de dados de treinamento  $\{(x_1, y_1), \dots, (x_n, y_n)\}$ , onde  $x$  representa o vetor de características e  $y$  o valor alvo, o SVR procura minimizar:

$$\text{minimize } \left(\frac{1}{2}\right) \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*)$$

$$\text{sujeito a: } y_i - f(x_i) \leq \varepsilon + \xi_i$$

$$f(x_i) - y_i \leq \varepsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \geq 0$$

Onde  $f(x) = \langle w, x \rangle + b$  é a função de regressão linear,  $\|w\|^2$  é o termo de regularização,  $C$  é o parâmetro de penalização que controla o equilíbrio entre a complexidade do modelo e o erro de treinamento, e  $\xi_i, \xi_i^*$  são variáveis de folga que permitem violações da margem  $\varepsilon$  (Smola & Schölkopf, 2004).

Uma das principais vantagens do SVM é sua capacidade de lidar com relações não-lineares através do chamado "truque do kernel". Essa técnica permite mapear implicitamente os dados para um espaço de características de maior dimensão, onde o problema pode se tornar linearmente separável ou mais facilmente modelável. Kernels comumente utilizados incluem o linear, polinomial, base radial (RBF) e sigmoide (Cortes & Vapnik, 1995).

No contexto florestal, o SVM tem sido aplicado com sucesso em diversas tarefas de estimação. Gonçalves (2010) utilizou SVM para estimar a biomassa aérea em florestas tropicais a partir de índices espectrais, obtendo resultados competitivos em comparação com outros algoritmos. Cordeiro (2020) avaliou diferentes configurações de SVM na modelagem do crescimento e produção de eucalipto, demonstrando desempenho estatisticamente satisfatório para implementação no manejo florestal.

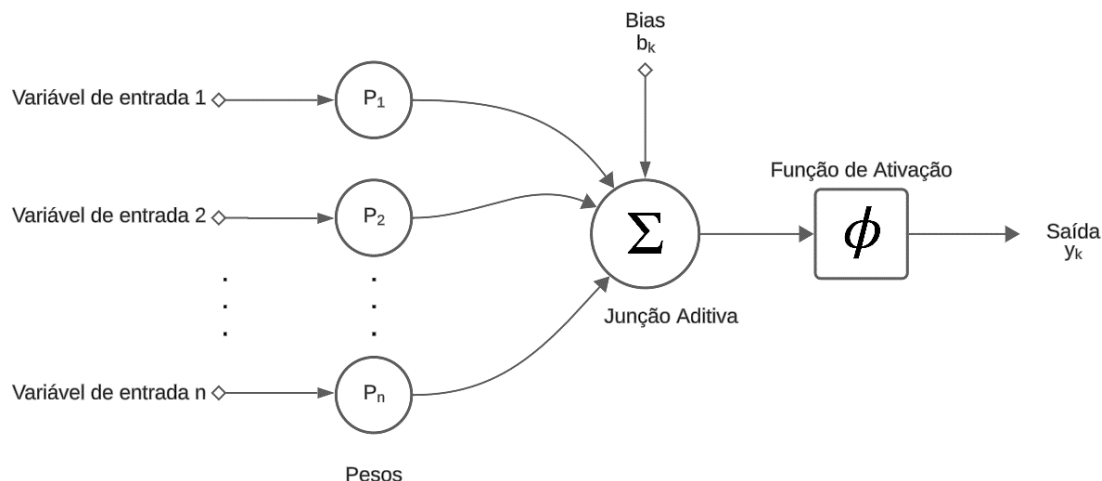
Entre as vantagens do SVM para aplicações florestais, destacam-se sua eficácia em conjuntos de dados de alta dimensionalidade (como os derivados de sensoriamento remoto hiperespectrais), robustez a ruídos e capacidade de generalização mesmo com amostras de treinamento relativamente pequenas. Além disso, o SVM tende a evitar o *overfitting* através do princípio da minimização do risco estrutural, equilibrando a complexidade do modelo com seu desempenho nos dados de treinamento (Smola & Schölkopf, 2004).

#### ***1.4.4 REDES NEURAIS ARTIFICIAIS (RNA)***

As Redes Neurais Artificiais (RNA) consistem em modelos projetados para imitar o processamento da informação realizado pelo cérebro humano, utilizando uma rede composta por unidades computacionais interconectadas, denominadas neurônios artificiais. Esses neurônios são organizados em camadas e possuem a capacidade de aprender padrões complexos a partir de dados de entrada, ajustando seus parâmetros internos de forma iterativa durante o processo de treinamento, com o objetivo de generalizar o conhecimento adquirido para novas situações (Haykin, 2001).

Cada neurônio artificial representa uma unidade básica de processamento dentro da rede e é composto por diversos elementos funcionais. As sinapses são representadas pelos pesos sinápticos, que determinam a importância de cada sinal de entrada. O somador é responsável por calcular a combinação linear ponderada dos sinais recebidos. Em seguida, a função de ativação é aplicada para restringir a amplitude da saída e introduzir não linearidade ao modelo, o que permite à RNA representar relações complexas entre as variáveis. O termo de bias atua

como um deslocamento da função de ativação, ajustando o limiar de resposta do neurônio. A saída de cada neurônio pode servir tanto como entrada para outros neurônios em camadas subsequentes quanto como resultado final da rede, conforme ilustrado na Figura 4 (Haykin, 2001)



*Figura 4 – Exemplo de um Neurônio*

Fonte: Adaptado de (Haykin, 2001)

As Redes Neurais Artificiais (RNAs) apresentam vantagens significativas em relação aos métodos tradicionais de modelagem, especialmente na capacidade de capturar relações não lineares complexas entre variáveis. Esse tipo de algoritmo é particularmente eficiente para descrever sistemas influenciados por múltiplos fatores interdependentes, como condições climáticas, características edáficas, densidade genética e variabilidade estrutural dos povoamentos florestais. As RNAs são capazes de identificar padrões ocultos nos dados que frequentemente não são detectados por modelos de regressão tradicionais (Bentes, 2018).

Além disso, as RNAs podem integrar variáveis de diferentes naturezas, combinando informações qualitativas, como genótipo, tipo de solo e espaçamento, com variáveis quantitativas, como diâmetro, altura e área basal. Outra característica relevante é a robustez frente a ruídos e falhas nos dados, bem como a capacidade de generalização, permitindo que o modelo apresente bom desempenho mesmo em situações não observadas durante o treinamento (Bentes, 2018).

No contexto florestal, diversos estudos têm evidenciado o potencial das RNAs em tarefas de predição e modelagem. Azevedo et al. (2020) empregaram diferentes estratégias de RNA para a modelagem de multivolumes do fuste de eucaliptos, observando que essas redes

apresentaram maior precisão e praticidade em comparação aos modelos de regressão tradicionais. De forma semelhante, Bayat et al. (2019) aplicaram RNAs para prever a sobrevivência e a mortalidade de árvores em florestas de faia oriental no norte do Irã, avaliando o desempenho dos modelos por meio das métricas RMSE e  $R^2$ . Os resultados demonstraram a superioridade das RNAs, reforçando sua eficácia como ferramenta preditiva na modelagem florestal. Niska et al. (2010) examinaram a capacidade técnica comuns de modelagem por RNA para predição de atributos florestais como o volume de fuste em nível de parcela experimental e de povoamento florestal. Esse estudo mostrou que as RNAs são métodos apropriados e precisos para a avaliação de atributos florestais, podendo ser utilizadas como alternativas para a regressão linear multivariada.

#### ***1.4.5 K-VIZINHOS PRÓXIMO***

O modelo de K-Vizinhos Mais Próximos (KNN) é um algoritmo baseado em medidas de similaridade entre observações. Fundamenta-se na premissa de que amostras com características semelhantes tendem a apresentar respostas também semelhantes. Segundo Cover & Hart (1967), que originalmente formularam o método, o princípio do KNN reside na proximidade entre vetores de características em um espaço multidimensional definido pelas variáveis explicativas.

O funcionamento do modelo consiste em identificar os  $k$  pontos mais próximos de uma observação, calculando a distância entre ela e os demais pontos do conjunto de dados. A métrica de distância mais comumente utilizada é a distância euclidiana, expressa pela Equação:

$$d(x_i, x_j) = \sqrt{\sum_{n=1}^p (x_{in} - x_{jn})^2}$$

Onde:  $x_i$  e  $x_j$  são os vetores de características das amostras;  $p$  é o número de variáveis. Outras métricas podem ser utilizadas, como as distâncias de Manhattan, Mahalanobis ou ponderadas, dependendo da natureza dos dados e do objetivo da modelagem.

Entre as principais vantagens do KNN destacam-se sua simplicidade conceitual e facilidade de implementação, além de não pressupor uma estrutura paramétrica fixa, o que o caracteriza como um método não paramétrico e flexível. Essa característica permite ao KNN adaptar-se a variações locais dos dados, capturando padrões complexos que outros modelos

podem não identificar. O algoritmo também apresenta boa interoperabilidade com diferentes tipos de variáveis e robustez em cenários com múltiplas fontes de dados. (Souza et al., 2019)

No contexto florestal, o uso do KNN tem se mostrado promissor. McRoberts et al. (2015) aplicaram o método para estimar a biomassa aérea média por unidade de área na Noruega, utilizando dados de inventário florestal e varredura a laser aerotransportada (ALS). Os resultados indicaram uma redução da dimensionalidade e um aumento de até 20% na variabilidade explicada no conjunto de referência. No Brasil, Amorim (2016) empregou o KNN combinado com dados Landsat-8 e variáveis topográficas para estimar a biomassa aérea de povoamentos de eucalipto, obtendo coeficientes de determinação ( $R^2$ ) superiores a 0,80, o que demonstra a eficiência do método para aplicações florestais de predição e mapeamento.

### **1.5 Aprendizado de máquina na modelagem florestal**

A integração de técnicas de aprendizado de máquina na modelagem florestal tem ganhado impulso significativo na última década, motivada tanto pelo aumento da disponibilidade de dados quanto pelos avanços nas capacidades computacionais. Diferentemente dos modelos empíricos tradicionais, que dependem de relações funcionais predefinidas, a AM oferece maior flexibilidade para capturar relações complexas e não-lineares entre variáveis preditoras e parâmetros florestais de interesse.

No contexto específico da modelagem de crescimento e produção florestal, os Modelos AM podem ser aplicados de diferentes formas: (i) como substitutos completos para modelos empíricos tradicionais, (ii) como componentes híbridos que complementam modelos existentes, ou (iii) como ferramentas auxiliares para a seleção de variáveis ou imputação de dados ausentes.

Binoti et al. (2014) demonstraram o potencial da RNA como alternativa aos modelos de crescimento e produção tradicionais. Comparando RNA com o modelo de Clutter para projeção da área basal e volume em povoamentos de eucalipto, os autores observaram que a RNA proporciona estimativas mais precisas e com distribuição mais homogênea dos resíduos. Costa Filho et al. (2019) corroboraram esses resultados, mostrando que a RNA pode capturar de forma mais eficaz a variabilidade espacial da produtividade florestal em comparação com modelos empíricos convencionais.

Melo (2019) avançou ainda mais nessa linha de pesquisa, investigando o potencial de diferentes algoritmos de ML, incluindo *Random Forest* e *Support Vector Machine*, para modelar

o crescimento e a produção em florestas inequidêneas de eucalipto. Os resultados indicaram que, embora os AM geralmente superem os modelos tradicionais em termos de precisão, a magnitude dessa vantagem varia conforme as características do conjunto de dados e a complexidade do fenômeno modelado.

Goycochea Casas (2021) comparou métodos de regressão tradicional e RNA na modelagem de crescimento e produção de povoamentos de eucalipto. Embora ambas as abordagens tenham apresentado desempenho satisfatório, o autor observou que as RNAs foram particularmente eficazes para capturar padrões não-lineares nas fases iniciais de desenvolvimento. No entanto, recomendou cautela ao realizar projeções a partir de idades muito jovens (próximas a dois anos), independentemente do método utilizado.

Uma abordagem particularmente promissora envolve a integração de variáveis derivadas de sensoriamento remoto com AM para melhorar as estimativas de crescimento e produção florestal. Araujo et al. (2023) avaliaram diferentes algoritmos de aprendizado de máquina para a prognose de volume, comparando o modelo de Clutter com técnicas como *Random Forest*, *Support Vector Machine*, k-vizinhos mais próximos e redes neurais artificiais. Os modelos foram ajustados com a inclusão de variáveis dendrométricas e índices espectrais derivados de sensoriamento remoto. Os resultados indicaram que a máquina de vetores de suporte, utilizando as mesmas variáveis preditoras do modelo de Clutter, apresentou uma modesta melhoria no desempenho. A inclusão de variáveis de sensoriamento remoto proporcionou um aprimoramento adicional, sendo o índice de vegetação VARI (*Visible Atmospherically Resistant Index*) aquele que resultou nas melhores estimativas.

Santos et al. (2023) propuseram uma metodologia inovadora baseada em dados do satélite Landsat para prever o crescimento de povoamentos de eucalipto, comparando-a com métodos tradicionais. Além de avaliar os benefícios da inclusão de variáveis derivadas de sensores remotos, o estudo investigou a possibilidade de reduzir o número de parcelas inventariadas, analisando três cenários amostrais distintos. Utilizando redes neurais artificiais e *Random Forest* para estimar o crescimento do povoamento, os autores concluíram que a amostragem do inventário florestal poderia ser reduzida em até 88% quando informações complementares de sensoriamento remoto são incorporadas à modelagem, sem comprometer significativamente a qualidade das estimativas.

Apesar dos resultados promissores, a aplicação de AM na modelagem florestal ainda enfrenta desafios importantes. Primeiro, muitos algoritmos de AM funcionam como "caixas-

pretas", dificultando a interpretação dos processos subjacentes e limitando a transferência de conhecimento para sistemas diferentes daqueles em que foram treinados. Segundo a qualidade das previsões depende fortemente da representatividade e qualidade dos dados de treinamento, podendo resultar em extrapolações inadequadas quando aplicados fora do domínio de calibração. Terceiro, a seleção e ajuste de hiperparâmetros podem ser processos complexos e computacionalmente intensivos, exigindo conhecimento especializado e recursos substanciais (Tsakiridis et al., 2020).

Para superar essas limitações, abordagens híbridas que combinam o conhecimento ecofisiológico incorporado em modelos baseados em processos com a flexibilidade e capacidade preditiva dos AM têm sido propostas como uma direção promissora para pesquisas futuras. Essas abordagens buscam aproveitar o melhor de ambos os mundos: a base teórica sólida dos modelos processuais e a adaptabilidade empírica dos algoritmos de aprendizado de máquina.

## **MATERIAL E MÉTODOS**

### **1.6 Material**

#### ***1.6.1 ÁREA DE ESTUDO***

A área de estudo está localizada na mesorregião do Vale do Rio Doce, estado de Minas Gerais, Brasil (Figura 5). Os talhões estão inseridos em onze diferentes municípios, sendo o município de Peçanha com o maior número de talhões e a área média de 39,295 ha. A altitude média da região é de 860 metros, com variação de 80 metros, com índices pluviométricos médio anual de 1000 mm a 1500 mm. As classes de solo predominante são latossolo vermelho-amarelo distrófico e cambissolo háplico com argila de atividade baixa distrófico (IBGE, 2003).

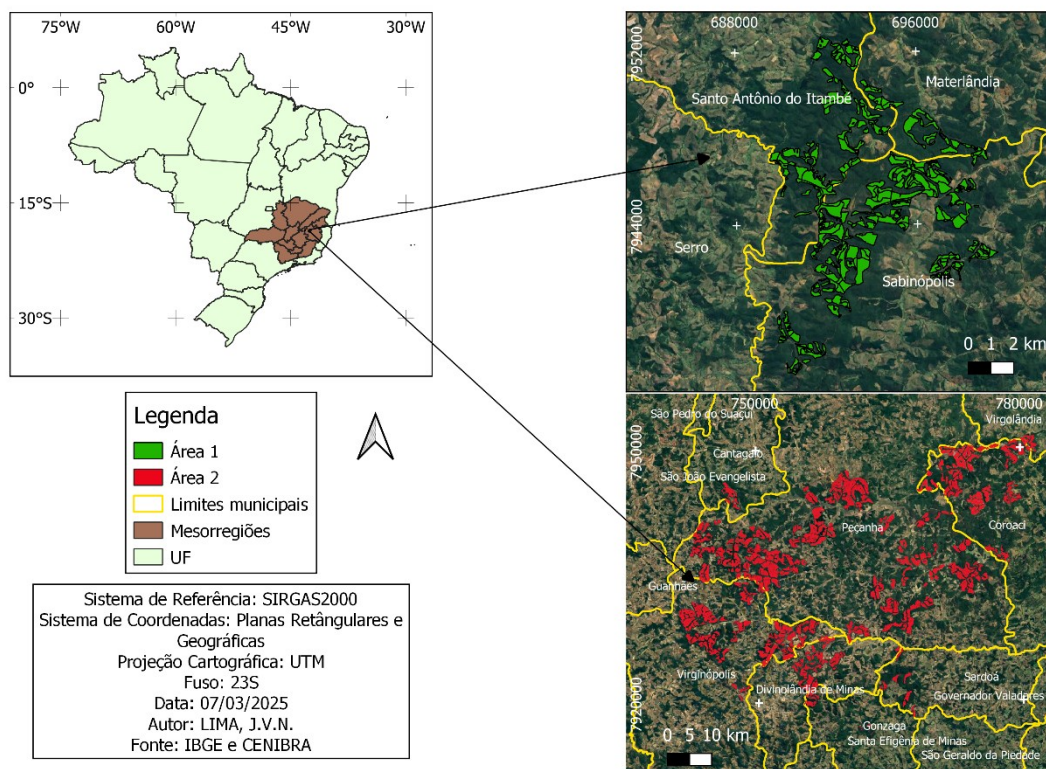


Figura 5 – Mapa da área de localização dos talhões

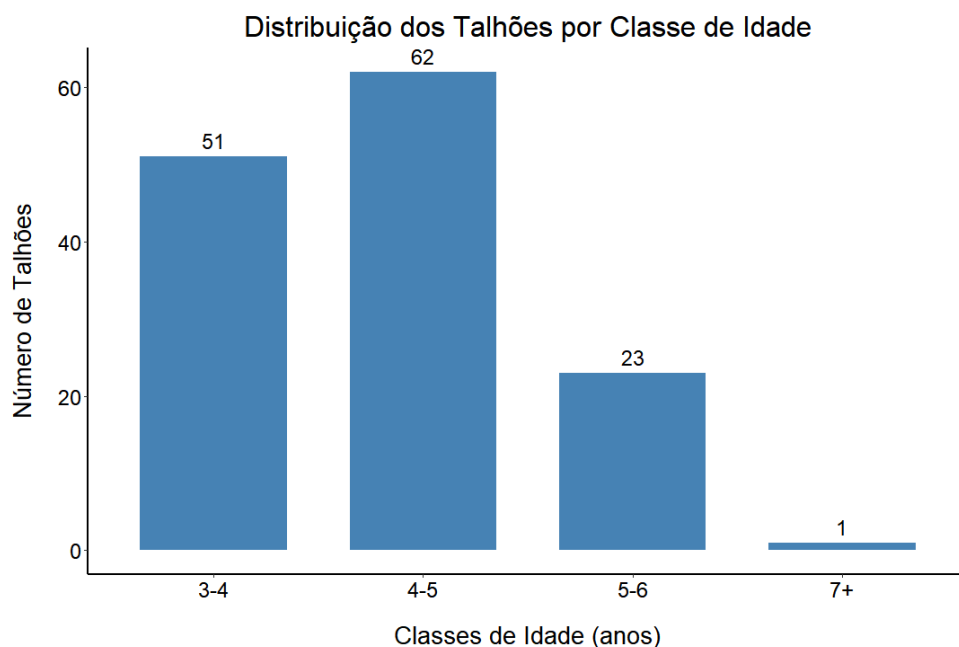
### 1.6.2 DADOS DE CAMPO

Foi utilizada uma base de dados derivada de inventário florestal contínuo provida pela companhia Celulose Nipo-Brasileira S/A (CENIBRA). O período contemplado pelas medições é de junho de 2013 a julho de 2019. Foram utilizados dados de talhões com pelo menos três medições anuais e sobrevivência acima de 90%. A Tabela 2 mostra os valores mínimos, máximos e a média da área basal e volume.

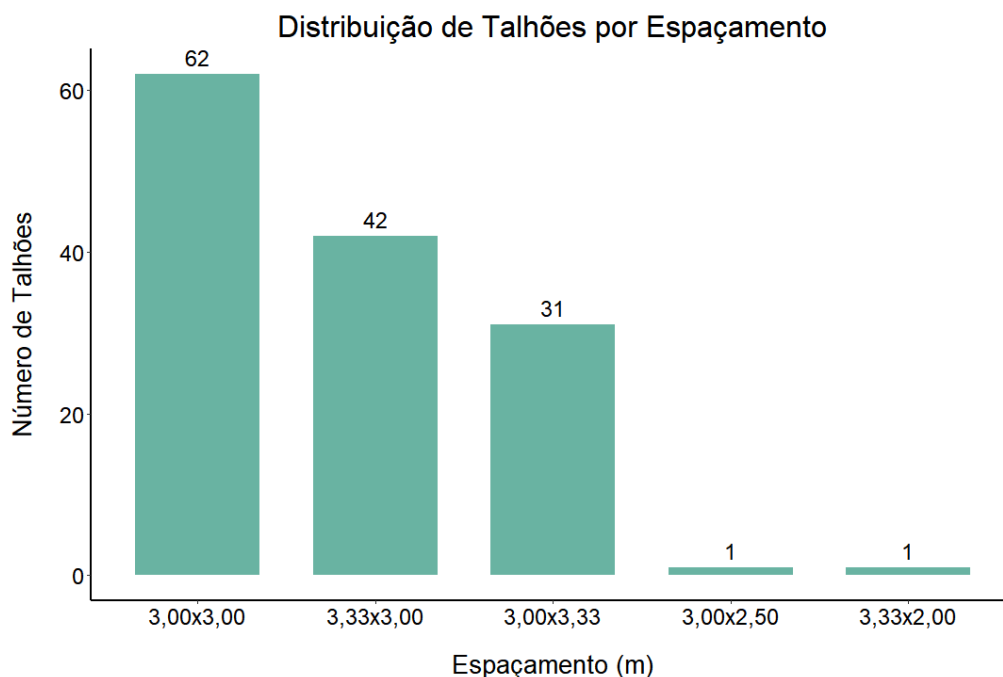
Tabela 2 – Estatísticas descritivas dos talhões

	<b>Altura Dominante (m)</b>	<b>Área Basal (m<sup>2</sup>/ha)</b>	<b>Volume (m<sup>3</sup>/ha)</b>
Mínimo	11,10	5,40	20,96
Máximo	36,70	32,38	495,33
Média	23,96	17,40	187,90
Desvio Padrão	4,66	4,93	82,56

As variáveis, em nível de unidade amostral, do inventário contínuo, incluem: data da medição, idade do talhão (em meses), espaçamento, área basal, altura dominante e volume comercial (até 5 cm de diâmetro mínimo) sem casca. A maioria dos talhões encontrava-se no segundo ciclo com idade média de 4 anos (Figura 6). Os talhões monoclonais, contemplam 28 diferentes genótipos plantados em sete diferentes espaçamentos: 3 x 2,5 m; 3 x 3 m; 3 x 3,30 m; 3 x 3,33 m; 3,33 x 2 m; 3,33 x 3 m e 4 x 2,5 m (Figura 7).



*Figura 6 – Distribuição dos talhões do inventário contínuo nas classes de idade.*



*Figura 7 – Distribuição dos talhões do inventário por espaçamento*

### 1.6.3 VARIÁVEIS DE SENSORIAMENTO REMOTO

O satélite Landsat 8 foi escolhido devido à sua cobertura histórica compatível com o período do inventário florestal, além de suas especificações técnicas favoráveis. O sensor OLI (*Operational Land Imager*) possui uma resolução espacial de 30 metros, o que é ideal para o monitoramento de vegetação em grandes áreas. Além disso, conta com uma resolução radiométrica de 16 bits e uma resolução temporal de 16 dias, permitindo a aquisição frequente de dados. As imagens da série Landsat destacam-se pela sua alta qualidade radiométrica e espacial, cobertura global e aquisição contínua, tornando-se uma ferramenta confiável para o monitoramento ambiental e estudos florestais.

As imagens Landsat 8 foram obtidas no site da USGS (*United States Geological Survey*) correspondentes ao período de 2013 a 2019, coincidindo com o período do inventário florestal. Foram selecionadas somente as imagens sem presença de nuvens e que possuíam nível 2 de processamento, o que garante que os pixels representem valores de reflectância corrigidos radiometricamente e atmosféricamente, diferentemente dos valores de números digitais (ND) do nível 1. O sensor OLI a bordo do Landsat 8 abrange desde as bandas do visível (azul, verde e vermelho) até as bandas do infravermelho médio (SWIR), conforme apresentado na Tabela 3.

TABELA 3 – Características das bandas espectrais

Bandas	Comprimento da onda	Resolução Espacial
Banda 1 Aerossol Costeiro	0,43 – 0,45 $\mu\text{m}$	30 m
Banda 2 Azul	0,45 – 0,51 $\mu\text{m}$	30 m
Banda 3 Verde	0,53 – 0,59 $\mu\text{m}$	30 m
Banda 4 Vermelho	0,64 – 0,67 $\mu\text{m}$	30 m
Banda 5 Infravermelho próximo	0,85 – 0,88 $\mu\text{m}$	30 m
Banda 6 SWIR	1,57 – 1,65 $\mu\text{m}$	30 m
Banda 7 SWIR	2,11 – 2,29 $\mu\text{m}$	30 m
Banda 8 Pancromática	0,50 – 0,68 $\mu\text{m}$	15 m
Banda 9 Cirrus	1,36 – 1,38 $\mu\text{m}$	30 m

## 1.7 Processamento dos dados

A metodologia foi seguida conforme o fluxograma da Figura 8.

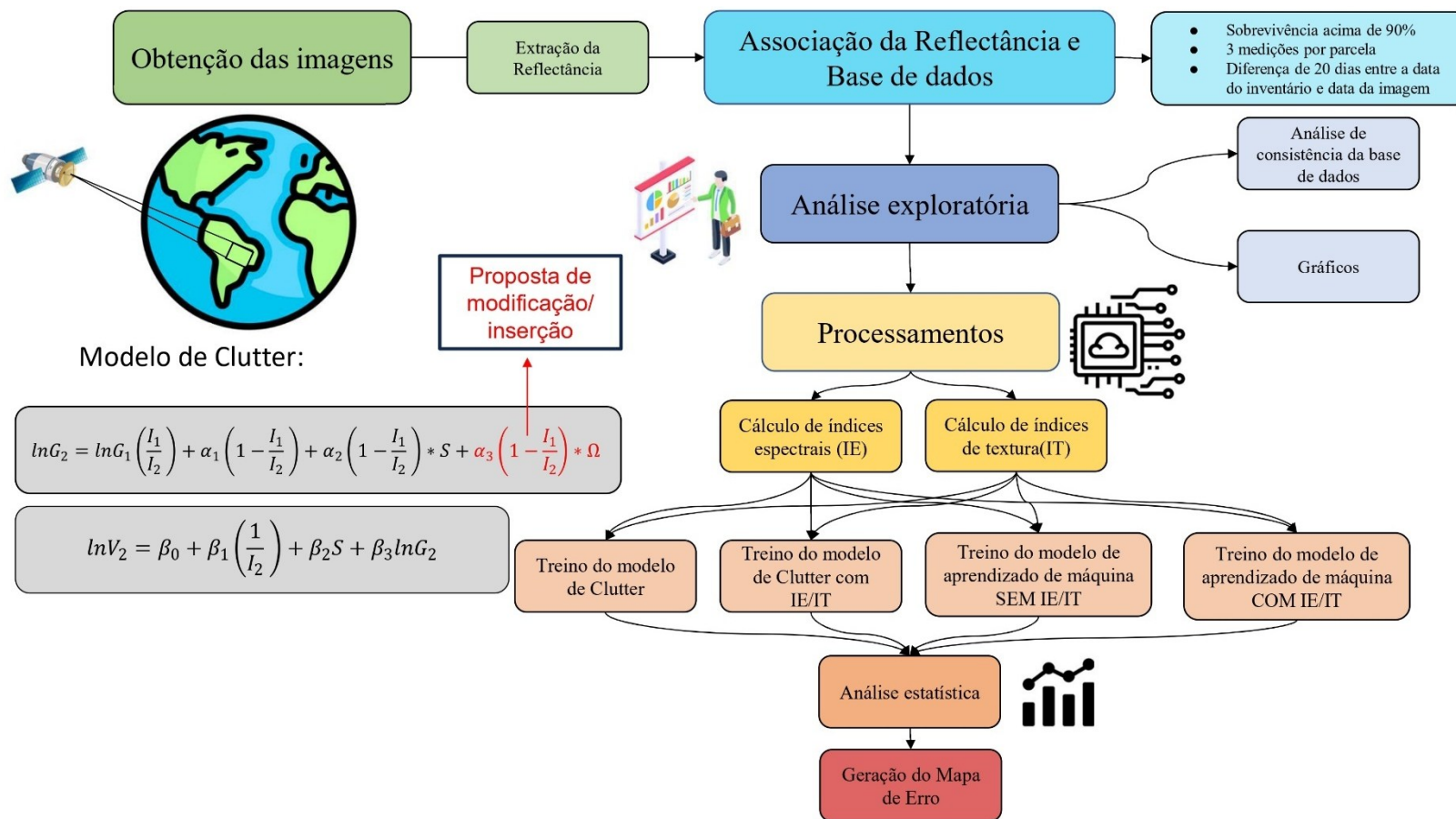
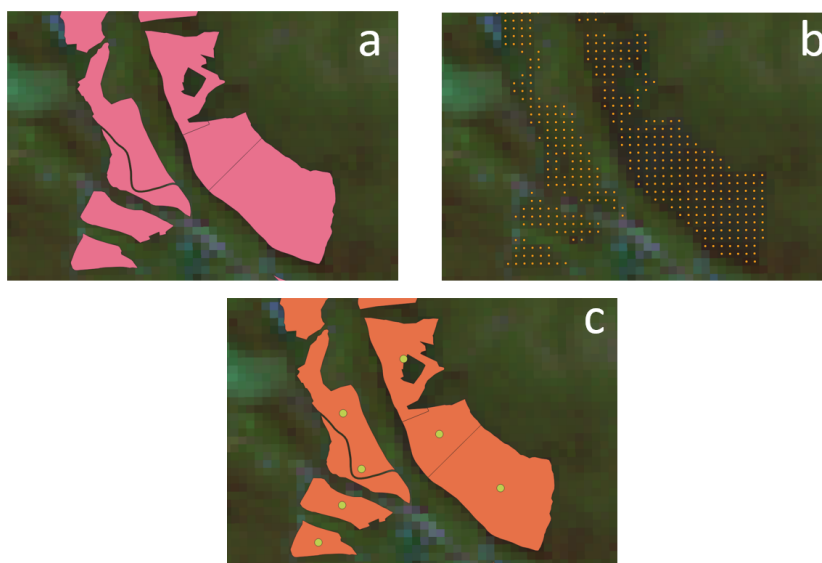


Figura 8 - Fluxograma das análises do estudo

A extração dos valores de reflectância foi realizada utilizando a linguagem de programação R com o uso do pacote *terra*. O procedimento envolveu: (a) a extração do centroide de cada pixel contido em cada polígono (talhão), (b) a atribuição dos valores de reflectância a cada ponto extraído e (c) o cálculo da média dos pontos interceptados pelo polígono (Figura 9). Este processo foi repetido para todas as imagens sem presença de nuvens, resultando em um valor médio de reflectância para cada banda e para cada talhão. O arquivo resultante foi exportado em formato CSV (*Comma Separete Value*) para posterior associação com a base de dados do inventário. Esse procedimento foi realizado para todas as imagens encontradas que não apresentava nuvens. Dessa forma foi obtido um valor de reflectância médio de todas as bandas para todos os talhões. Esse procedimento foi realizado devido à ausência das parcelas georreferenciadas.



*Figura 9 - Processo de extração da reflectância*

A associação dos dados de reflectância com o inventário florestal foi realizada no ambiente virtual do (Figura 10), utilizando dois arquivos de entrada: (i) um contendo os valores médios de reflectância para cada talhão e (ii) outro com os dados do inventário florestal. Para garantir a melhor correspondência temporal, a associação considerou a menor diferença possível entre as datas do inventário e as datas das imagens, de modo que cada registro do inventário fosse vinculado à imagem mais próxima em termos de data. Foi considerado também nessa associação, a sobrevivência acima de 90% e três medições de parcelas. O arquivo resultante foi gerado no formato XLSX, contendo os valores médios de reflectância associados a cada talhão. Dessa forma, cada medição do inventário foi relacionada aos valores de

reflectância das bandas espectrais (do azul ao SWIR 2), permitindo a realização da análise exploratória dos dados.

A análise exploratória dos dados foi inicialmente conduzida no Excel, com o objetivo de verificar a consistência das informações. Por exemplo, foram identificadas inconsistências como a redução na altura média em algumas parcelas e variações abruptas no DAP médio em medições sucessivas. Além disso, outliers decorrentes da interferência de nuvens nas imagens foram removidos com base em um critério estatístico de 1 desvio-padrão acima e abaixo da média da reflectância das bandas espectrais. Em seguida, foram construídos gráficos bivariados para a exploração da relação entre as variáveis dendrométricas e entre as variáveis dendrométricas e variáveis de sensoriamento remoto. Essas análises forneceram subsídios para a seleção das variáveis a serem incluídas nos modelos de Clutter e de aprendizado de máquina.

Para cada ano das séries temporais foram calculados diferentes índices de vegetação (Tabela 4) para serem usados como possíveis preditores do modelo. Esses índices possuem relação direta com variáveis dendrométricas e biofísicas de povoamentos florestais, sendo, portanto, fundamentais para análises em ambientes silviculturais. Além disso, são amplamente utilizados na literatura atual, o que reforça sua adequação para estudos em florestas plantadas. Adicionalmente, foram calculados índices texturais (Tabela 5). Os índices de texturas são baseados nas métricas GLCM (*Gray Level Co-occurrence*) ou matriz de coocorrência de nível de cinza e são amplamente utilizados em sensoriamento remoto para quantificar a variação espacial dos tons de cinza em uma imagem orbital. São bastante úteis para distinguir padrões de cobertura vegetal, identificar variações estruturais em dossel arbóreo (Haralick et al., 1987).

*Tabela 4 – Índices de Vegetação usados*

<b>Índice de vegetação</b>	<b>Equação</b>	<b>Referência</b>
Índice de Vegetação por Diferença Normalizada	$NDVI = \frac{NIR - RED}{NIR + RED}$	(Rouse Jr et al., 1973)
Índice de Vegetação por Diferença Normalizada Verde	$GNDVI = \frac{NIR - GREEN}{NIR + GREEN}$	(Gitelson et al., 1996)

Índice de vegetação ajustado ao solo	$SAVI = \left( \frac{NIR - RED}{NIR + RED + L} \right) + (1 + L)$	(Huete, 1988)
Razão Simples	$SR = \left( \frac{NIR}{RED} \right)$	(Jordan, 1969)
Índice de refletância de queima normalizado	$NBRI = \frac{NIR - SWIR2}{NIR + SWIR2}$	(García & Caselles, 1991)
Índice de Vegetação Proporcional	$RVI = \left( \frac{RED}{NIR} \right)$	(Broge & Leblanc, 2001)
Índice de Vegetação Diferencial	$DVI = 2,4 * NIR - RED$	(Richardson, 1977)
Índice de Vegetação Aprimorado	$EVI = 2,5 * \frac{(NIR - RED)}{(NIR + 6 * RED - 7,5 * BLUE + 1)}$	(Liu & Huete, 1995)

*Tabela 5 – Índices de Vegetação usados*

Índice de textura	Equação
Energia	$\sum_{i,j} g(i,j)^2$
Entropia	$\sum_{i,j} g(i,j) \log g(i,j)$
Correlação	$\sum_{i,j} \frac{(i - \mu)(j - \mu)g(i,j)}{\sigma^2}$
Momento de Diferença Inversa (IDM)/Homogeneidade	$\sum_{i,j} \frac{1}{1 + (i - j)^2} g(i,j)$
Inércia/Contraste	$\sum_{i,j} (i - j)^2 g(i,j)$
Sombra do Cluster (CS)	$\sum_{i,j} ((i - \mu)(j - \mu))^3 g(i,j)$
Proeminência do Cluster (CP)	$\sum_{i,j} ((i - \mu)(j - \mu))^4 g(i,j)$

Correlação de Haralick (HC) 
$$\sum_{i,j} \frac{(i - \mu)(j - \mu)g(i,j) - \mu_t^2}{\sigma_t^2}$$

---

Fonte: (*Feature extraction — Orfeo ToolBox 9.1.0 documentation*, [s.d.])

Onde  $g(i, j)$  é o valor normalizado da matriz de coocorrência GLCM,  $\mu$  a média e  $\sigma$  o desvio-padrão das linhas e colunas da GLCM.

### 1.7.1 AJUSTE DO MODELO DE CLUTTER

O modelo de prognose de volume adotado foi o de Clutter (Campos & Leite, 2006). Este modelo consiste em um sistema de duas equações: a primeira estima a área basal em uma idade de interesse e a segunda estima o volume nessa mesma idade. A estimativa da área basal é um preditor da equação de volume como apresentado abaixo:

$$\ln G_2 = \ln G_1 \left( \frac{I_1}{I_2} \right) + \alpha_1 \left( 1 - \frac{I_1}{I_2} \right) + \alpha_2 \left( 1 - \frac{I_1}{I_2} \right) * S$$

$$\ln V_2 = \beta_0 + \beta_1 \left( \frac{1}{I_2} \right) + \beta_2 S + \beta_3 \ln G_2$$

Onde:  $\ln G_1$  é a área basal atual,  $I_1$  é a idade atual em meses,  $I_2$  é a idade de interesse em meses,  $S$  é o índice de sítio em metros,  $\ln G_2$  é a área basal na idade interesse em  $m^2$ ,  $\ln V_2$  é o volume do talhão na idade de interesse em  $m^3$ ,  $\alpha_1$ ,  $\alpha_2$ ,  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$  são coeficientes do modelo. Para a realização do ajustamento do modelo, foi utilizado o Método Mínimos Quadrado (MMQ) em dois estágios, conforme Leite e Campos (2017).

Para estimativa do índice de sítio usado como variável do modelo de Clutter foram avaliados o modelo de Schumacher, Logístico e Gompertz (Tabela 6). O índice de sítio é dado pela altura dominante em uma idade de referência, geralmente aos 7 anos para eucalipto no Brasil. É um índice de capacidade produtiva de um talhão e utilizado dentro do manejo florestal para classificar e comparar diferentes áreas de povoamento florestal com base no potencial crescimento das árvores. A escolha do melhor modelo se deu através do cálculo das medidas de acurácia REMQ,  $R^2$ , e EMA calculadas via validação cruzada. Já o modelo de Clutter foi ajustado usando no R usando o pacote dplyr e caret seguindo o método dos mínimos quadrado em dois estágios, conforme Leite e Campos (2017).

*Tabela 6 – Modelos de índice de sítio pelo método da curva-guia*

<b>Modelo</b>	<b>Equação base</b>	<b><i>Equação de sítio</i> Método da curva-guia</b>
Schumacher	$\ln(Hdom) = \beta_0 + \beta_1 \left(\frac{1}{I_1}\right)$	$\ln(S) = \ln(Hdom) - \beta_1 \left[\left(\frac{1}{I_1}\right) - \left(\frac{1}{I_2}\right)\right]$
Logístico	$Hdom = \frac{\beta_0}{(1 + \beta_1 e^{(-\beta_2 I_1)})}$	$S = Hdom * \frac{(1 + \beta_1 e^{(-\beta_2 I_1)})}{(1 + \beta_1 e^{(-\beta_2 I_2)})}$
Gompertz	$Hdom = \beta_0 \left[e^{-e^{(\beta_1 - \beta_2 I_1)}}$	$S = Hdom * \frac{(e^{-e^{(\beta_1 - \beta_2 I_2)}})}{(e^{-e^{(\beta_1 - \beta_2 I_1)}})}\right]$

Para incorporar a variável que representam a capacidade fotossintética, os índices espectrais e texturais calculados foram adicionados à equação da área basal. A inclusão dos índices espectrais foi definida com base em análises preliminares e em evidências apresentadas na literatura (Pereira et al., 2016), as quais indicaram sua relevância para a melhoria do ajuste do modelo, em virtude de sua relação direta com a área foliar e com a biomassa do povoamento. Assim, os índices foram inseridos empiricamente na equação correspondente à área basal.

$$\ln G_2 = \ln G_1 \left(\frac{I_1}{I_2}\right) + \alpha_1 \left(1 - \frac{I_1}{I_2}\right) + \alpha_2 \left(1 - \frac{I_1}{I_2}\right) * S + \alpha_3 \left(1 - \frac{I_1}{I_2}\right) * \Omega$$

Onde  $\Omega$  corresponde ao índice espectral,  $\alpha_3$  corresponde ao parâmetro de ajuste da equação, e foi adicionado  $\left(1 - \frac{I_1}{I_2}\right)$  pois quando  $I_2 = I_1$  a área basal corresponderá a mesma área basal na data do inventário e o modelo continuará consistente. Foram testados os seguintes IE e IT: *NDVI, GNDVI, SAVI, SR, RVI, DVI, NBRI, EVI, Energy, Entropy, Correlation, IDM, Inertia, CS, CP, HC*.

A inclusão dessas formas busca representar mais fielmente o comportamento da vegetação captado pelos sensores orbitais, que muitas vezes apresenta crescimento assintótico ou respostas parabólicas frente a diferentes fases de desenvolvimento do povoamento. O ajuste do modelo de Clutter modificado também foi realizado através do método dos mínimos quadrados em dois estágios. Para a equação de volume, optou-se por não incluir o parâmetro  $\Omega$ , uma vez que sua adição resultou em um aumento do RMSE%, indicando uma redução na precisão do modelo. Além disso, o estudo de Pereira et al., (2016) incorporou uma componente espacial na equação de área basal e identificou uma dependência espacial dessa variável e o volume não apresentou correlação espacial.

### 1.7.2 TREINAMENTO DOS MODELOS DE APRENDIZADO DE MÁQUINA

Os métodos RF, SVM, XGBoost, RNA e KNN foram implementados no software R (versão 4.4.1), com o uso dos pacotes ranger, e1071, kernlab, xgboost e nnet, respectivamente. Todas as variáveis foram normalizadas por meio da função de normalização mínimo-máximo, a fim de padronizar as escalas e evitar influência desproporcional de variáveis com magnitudes distintas e ao final foram convertidas para a unidade original para o cálculo das estatísticas de qualidade.

$$z = \frac{x - \min(x)}{[\max(x) - \min(x)]}$$

$$x = z [\max(x) - \min(x)] + \min(x)$$

Onde  $z$  é o valor normalizado,  $x$  é o valor original da variável,  $\max(x)$  e  $\min(x)$  são, respectivamente, os valores máximo e mínimo observados para a variável.

Como variáveis preditoras, adotaram-se as mesmas variáveis do modelo de Clutter, a saber: idade atual, idade de interesse, área basal atual e índice de sítio. Os parâmetros dos modelos foram ajustados por meio de uma busca em grade (“grid search”), na qual se avaliaram diferentes combinações de hiperparâmetros. Ao final, selecionou-se o modelo com o melhor desempenho e os respectivos hiperparâmetros para a geração do mapa de erro. Os hiperparâmetros testados foram:

- XGBoost:
  - Número de árvores: 100 e 200
  - Profundidade máxima das árvores: 2, 4 e 5 (profundidade rasa, profunda e ainda mais profunda, respectivamente)
  - Taxa de aprendizado: 0,01; 0,1 e 0,3 (taxa de aprendizagem baixa, média e alta, respectivamente)
  - Valor mínimo para redução de perda: 0 (não há restrição para divisão)
  - Proporção de coluna: 1 (todas as variáveis são consideradas)
  - Peso mínimo da soma dos gradientes: 1 (árvores mais complexas)
  - Proporção das amostras: 1 (todas as amostras são usadas)
- SVM:
  - Parâmetro de regularização variando de  $2^{-5}$  até  $2^2$
  - Parâmetro do kernel radial variando de  $2^{-5}$  até  $2^2$

- RF
    - Número de variáveis: 2, 4 e 5 (2, 4 ou 5 variáveis aleatórias para decidir a divisão)
    - Critério de divisão do nós: variância
    - Número de observações em um nó terminal: 5 e 10
  - RNA
    - Número de neurônio testados: 5, 10 e 15
    - Parâmetro de regularização: 0,01; 0,1 e 0,5
    - Número de camada oculta: 1
    - A função de ativação: logística
- $$f(x) = \frac{1}{1 + e^{-x}}$$
- Número máximo de iterações: 100
  - KNN
    - Número de vizinhos considerados: 3, 5, 7, 9, 11, 13 e 15
    - Distância utilizada: distância Euclidiana

Na versão ampliada dos modelos, incorporaram-se índices espectrais e de texturas às variáveis preditoras do modelo de Clutter. As combinações testadas incluíram:

1. **IE únicos:** NDVI, GNDVI, SAVI, SR, RVI, DVI, NBRI, EVI, Energy, Entropy, Correlation, IDM, CShade, CProminence, HC, Inertia
2. **Combinações binárias:** NDVI+GNDVI, NDVI+SAVI, NDVI+SR, GNDVI+SAVI, GNDVI+SR, SAVI+SR
3. **Combinações ternárias:** NDVI + GNDVI + SAVI, NDVI + GNDVI + SR, GNDVI + SAVI + SR
4. **Combinação quaternária:** NDVI + GNDVI + SAVI + SR, RVI + DVI + NBRI + EVI, Energy + Entropy + Correlation + IDM, CShade + CProminence + HC + Inertia.
5. **Combinação de oito:** NDVI + GNDVI + SAVI + SR + RVI + DVI + NBRI + EVI, Energy + Entropy + Correlation + IDM + CShade + CProminence + HC + Inertia.

### 1.7.3 AVALIAÇÃO DOS MODELOS

Para a avaliação dos modelos de sítio (usados dentro do modelo de Clutter), do modelo de Clutter e dos modelos de aprendizado de máquina, foi adotado o método de *Cross-Validation* (CV) utilizando 5 *folds* como demonstrado na Figura 11.



Figura 11 – modelo de validação CV

A validação cruzada (*Cross Validation* – CV) é uma das técnicas mais comuns em modelagem estatística e aprendizado de máquina, utilizada para avaliar o desempenho e a capacidade de generalização dos modelos. Nesse procedimento, o conjunto de dados é dividido aleatoriamente em cinco subconjuntos (*folds*), dos quais quatro são utilizados para o treinamento do modelo e um para a validação. Esse processo é repetido até que todos os subconjuntos tenham sido utilizados como conjunto de validação. A aplicação da CV reduz o risco de enviesamento dos conjuntos de teste, tornando as métricas de desempenho, como RMSE,  $R^2$  e MAE, mais estáveis e representativas. Além disso, contribui para a detecção do overfitting (sobreajuste dos dados) e pode ser empregada tanto em modelos paramétricos, empíricos quanto em modelos de aprendizado de máquina (Browne, 2000). Adicionalmente, foram calculadas as estatísticas de desvio-padrão, valores mínimos e máximos para todos os *folds*, de modo a fornecer uma referência sobre a variação das métricas descritivas ao longo da validação cruzada.

### 1.8 Critérios de seleção de modelos

Para as análises de estatística de qualidade dos modelos, foram utilizadas as estatísticas de Raiz do Erro Médio Quadrático (RQEM), Raiz do Erro Médio Quadrático (RQEM%), correlação de Pearson ( $r$ ), coeficiente de determinação ( $R^2$ ) e Viés%.

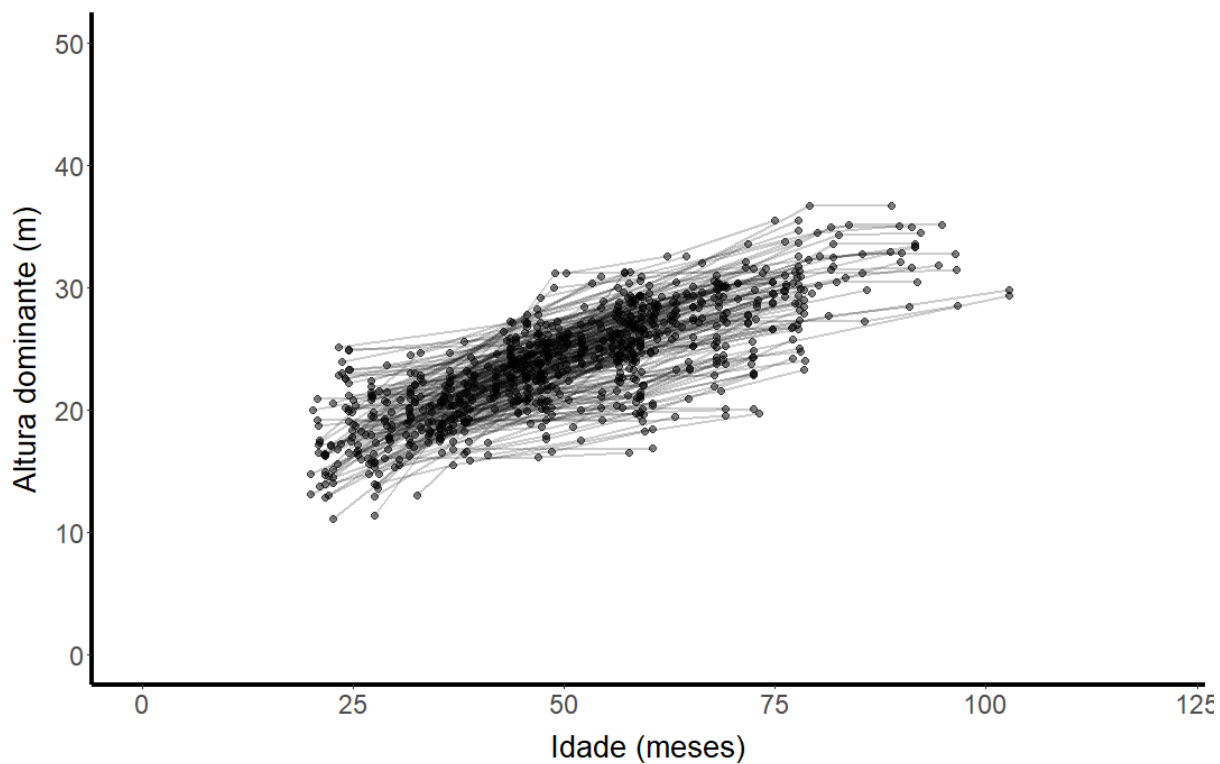
$$\begin{aligned}
REM Q &= \sqrt{\frac{\sum_{i=1}^n (Vpredito_i - Vreal_i)^2}{n}} \\
REM Q\% &= \frac{\sqrt{\frac{\sum_{i=1}^n (Vpredito_i - Vreal_i)^2}{n}}}{\sum_{i=1}^n Vreal_i} * n * 100 \\
Viés\% &= \frac{\sum_{i=1}^n (Vpredito_i - Vreal_i)}{\sum_{i=1}^n Vreal_i} * 100 \\
r &= \frac{\sum_{i=1}^n (Vpredito_i - \bar{Vpredito})(Vreal_i - \bar{Vreal})}{\sqrt{\left(\sum_{i=1}^n (Vpredito_i - \bar{Vpredito})^2\right) \left(\sum_{i=1}^n (Vreal_i - \bar{Vreal})^2\right)}} \\
R^2 &= 1 - \frac{\sum_{i=1}^n (Vreal_i - Vpredito_i)^2}{\sum_{i=1}^n (Vreal_i - \bar{Vreal})^2}
\end{aligned}$$

O REMQ mede o erro médio entre os valores observados e preditos. De forma simples mostra o quão longe, em média, as previsões estão da realidade. REMQ% é o REMQ em porcentagem e permite comparar modelos em diferentes escalas, facilitando a interpretação prática dos erros. Viés% indica se os modelos tendem a superestimar ou subestimar os valores. Um viés mais baixo indica que o modelo não possui tendência sistemáticas.  $r$  é o coeficiente de correlação de Pearson onde mede a força da relação linear entre os valores observados e valores preditos.  $R^2$  mostra a proporção da variabilidade dos dados observados explicadas pelo modelo. As análises estatísticas foram realizadas para todos os processamentos, desde a seleção do índice de sítio até nos ajustes do modelo de Clutter e modelo de aprendizado de máquina.

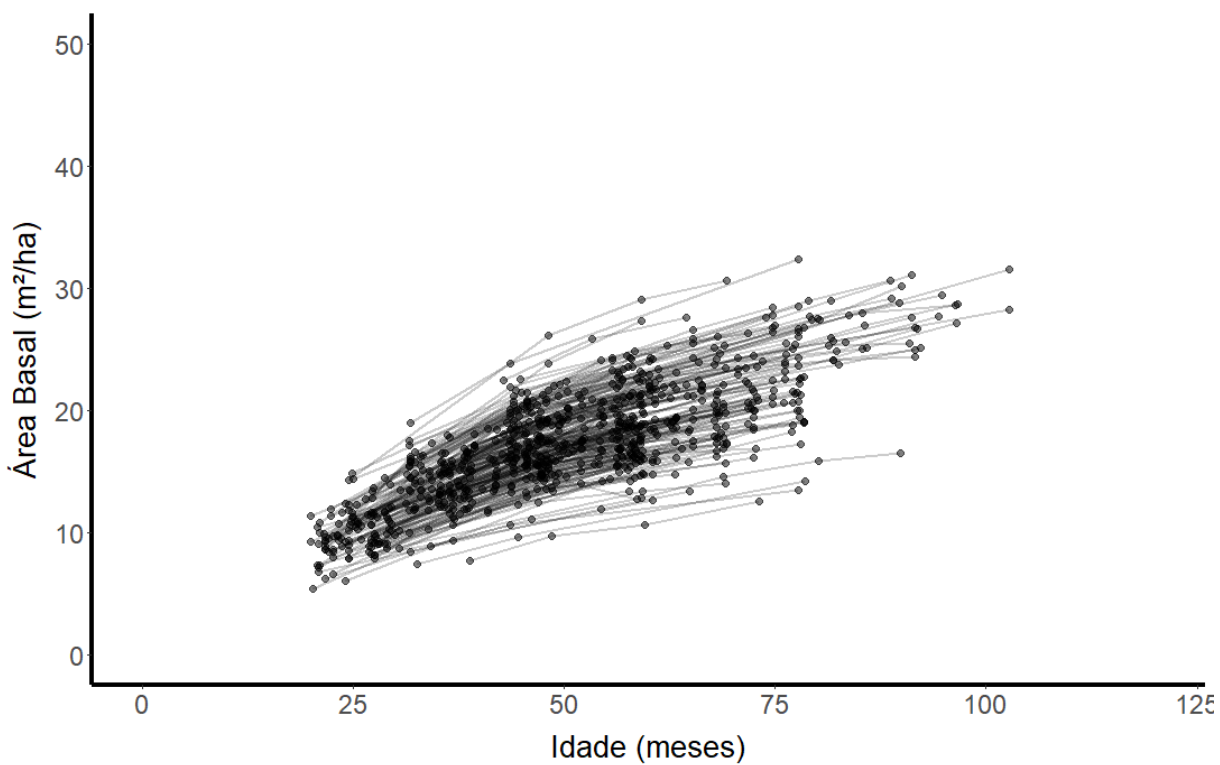
## RESULTADOS

### 1.9 Análise exploratória

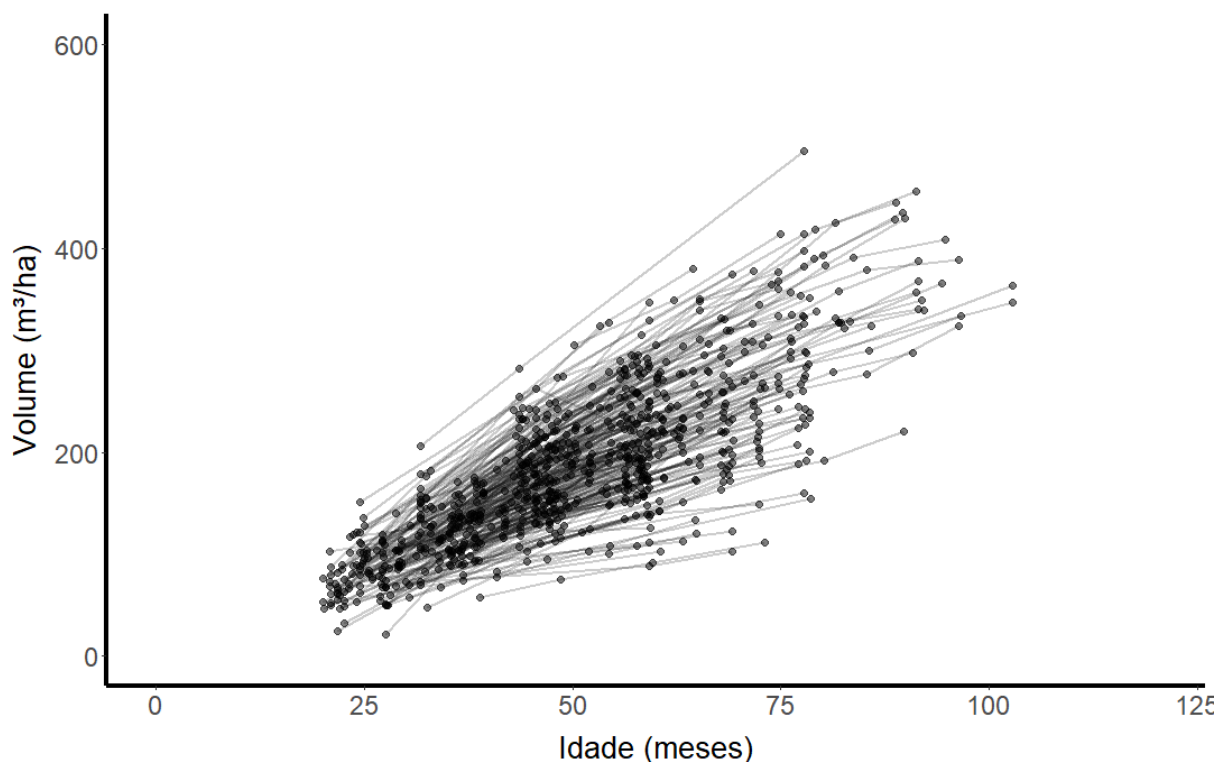
Nas Figuras 12, 13 e 14 são apresentados os gráficos correspondentes às variáveis biométricas altura dominante, área basal e volume, respectivamente. Em todos os gráficos, as linhas que conectam os pontos representam a trajetória temporal de cada talhão ao longo do período analisado.



*Figura 12 – Gráfico da altura dominante e idade*



*Figura 13 – Gráfico da área basal e idade*



*Figura 14 – Gráfico de volume e idade*

A análise exploratória da variável altura dominante indica um padrão de dispersão coerente com a dinâmica de crescimento florestal, caracterizado pelo aumento dos valores em função da idade e pela tendência à estabilização a partir de aproximadamente 70 meses. Observa-se, contudo, elevada dispersão dos dados, refletindo a heterogeneidade dos talhões avaliados. Para a área basal, verificou-se comportamento semelhante e esperado ao longo do tempo, com rápido incremento nos primeiros anos de desenvolvimento, seguido de estabilização após cerca de 60 meses. Em relação ao volume, não foi observada redução na taxa de crescimento, sendo predominante um comportamento aproximadamente linear em função do tempo. Adicionalmente, os gráficos evidenciam significativa dispersão dos dados ao longo da série temporal, associada às diferenças entre talhões e genótipos analisados.

A Figura 15 apresenta a relação entre os índices espectral e de textura, representados no eixo y, e a idade dos talhões, expressa em anos no eixo x. Enquanto a Figura 16 apresenta a relação dos índices e do volume. A linha azul da Figura 15 e 16 representa a tendência média das observações ao longo do tempo, obtida por meio de um ajuste suavizado (smooth).

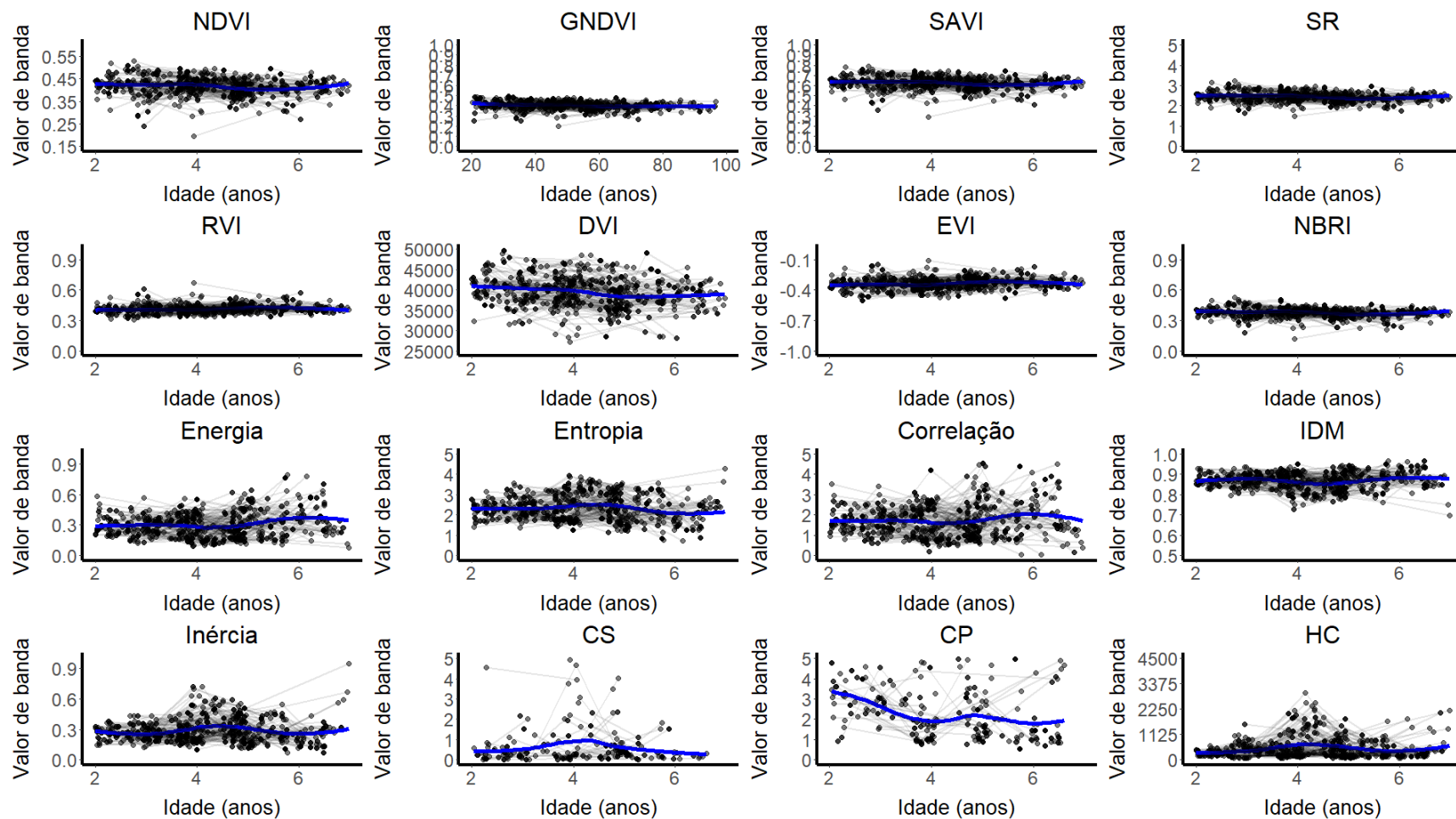


Figura 15 – Gráfico da relação entre IE/IT e Idade

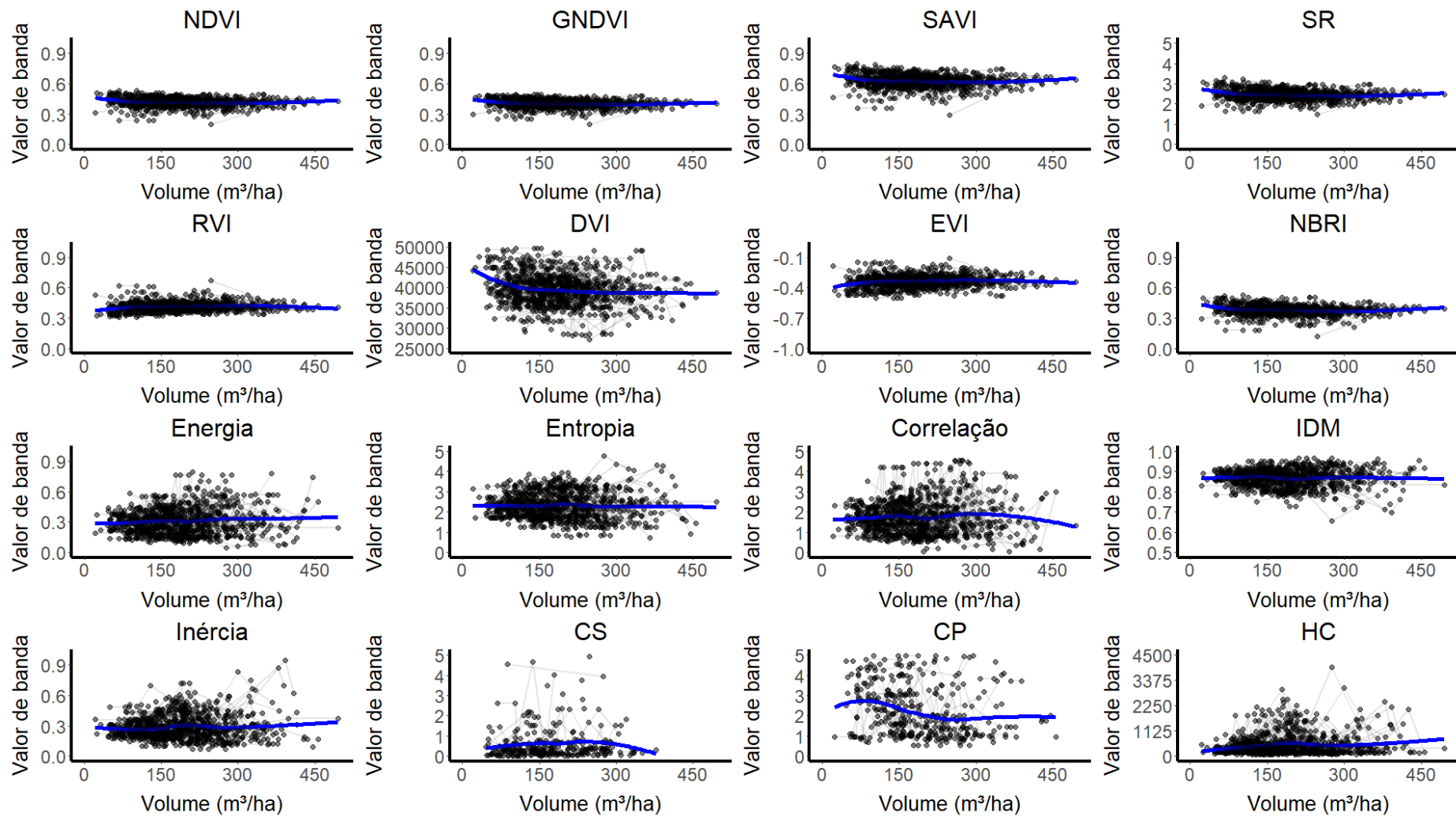


Figura 16 – Gráfico da relação entre IE/IT e Volume.

Na Figura 15, observa-se que os índices espectrais apresentam variação limitada ao longo da idade, com tendências suaves de aumento ou estabilização após os primeiros anos. Esse comportamento é típico de povoamentos florestais homogêneos e manejados, nos quais o rápido fechamento do dossel ocorre ainda nas fases iniciais do crescimento. O NDVI, SAVI e EVI exibem valores relativamente elevados já a partir dos primeiros anos, mantendo-se estáveis ao longo do tempo. Esse padrão indica que o aumento inicial da área foliar promove rápida elevação da reflectância no infravermelho próximo e redução no vermelho, refletindo maior interceptação de radiação fotossinteticamente ativa e maior eficiência fotossintética.

Os índices de textura apresentam maior dispersão e variabilidade ao longo da idade, refletindo mudanças na heterogeneidade espacial do dossel. O aumento gradual da Energia e do IDM (Homogeneidade) indica que, com o avanço da idade, o povoamento tende a se tornar mais estruturalmente uniforme, com copas mais contínuas e redução de variações locais de tons de cinza.

Já na Figura 16, de forma geral, observa-se que os índices espectrais apresentam tendências fracas a moderadas em relação ao volume, com padrões não lineares e grande dispersão dos dados, especialmente em volumes mais elevados. NDVI, GNDVI e SAVI exibem comportamento de leve decréscimo ou estabilização à medida que o volume aumenta. Esse padrão é consistente com o fenômeno de saturação espectral, no qual incrementos adicionais de biomassa e área foliar não resultam em variações proporcionais nos índices baseados na razão entre o infravermelho próximo e o vermelho (ou verde). Em volumes baixos a intermediários, os índices apresentam maior sensibilidade, refletindo o aumento da área foliar, maior interceptação da radiação fotossinteticamente ativa e maior acúmulo de biomassa. Entretanto, em talhões com volumes elevados, a capacidade discriminatória desses índices diminui substancialmente.

Em contraste com os índices espectrais, os índices de textura apresentam respostas mais consistentes ao longo do gradiente de volume, refletindo alterações na organização espacial do dossel. A Energia e o IDM (Homogeneidade) tendem a aumentar com o volume, indicando que povoamentos mais volumosos apresentam estrutura espacial mais uniforme, com copas contínuas e menor variação local de tons de cinza. A Entropia mantém valores relativamente constantes, com leve aumento em volumes intermediários, sugerindo maior heterogeneidade estrutural durante fases de crescimento mais intenso, seguida de estabilização em povoamentos maduros. Esses resultados evidenciam que os índices de textura capturam

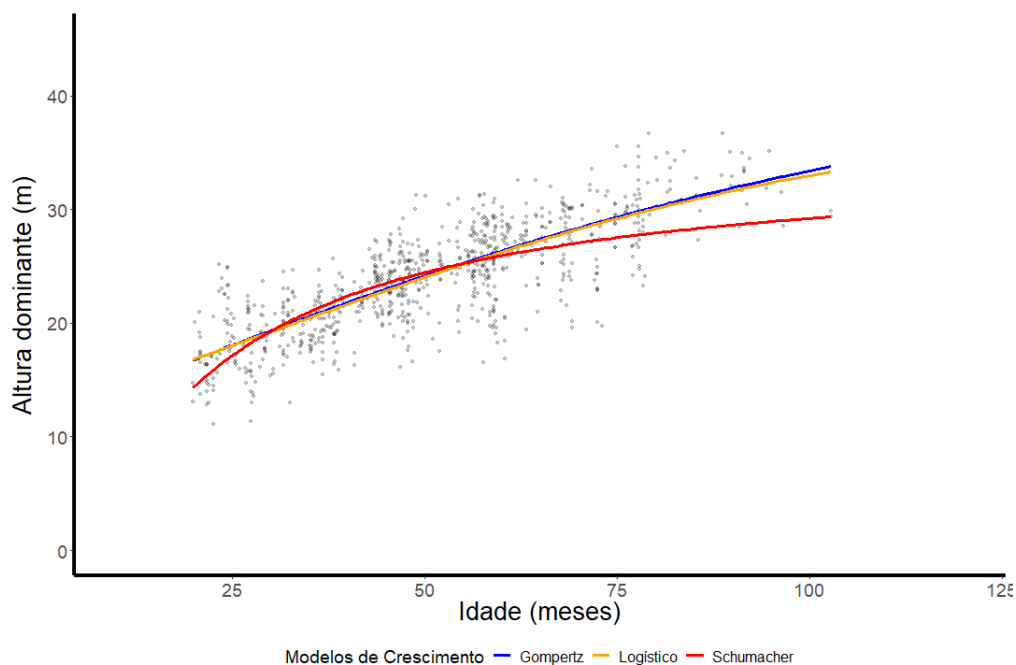
informações estruturais do povoamento que não são diretamente representadas pelos índices espectrais médios. De modo geral, todos os índices apresentaram baixa correlação com a idade, sugerindo pouca variação espectral ao longo dos anos.

### 1.9.1 SELEÇÃO DO MODELO DE ÍNDICE DE SÍTIO

A Tabela 7 apresenta os coeficientes estimados dos modelos de índice de sítio, assim como as respectivas estatísticas de qualidade do ajuste, para os três modelos de equação de sítio descritos na Tabela 5. Os resultados demonstraram alta significância estatística para todos os coeficientes avaliados, com valores de p inferiores a 0,0001, indicando que os parâmetros dos modelos apresentam contribuição significativa para as estimativas. Na comparação de desempenho entre os modelos testados, o modelo de Gompertz e Logístico mostraram-se mais preciso, apresentando menor valor de REMQ% para predição do índice de sítio. A Figura 17 mostra as curvas dos modelos de sítio ajustado.

*Tabela 7 – Coeficientes estimados e estatística de qualidade de cada modelo. Os valores em parênteses indicam o p-valor*

Modelo	Coeficientes Estimados			Estatísticas de qualidade			
	$\beta_1$	$\beta_2$	$\beta_3$	REMQ%	Bias	r	R <sup>2</sup>
Schumacher	35,533 (<0,0001)	-178,209 (<0,0001)	-	12,57	0,208	0,76	0,57
Gompertz	43,3130 (<0,0001)	0,2713 (<0,0001)	0,0161 (<0,0001)	11,87	0,002	0,79	0,62
Logístico	38,5921 (<0,0001)	21,431 (<0,0001)	0,0250 (<0,0001)	11,87	0,002	0,79	0,62



*Figura 17 – curvas dos índices de sítio*

Observa-se que todos os modelos são capazes de representar a tendência geral de crescimento em altura, caracterizada por um rápido incremento inicial seguido de redução gradual da taxa de crescimento com o avanço da idade, comportamento típico de espécies florestais de rápido crescimento.

### ***1.9.2 COMPARAÇÃO ENTRE O MODELO DE EM SUA ESTRUTURA TRADICIONAL E ADICIONADO DE IE/IT***

A Figura 18 apresenta a média da métrica de acurácia REMQ% para a estimativa da área basal (a) e do volume (b). Os resultados indicam elevada precisão dos modelos, com valores de REMQ% inferiores a 10%, coeficiente de determinação ( $R^2$ ) superior a 80% e coeficiente de correlação ( $r$ ) acima de 90% na predição da área basal. Para o volume, o REMQ% manteve-se inferior a 10%, com valores de  $R^2$  e  $r$  superiores a 90%, evidenciando alto desempenho preditivo.

Na Tabela 8 (Apêndice), observa-se que os valores de  $p$  foram inferiores a 0,01 para todos os índices espectrais (IE), indicando significância estatística dos coeficientes estimados no modelo. Em relação aos índices de textura (IT), apenas as métricas Entropy, CP, HC, IDM, CS e Inertia não apresentaram valores de  $p$  significativos. Os demais parâmetros apresentados na Tabela 11 mostraram significância estatística no modelo. Além disso, conforme ilustrado na

Figura 18, a incorporação dos índices espectrais e de textura não resultou em diferenças expressivas em relação ao modelo de Clutter tradicional, indicando desempenho semelhante entre as abordagens.

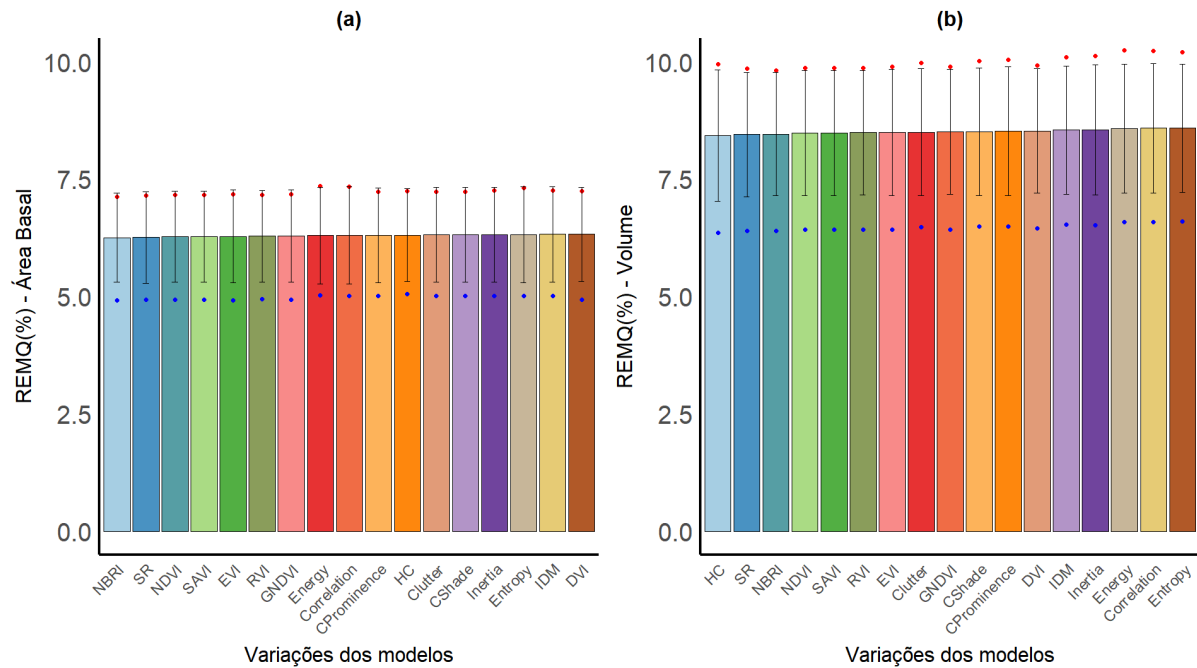


Figura 18 – Gráfico de valores máximos, médios e mínimos da raiz quadrada do erro médio quadrático para as variáveis área basal (a) e volume (b) estimados pelo modelo de Clutter em sua estrutura tradicional e com a inserção de índices espectrais e texturais.

### 1.9.3 MODELOS DE APRENDIZADO DE MÁQUINA

A Figura 19 apresenta o gráfico de REMQ% dos modelos de aprendizagem de máquina, utilizando apenas as variáveis do modelo de Clutter com inserção dos IE/ITs. A Figura 21 apresenta o mapa de erro dos volumes preditos pelo melhor modelo de aprendizagem de máquina, que foi o SVM, utilizando os hiperparâmetros:  $\sigma = 0,25$ ;  $C = 4$ , e a figura 22 é apresentado o erro em modulo do modelo em relação a idade dos talhões.

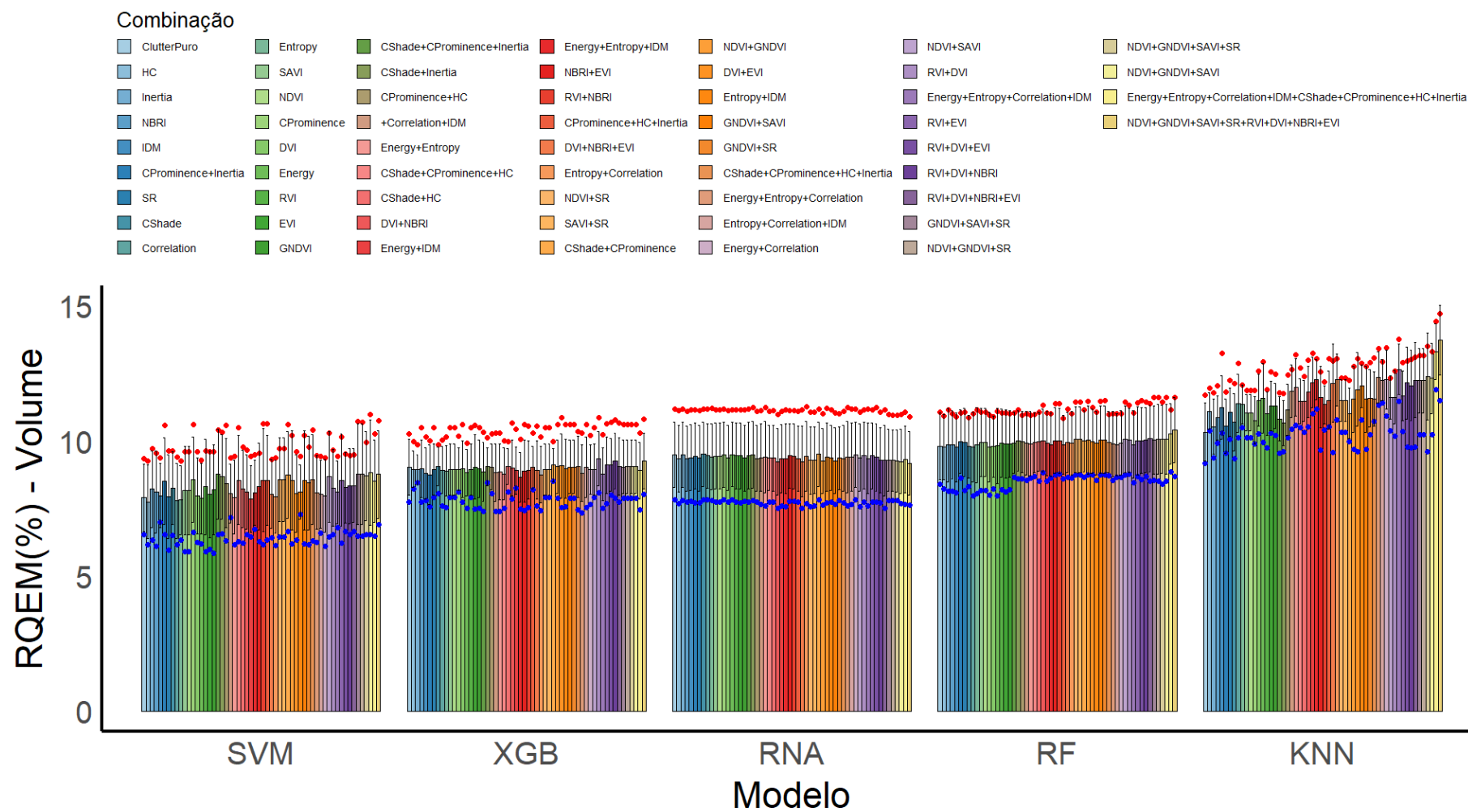


Figura 19 – Gráfico de valores máximos, médios e mínimos da raiz quadrada do erro médio quadrático para o Volume.

Observa-se na Figura 20 que os modelos de AM somente com as variáveis usuais do modelo de Clutter, ou seja, mesmo sem a inclusão de variáveis de sensoriamento remoto, já obtiveram uma melhora na acurácia da prognose. A figura apresenta o desempenho dos modelos de aprendizado de máquina SVM, XGBoost, RNA, RF e KNN, avaliados por meio do REMQ% do volume, considerando diferentes combinações de variáveis dendrométricas, índices espectrais e índices de textura. As barras representam as distribuições do erro associadas a cada combinação, enquanto os pontos indicam medidas centrais e dispersão. Observa-se que os modelos apresentam desempenhos distintos, evidenciando diferenças na capacidade de capturar relações complexas entre as variáveis explicativas e o volume florestal.

Dentro os modelos testados, o SVM apresentou os menores valores médios de REMQ, situando-se predominantemente entre aproximadamente 6% e 9%, indicando elevada capacidade de generalização. Esse desempenho sugere que o SVM foi eficiente em modelar relações não lineares entre variáveis dendrométricas e espectrais, mesmo em um contexto de alta dimensionalidade. XGBoost e RNA apresentaram desempenhos intermediários, com valores de REMQ levemente superiores aos do SVM, porém ainda relativamente estáveis entre as diferentes combinações de variáveis. Isso indica boa robustez, embora com menor sensibilidade a ganhos incrementais decorrentes da inclusão de múltiplos IE e IT. Random Forest apresentou REMQs ligeiramente mais elevados e maior variabilidade, o que pode estar associado à sua sensibilidade à correlação entre variáveis e à presença de ruído, especialmente quando muitas combinações espectrais e texturais são incluídas simultaneamente. KNN apresentou os maiores valores de REMQ e maior dispersão, evidenciando menor capacidade de generalização. Esse comportamento é consistente com a natureza do algoritmo, que é fortemente dependente da distribuição local dos dados e sensível à dimensionalidade do espaço de atributos.

## DISCUSSÕES

A melhora da capacidade preditiva dos modelos de aprendizado de máquinas em relação ao sistema de equações usual de Clutter, ainda que discreta, sugere que estes modelos tem potencial de melhor captar relações complexas entre as variáveis. O sistema de equações de Clutter, sendo um sistema de regressão linear, presume que a forma da relação entre as variáveis resposta e explicativas seja explicitada no ajuste ao passo os modelos de aprendizado de máquinas compõem esta forma de relação entre variáveis ao longo do treinamento

Em relação à adição dos índices espectrais e de textura ao modelo de Clutter e aos modelos de AM, apesar dos ganhos observados, a magnitude das melhorias também foi sutil. Quando as variáveis espectrais foram inseridas empiricamente no modelo de Clutter, as variações na capacidade preditiva do modelo foram praticamente imperceptíveis, com diferenças na segunda casa decimal. Quando as mesmas variáveis foram utilizadas em modelos de aprendizado de máquina, o ganho foi mais perceptível, embora ainda pequeno chegando a 0,7%, o que indica que esses algoritmos capturam melhor, ainda que minimamente, efeito das variáveis espectrais sobre a variável resposta.

Dentre os índices testados, o SR, o NBRI e o HC destacaram-se por apresentarem os menores erros, refletindo diretamente o vigor fisiológico das árvores. Árvores com folhas mais saudáveis realizam fotossíntese de forma mais eficiente, resultando em maior produção de fotoassimilados que é a base fisiológica para o acúmulo de biomassa e volume. O teor de clorofila, por sua vez, é um indicador importante da produtividade florestal, uma vez que regiões com menor concentração de clorofila tendem a apresentar menor crescimento (Azevedo, 2022). A absorção de clorofila ocorre predominantemente no comprimento de onda de 680 nm (faixa do vermelho), enquanto a região entre 690 e 740 nm (borda vermelha e infravermelho próximo) é altamente sensível à presença desse pigmento, constituindo-se em uma métrica relevante para avaliar o estado fisiológico das plantas (Coops, 2015). Assim, o aumento da banda do infravermelho próximo está associado a uma maior área foliar, com maior refletância no NIR e menor no vermelho, o que indica copas mais fechadas, maior eficiência fotossintética e, conseqüentemente, maior acúmulo de biomassa e volume.

Contudo, índices baseados no infravermelho próximo, como o NDVI, apresentam limitação de saturação em florestas densas (Berger et al., 2019). Nesses casos, mesmo com o aumento da biomassa, o índice tende a estabilizar ou crescer mais lentamente, tornando-se menos sensível às variações de vigor em níveis elevados de fechamento de copa. Essa limitação justifica o uso de índices alternativos, como o SR e o NBRI, que são menos suscetíveis à saturação, além da adoção de índices de textura, capazes de representar a homogeneidade estrutural dos talhões. Outra abordagem promissora consiste na utilização de séries históricas de índices espectrais, como o NDVI acumulado, em conjunto com variáveis bioclimáticas. Quando processadas por algoritmos de Floresta Aleatória, essas informações têm demonstrado capacidade de explicar mais de 90% da variabilidade do volume de madeira, além de contribuir significativamente para a estimativa e a contabilização de estoques de carbono em escalas regional e até nacional. (le Maire et al., 2011).

Entretanto, o pequeno ganho marginal observado nas estimativas dos modelos com a inclusão dos índices espectrais e texturais pode estar associado à ausência de georreferenciamento das parcelas de inventário. A utilização de valores médios dos índices por talhão tende a reduzir a variabilidade espacial dos dados, considerando que os talhões florestais apresentam heterogeneidade intrínseca. Assim, a agregação por média pode suprimir diferenças locais relevantes, limitando a capacidade dos modelos em captar variações intra-talhão. Nesse sentido, estudos futuros devem priorizar a associação dos índices espectrais diretamente à área das parcelas amostrais, em detrimento de médias por talhão, a fim de preservar a variabilidade espacial e, consequentemente, aumentar a precisão das estimativas.

Ressalta-se que foram utilizadas imagens do satélite Landsat 8, de acesso gratuito, o que confere viabilidade operacional ao fluxo metodológico proposto para aplicação em empresas florestais. Ademais, a incorporação de sensores com maior resolução espacial e espectral tende a potencializar os resultados, contribuindo para o aprimoramento do desempenho dos modelos preditivos. A resolução espacial e, principalmente, a resolução temporal do Landsat 8 representaram limitações relevantes neste estudo. Para um monitoramento contínuo e eficiente de plantios florestais, é desejável o uso de satélites com revisita mais frequente e maior nível de detalhe espacial. Nesse contexto, o sensoriamento remoto, aliado ao inventário florestal, mostra-se uma ferramenta poderosa, oferecendo uma visão abrangente de grandes áreas, inclusive aquelas de difícil acesso. Contudo, o sensoriamento remoto não substitui o inventário tradicional, mas o complementa, auxiliando na espacialização de variáveis como volume e área basal e aprimorando o planejamento de colheita e manejo. Assim, consolida-se como um componente essencial da silvicultura de precisão, atendendo às demandas crescentes do setor florestal moderno (Santos et al., 2023). **Dedicar um parágrafo a respeito da saturação (“estouro”) dos índices espectrais e texturais. Citar referências que mostram isso. Mencionar a existência de outros índices ou sensores que possam minimizar esta questão.**

Estudos anteriores reforçam a aplicabilidade dessas metodologias. Santos et al. (2023), na região do Vale do Rio Doce, utilizaram RNA e RF para prognose do crescimento de povoamentos de *Eucalyptus grandis* × *E. urophylla* com dados Landsat 7 e obtiveram bons RQEM%, mesmo com redução do esforço amostral, onde variaram de 7,9% a 14,5% para predição contínua e 6,8% a 11,8% para predição. Dos Reis et al. (2019) combinaram dados Landsat 8, RADAR e MDE com *Random forest* para estimar volume de eucaliptos, obtendo RMSE% de 12,88%. Alcântara et al. (2018), ao estimar rendimento volumétrico de eucaliptos

em Minas Gerais com RNA, incorporando variáveis ambientais e edáficas, alcançaram RMSE% de 22,22%. Esses resultados corroboram as evidências encontradas neste estudo quanto à viabilidade de integrar dados de sensoriamento remoto e técnicas de aprendizado de máquina na modelagem florestal.

## CONCLUSÃO

A inclusão de índices espectrais e texturais, especialmente o SR, NBRI e HC, promoveu uma melhora modesta na acurácia das projeções volumétricas. Entretanto, foi na aplicação dos modelos de aprendizado de máquinas que se observou um ganho um pouco mais expressivo de desempenho. As máquinas de vetor de suporte destacaram-se como o melhor modelo, superando os outros modelos testados, tais quais Florestas Aleatórias, K-Vizinho Próximo, redes neurais artificiais e Aumento de Gradiente Extremo.

Em síntese, os achados sugerem que inclusão de variáveis de sensoriamento remoto pode melhorar, ainda que discretamente, a capacidade preditiva dos modelos tradicionais. Os algoritmos de inteligência computacional, como os modelos de aprendizado de máquinas, mostraram-se mais eficientes na incorporação dessas variáveis, indicando um caminho promissor para o aprimoramento dos modelos de crescimento e produção florestal.

## REFERENCIAL TEÓRICO

Alfenas, A. C., Por, U. F. de V., Zauza, E. a. V., Mafia, R. G., & Assis, T. F. de. (2004). *Clonagem e doenças do Eucalipto*. <https://agris.fao.org/search/en/providers/122621/records/6473969f68b4c299a3fb5c2a>

Almeida, M. B. F. de. (2022). *Métodos de sensoriamento remoto orbital para o diagnóstico e monitoramento dos níveis de degradação de pastagens tropicais: Uma contribuição ao plano de agricultura de baixa emissão de carbono*. <https://www.bdt.d.uerj.br:8443/handle/1/20404>

Amorim, D. C. R. (2016). *QUANTIFICAÇÃO DE BIOMASSA E ALTURA DAS ÁRVORES EM UMA FLORESTA DE MATA ATLÂNTICA EM BELO HORIZONTE - MG, UTILIZANDO A TECNOLOGIA LIDAR*.

Araujo, R. M. C., Soares, A. A. V., Silva, A. A. L., Martins, B. F., & Carneiro, M. G. (2023). Modeling of growth and production of Eucalyptus spp stands using supervised machine

learning. *Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*, 1–14. <https://doi.org/10.5753/eniac.2023.232893>

Azevedo, H. C. R. de. (2022). *Modelos locais de produção de látex em seringueira utilizando variáveis dendrométricas e de sensoriamento remoto*. <https://repositorio.ufu.br/handle/123456789/35759>

Baesso, R. C. E., Ribeiro, A., & Silva, M. P. (2010). IMPACTO DAS MUDANÇAS CLIMÁTICAS NA PRODUTIVIDADE DO EUCALIPTO NA REGIÃO NORTE DO ESPÍRITO SANTO E SUL DA BAHIA. *Ciência Florestal*, 20, 335–344. <https://doi.org/10.5902/198050981856>

Bayat, M., Ghorbanpour, M., Zare, R., Jaafari, A., & Thai Pham, B. (2019). Application of artificial neural networks for predicting tree survival and mortality in the Hyrcanian forest of Iran. *Computers and Electronics in Agriculture*, 164, 104929. <https://doi.org/10.1016/j.compag.2019.104929>

Bentes, B. A. H. F. (2018). *Aplicação de redes neurais artificiais para estimativa de volume de árvores em pé em uma floresta ombrófila densa na Amazônia Oriental*. <https://repositorio.ufra.edu.br/jspui/handle/123456789/435>

Berger, R., Silva, J. A. A. da, Ferreira, R. L. C., Candeias, A. L. B., & Rubilar, R. (2019). Índices de vegetação para a estimativa do índice de área foliar em plantios clonais de *Eucalyptus saligna* Smith. *Ciência Florestal*, 29, 885–899. <https://doi.org/10.5902/1980509816942>

Binkley, D., Senock, R., Bird, S., & Cole, T. G. (2003). Twenty years of stand development in pure and mixed stands of *Eucalyptus saligna* and nitrogen-fixing *Facaltaria moluccana*. *Forest Ecology and Management*, 182(1), Artigo 1. [https://doi.org/10.1016/S0378-1127\(03\)00028-8](https://doi.org/10.1016/S0378-1127(03)00028-8)

Binoti, D. H. B., Binoti, M. L. M. da S., & Leite, H. G. (2014). CONFIGURAÇÃO DE REDES NEURAIS ARTIFICIAIS PARA ESTIMAÇÃO DO VOLUME DE ÁRVORES. *Revista Ciência da Madeira (Brazilian Journal of Wood Science)*, 5(1), Artigo 1.

Brandão, M. L. S. M., Iwata, B. de F., Alencar, G. da S., Carvalho, S. P., Almeida, K. de S., Silva, C. M. A. da, Rocha, I. L., & Sousa, M. C. de B. (2020). Biomassa aérea e NDVI em zona ecotonal cerrado-caatinga da Flona de Palmares, Altos, Piauí, Brasil. *Revista Ibero-*

*Americana de Ciências Ambientais*, 11(5), Artigo 5. <https://doi.org/10.6008/CBPC2179-6858.2020.005.0042>

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>

Brink, H., Richards, J., & Fetherolf, M. (2016). *Real-World Machine Learning*. Simon and Schuster.

Broge, N. H., & Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*, 76(2), 156–172. [https://doi.org/10.1016/S0034-4257\(00\)00197-8](https://doi.org/10.1016/S0034-4257(00)00197-8)

Browne, M. W. (2000). Cross-Validation Methods. *Journal of Mathematical Psychology*, 44(1), 108–132. <https://doi.org/10.1006/jmps.1999.1279>

Burkhart, H. E., & Tomé, M. (2012). *Modeling forest trees and stands*. Springer Science & Business Media.

Campos, J. C. C., & Leite, H. G. (2006). *Mensuração florestal: Perguntas e respostas*. UFV.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>

Chen, Z.-Y., Zhang, T.-H., Zhang, R., Zhu, Z.-M., Yang, J., Chen, P.-Y., Ou, C.-Q., & Guo, Y. (2019). Extreme gradient boosting model to estimate PM<sub>2.5</sub> concentrations with missing-filled satellite data in China. *Atmospheric Environment*, 202, 180–189. <https://doi.org/10.1016/j.atmosenv.2019.01.027>

Clutter, J. L. (1963). Compatible Growth and Yield Models for Loblolly Pine. *Forest Science*, 9(3), 354–371. <https://doi.org/10.1093/forestscience/9.3.354>

Coops, N. C. (2015). Characterizing Forest Growth and Productivity Using Remotely Sensed Data. *Current Forestry Reports*, 1(3), 195–205. <https://doi.org/10.1007/s40725-015-0020-x>

Cordeiro, M. A. (2020). *Modelagem do crescimento e produção utilizando máquina de vetor de suporte e redes neurais artificiais*. <http://tede.unicentro.br:8080/jspui/handle/jspui/1312>

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20, 273–297.

Cosenza, D. N., Korhonen, L., Maltamo, M., Packalen, P., Strunk, J. L., Næsset, E., Gobakken, T., Soares, P., & Tomé, M. (2021). *Comparison of linear regression, k-nearest neighbour and random forest methods in airborne laser-scanning-based prediction of growing stock*. <https://dx.doi.org/10.1093/forestry/cpaa034>

Cosenza, D. N., Saarela, S., Strunk, J., Korhonen, L., Maltamo, M., & Packalen, P. (2025). Effects of model-overfit on model-assisted forest inventory in boreal forests with remote sensing data. *Forestry: An International Journal of Forest Research*, 98(4), 507–521. <https://doi.org/10.1093/forestry/cpae055>

Costa Filho, S. V. S. D., Arce, J. E., Montaña, R. A. N. R., & Pelissari, A. L. (2019). Configuração de algoritmos de aprendizado de máquina na modelagem florestal: Um estudo de caso na modelagem da relação hipsométrica. *Ciência Florestal*, 29(4), 1501–1515. <https://doi.org/10.5902/1980509828392>

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>

Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783–2792.

de Azevedo, G. B., Tomiazzi, H. V., Azevedo, G. T. de O. S., Teodoro, L. P. R., Teodoro, P. E., de Souza, M. T. P., Batista, T. S., de Jesus Eufrade-Junior, H., & Guerra, S. P. S. (2020). Multi-volume modeling of Eucalyptus trees using regression and artificial neural networks. *PLoS ONE*, 15(9), e0238703. <https://doi.org/10.1371/journal.pone.0238703>

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120, 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>

*Feature extraction—Orfeo ToolBox 9.1.0 documentation.* ([s.d.]). Recuperado 6 de outubro de 2025, de <https://www.orfeo-toolbox.org/CookBook/recipes/featextract.html?highlight=textur>

Figueiredo, E. O., d'Oliveira, M. V. N., Fearnside, P. M., & Papa, D. de A. (2014). MODELS TO ESTIMATIVE VOLUME OF INDIVIDUAL TREES BY MORPHOMETRY OF CROWNS OBTAINED WITH LIDAR. *CERNE*, 20(4), Artigo 4.

Fiorentin, L. D., Corte, A. P. D., Sanquetta, C. R., & Behling, A. (2015). *QUANTIFICAÇÃO E MODELAGEM DA BIOMASSA E CARBONO DA REGENERAÇÃO NATURAL EM ÁREA DE FLORESTA OMBRÓFILA MISTA.*

García, M. J. L., & Caselles, V. (1991). Mapping burns and natural reforestation using thematic Mapper data. *Geocarto International*, 6(1), 31–37. <https://doi.org/10.1080/10106049109354290>

Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, 58(3), 289–298. [https://doi.org/10.1016/S0034-4257\(96\)00072-7](https://doi.org/10.1016/S0034-4257(96)00072-7)

Gómez, C., Wulder, M. A., Montes, F., & Delgado, J. A. (2012). Modeling Forest Structural Parameters in the Mediterranean Pines of Central Spain using QuickBird-2 Imagery and Classification and Regression Tree Analysis (CART). *Remote Sensing*, 4(1), Artigo 1. <https://doi.org/10.3390/rs4010135>

Gonçalves, A. R. (2010). Máquina de vetores suporte. *Acesso em*, 21. <https://andreric.github.io/files/pdfs/svm.pdf>

Gonçalves, J. L. de M., Alvares, C. A., Higa, A. R., Silva, L. D., Alfenas, A. C., Stahl, J., Ferraz, S. F. de B., Lima, W. de P., Brancalion, P. H. S., Hubner, A., Bouillet, J.-P. D., Laclau, J.-P., Nouvellon, Y., & Epron, D. (2013). Integrating genetic and silvicultural strategies to minimize abiotic and biotic constraints in Brazilian eucalypt plantations. *Forest Ecology and Management*, 301, 6–27. <https://doi.org/10.1016/j.foreco.2012.12.030>

Gower, S. T., Kucharik, C. J., & Norman, J. M. (1999). Direct and indirect estimation of leaf area index, fAPAR, and net primary production of terrestrial ecosystems. *Remote sensing of environment*, 70(1), 29–51.

Goycochea Casas, G. (2021). *Modelagem do crescimento e da produção de povoamentos de eucalipto utilizando regressão, redes neurais artificiais e redes profundas*.

Haralick, R. M., & Shanmugam, K. (1973). Computer Classification of Reservoir Sandstones. *IEEE Transactions on Geoscience Electronics*, 11(4), 171–177. <https://doi.org/10.1109/TGE.1973.294312>

Haralick, R. M., Sternberg, S. R., & Zhuang, X. (1987). Image Analysis Using Mathematical Morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(4), 532–550. <https://doi.org/10.1109/TPAMI.1987.4767941>

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (Vol. 2). Springer.

Haykin, S. (2001). *Redes Neurais: Princípios e Prática*. Bookman Editora.

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3), 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)

IBÁ. (2022). <https://www.iba.org/datafiles/publicacoes/relatorios/relatorio-anual-iba2022-compactado.pdf>

Joibary, S. S. (2013). Forest attributes estimation using aerial laser scanner and TM data. *Forest systems*, 22(3), 484–496.

Jordan, C. F. (1969). Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology*, 50(4), 663–666. <https://doi.org/10.2307/1936256>

Kelsey, K. C., & Neff, J. C. (2014). Estimates of Aboveground Biomass from Texture Analysis of Landsat Imagery. *Remote Sensing*, 6(7), 6407–6422. <https://doi.org/10.3390/rs6076407>

Laclau, J.-P., Gonçalves, J. L. D. M., & Stape, J. L. (2013). Perspectives for the management of eucalypt plantations under biotic and abiotic stresses. *Forest Ecology and Management*, 301, 1–5. <https://doi.org/10.1016/j.foreco.2013.03.007>

Landsberg, J., & Sands, P. (2011). *Physiological ecology of forest production: Principles, processes and models* (Vol. 4). Amsterdam.

Larcher, W. (2003). *Physiological plant ecology: Ecophysiology and stress physiology of functional groups*. Springer Science & Business Media.

le Maire, G., Marsden, C., Nouvellon, Y., Grinand, C., Hakamada, R., Stape, J.-L., & Laclau, J.-P. (2011). MODIS NDVI time-series allow the monitoring of *Eucalyptus* plantation biomass. *Remote Sensing of Environment*, 115(10), 2613–2625. <https://doi.org/10.1016/j.rse.2011.05.017>

Li, W., Radke, J., Liu, D., & Gong, P. (2012). Measuring Detailed Urban Vegetation with Multisource High-Resolution Remote Sensing Imagery for Environmental Design and Planning. *Environment and Planning B: Planning and Design*, 39(3), 566–585. <https://doi.org/10.1068/b37135>

Lima, E. D. S., Souza, Z. M. D., Oliveira, S. R. D. M., Montanari, R., & Farhate, C. V. V. (2022). RANDOM FOREST MODEL TO PREDICT THE HEIGHT OF EUCALYPTUS. *Engenharia Agrícola*, 42(spe), e20210153. <https://doi.org/10.1590/1809-4430-eng.agric.v42nepe20210153/2022>

Liu, H. Q., & Huete, A. (1995). A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Transactions on Geoscience and Remote Sensing*, 33(2), 457–465. <https://doi.org/10.1109/TGRS.1995.8746027>

Lu, D., Chen, Q., Wang, G., Liu, L., Li, G., & Moran, E. (2016). A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. *International Journal of Digital Earth*, 9(1), 63–105.

McRoberts, R. E., Næsset, E., & Gobakken, T. (2015). Optimizing the k-Nearest Neighbors technique for estimating forest aboveground biomass using airborne laser scanning data. *Remote Sensing of Environment*, 163, 13–22. <https://doi.org/10.1016/j.rse.2015.02.026>

Melo, A. M. B. (2019). *Modelos de árvores isoladas e redes neurais para mensuração de crescimento em florestas inequidâneas* [Tese de Doutorado]. Universidade Federal de Viçosa.

Melo, E. de A., Calegario, N., Mendonça, A. R. de, Possato, E. L., Alves, J. de A., & Isaac, M. A. (2017). MODELAGEM NÃO LINEAR DA RELAÇÃO HIPSOMÉTRICA E DO CRESCIMENTO DAS ÁRVORES DOMINANTES E CODOMINANTES DE *Eucalyptus* sp. *Ciência Florestal*, 27, 1325–1338. <https://doi.org/10.5902/1980509829895>

Miguel, E. P., Rezende, A. V., Leal, F. A., Matricardi, E. A. T., Vale, A. T. D., & Pereira, R. S. (2015). Redes neurais artificiais para a modelagem do volume de madeira e biomassa do

cerradão com dados de satélite. *Pesquisa Agropecuária Brasileira*, 50(9), 829–839. <https://doi.org/10.1590/S0100-204X2015000900012>

Niska, H., Skon, J.-P., Packalen, P., Tokola, T., Maltamo, M., & Kolehmainen, M. (2010). Neural Networks for the Prediction of Species-Specific Plot Volumes Using Airborne Laser Scanning and Aerial Photographs. *IEEE Transactions on Geoscience and Remote Sensing*, 48(3), 1076–1085. <https://doi.org/10.1109/TGRS.2009.2029864>

*Pedologia* | IBGE. (2003). <https://www.ibge.gov.br/geociencias/informacoes-ambientais/pedologia/10871-pedologia.html?=&t=downloads>

Pereira, J. C., Dias, P. A. S., Mergulhão, R. C., Thiersch, C. R., & Faria, L. C. (2016). Modelo de crescimento e produção de Clutter adicionado de uma variável latente para predição do volume em um plantio de *Eucalyptus urograndis* com variáveis correlacionadas espacialmente. *Scientia Forestalis*, 44(110). <https://doi.org/10.18671/scifor.v44n110.12>

Ponzoni, F. J., Shimabukuro, Y. E., & Kuplich, T. M. (2015). *Sensoriamento remoto da vegetação*. Oficina de textos.

Pretzsch, H. (2009). *Forest dynamics, growth, and yield: From Measurement to Model*—Springer Berlin Heidelberg.

Richardson, A. J. (1977). *Distinguishing Vegetation from Soil Background Information*. 43(12), 1541–1552.

Rosa, M. R. (2021). *Metodologia de classificação de uso e cobertura da terra para análise de três décadas de ganho e perda anual da cobertura florestal nativa na mata atlântica* [Tese, Universidade de São Paulo]. <https://doi.org/10.11606/T.8.2021.tde-16072021-150114>

Rouse Jr, J. W., Haas, R., & Schell, J. A. (1973). Deering, DW Monitoring vegetation systems in the Great Plains with ERTS. *Earth resources technology satellite-1 symposium*, 3, 307â.

Roy, D. P., Wulder, M. A., Loveland, T. R., C.e., W., Allen, R. G., Anderson, M. C., Helder, D., Irons, J. R., Johnson, D. M., Kennedy, R., Scambos, T. A., Schaaf, C. B., Schott, J. R., Sheng, Y., Vermote, E. F., Belward, A. S., Bindschadler, R., Cohen, W. B., Gao, F., ... Zhu, Z. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145, 154–172. <https://doi.org/10.1016/j.rse.2014.02.001>

Sanquetta, C. R., Corte, A. P. D., Pelissari, A. L., Tomé, M., Maas, G. C. B., & Sanquetta, M. N. I. (2018). DINÂMICA EM SUPERFÍCIE, VOLUME, BIOMASSA E CARBONO NAS FLORESTAS PLANTADAS BRASILEIRAS: 1990-2016. *BIOFIX Scientific Journal*, 3(1), 152. <https://doi.org/10.5380/biofix.v3i1.58384>

Santos, J. S., Mendonça, A. R. de, Gonçalves, F. G., Silva, G. F. da, Almeida, A. Q. de, Carvalho, S. de P. C. e, Silva, J. P. M., Carvalho, R. C., Silva, E. F. da, & Aguiar, M. O. (2023). Predicting eucalyptus plantation growth and yield using Landsat imagery in Minas Gerais, Brazil. *Ecological Informatics*, 75, 102120. <https://doi.org/10.1016/j.ecoinf.2023.102120>

Silva, G. F. da, Curto, R. D. A., Soares, C. P. B., & Piassi, L. de C. (2012). Avaliação de métodos de medição de altura em florestas naturais. *Revista Árvore*, 36, 341–348. <https://doi.org/10.1590/S0100-67622012000200015>

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14, 199–222.

Souza, D. V., Nievola, J. C., Santos, J. X., Wojciechowski, J., Gonçalves, A. L., Corte, A. P. D., & Sanquetta, C. R. (2019). K-Nearest Neighbor Regression in the Estimation of Tectona Grandis Trunk Volume in the State of Pará, Brazil. *Journal of Sustainable Forestry*, 38(8), 755–768. <https://doi.org/10.1080/10549811.2019.1607391>

Stape, J. L., Binkley, D., Ryan, M. G., Fonseca, S., Loos, R. A., Takahashi, E. N., Silva, C. R., Silva, S. R., Hakamada, R. E., Ferreira, J. M. de A., Lima, A. M. N., Gava, J. L., Leite, F. P., Andrade, H. B., Alves, J. M., Silva, G. G. C., & Azevedo, M. R. (2010). The Brazil Eucalyptus Potential Productivity Project: Influence of water, nutrients and stand uniformity on wood production. *Forest Ecology and Management*, 259(9), 1684–1694. <https://doi.org/10.1016/j.foreco.2010.01.012>

Tsakiridis, N. L., Keramaris, K. D., Theocharis, J. B., & Zalidis, G. C. (2020). Simultaneous prediction of soil properties from VNIR-SWIR spectra using a localized multi-channel 1-D convolutional neural network. *Geoderma*, 367, 114208. <https://doi.org/10.1016/j.geoderma.2020.114208>

Vapnik, V. (2013). *The Nature of Statistical Learning Theory*. Springer Science & Business Media.

Velasco Pereira, E. A., Varo Martínez, M. A., Ruiz Gómez, F. J., & Navarro-Cerrillo, R. M. (2023). Temporal Changes in Mediterranean Pine Forest Biomass Using Synergy Models of ALOS PALSAR-Sentinel 1-Landsat 8 Sensors. *Remote Sensing*, 15(13), 3430. <https://doi.org/10.3390/rs15133430>

Viana, H., Lopes, D., & Aranha, J. (2016). Predição de biomassa arbustiva lenhosa empregando dados de inventário e o índice de diferença normalizada extraído em imagens Landsat 5 TM. *Millenium-Journal of Education, Technologies, and Health*, 37, Artigo 37. <https://revistas.rcaap.pt/millenium/article/download/8280/5892>

White, J. C., Coops ,Nicholas C., Wulder ,Michael A., Vastaranta ,Mikko, Hilker ,Thomas, & and Tompalski, P. (2016). Remote Sensing Technologies for Enhancing Forest Inventories: A Review. *Canadian Journal of Remote Sensing*, 42(5), 619–641. <https://doi.org/10.1080/07038992.2016.1207484>

Xu, J., Quackenbush, L. J., Volk, T. A., & Im, J. (2020). Forest and Crop Leaf Area Index Estimation Using Remote Sensing: Research Trends and Future Directions. *Remote Sensing*, 12(18), Artigo 18. <https://doi.org/10.3390/rs12182934>

Zou, M., Jiang, W.-G., Qin, Q.-H., Liu, Y.-C., & Li, M.-L. (2022a). Optimized XGBoost Model with Small Dataset for Predicting Relative Density of Ti-6Al-4V Parts Manufactured by Selective Laser Melting. *Materials*, 15(15), Artigo 15. <https://doi.org/10.3390/ma15155298>

Zou, M., Jiang, W.-G., Qin, Q.-H., Liu, Y.-C., & Li, M.-L. (2022b). Optimized XGBoost Model with Small Dataset for Predicting Relative Density of Ti-6Al-4V Parts Manufactured by Selective Laser Melting. *Materials*, 15(15), Artigo 15. <https://doi.org/10.3390/ma15155298>

## APENDICE

*Tabela 8– Estimativas dos coeficientes da equação área basal do modelo de Clutter para base de dados 100%. Também é apresentado estatística de qualidade da Validação Cruzada.*

Entrada de IE/IT	Coeficientes estimados			Estatística de qualidade			
Modelo	$\alpha 1$	$\alpha 2$	$\alpha 3$	RQEM%	Bias	r	R <sup>2</sup>
Clutter usual	2,284136	0,037946	NA	6,31	0,01	0,92	0,81
Clutter + NDVI	2,030995	0,036274	0,72738	6,26	0,03	0,93	0,86
Clutter + GNDVI	2,048182	0,036611	0,68044	6,25	0,04	0,93	0,86
Clutter + SAVI	2,030996	0,036274	0,48493	6,28	0,03	0,93	0,86
Clutter + SR	1,974079	0,036104	0,14858	6,28	0,04	0,93	0,86
Clutter + RVI	2,598044	0,036379	-0,64526	6,28	0,03	0,93	0,84
Clutter + DVI	2,010441	0,038146	0,00001	6,28	0,08	0,93	0,87
Clutter + NBRI	2,05384	0,035878	0,77260	6,31	0,03	0,93	0,85
Clutter + EVI	2,140198	0,036693	-0,53322	6,29	0,02	0,92	0,83
Clutter + Energy	2,363668	0,037825	-0,25516	6,31	0,01	0,92	0,82
Clutter + Entropy	2,221749	0,037837	0,02824	6,31	-0,01	0,92	0,83
Clutter + Correlation	2,350172	0,037748	-0,03539	6,33	0,01	0,92	0,82
Clutter + IDM	2,322118	0,037911	-0,04238	6,32	-0,01	0,92	0,81
Clutter + Inertia	2,286939	0,037845	0,00076	6,31	0,1	0,95	0,9
Clutter + Cshade	2,288042	0,037774	0,00004	6,30	0,01	0,92	0,81
Clutter + Cprominence	2,305727	0,037912	-0,00005	6,30	0,01	0,92	0,81
Clutter + HC	2,280803	0,037737	0,03335	6,32	-0,02	0,92	0,81

*Tabela 9 – Valores do p-valor dos coeficientes da equação da área basal*

Entrada de IE/IT	Coeficientes estimados		
Modelo	p-valor de $\alpha_1$	p-valor de $\alpha_2$	p-valor de $\alpha_3$
Clutter usual	6,62E-103	1,88E-36	NA
Clutter + NDVI	6,00E-65	6,19E-34	4,79E-05
Clutter + GNDVI	5,08E-62	2,62E-34	5,71E-04
Clutter + SAVI	6,00E-65	6,19E-34	4,79E-05
Clutter + SR	2,86E-60	5,76E-34	3,07E-06
Clutter + RVI	3,03E-78	5,61E-34	1,63E-04
Clutter + DVI	4,91E-47	3,95E-37	3,94E-03
Clutter + NBRI	6,03E-75	1,34E-33	1,24E-06
Clutter + EVI	3,86E-85	8,16E-35	7,16E-05
Clutter + Energy	1,91E-100	1,46E-36	7,54E-03
Clutter + Entropy	4,74E-86	2,55E-36	1,26E-01
Clutter + Correlation	6,34E-101	2,31E-36	1,39E-02
Clutter + IDM	4,75E-17	3,97E-36	8,81E-01
Clutter + Inertia	1,07E-102	4,06E-36	5,81E-01
Clutter + Cshade	6,44E-103	4,89E-36	3,45E-01
Clutter + Cprominence	3,15E-103	1,61E-36	6,04E-02
Clutter + HC	2,09E-102	8,63E-36	4,57E-01

*Tabela 10 – Estimativas dos coeficientes da equação Volume do modelo de Clutter para base de dados 100%. Também é apresentado estatística de qualidade da Validação Cruzada.*

Entrada de IE/IT	Coeficientes estimados				Estatística de qualidade			
Modelo	β0	β1	β2	β3	RQEM%	Bias	r	R²
Clutter Puro	1,5623	-16,6823	0,0269	1,1058	8,51	1,06	0,97	0,93
Cluter + NDVI	1,5600	-16,6544	0,0269	1,1067	8,49	1,02	0,97	0,94
Clutter + GNDVI	1,5631	-16,6924	0,0269	1,1055	8,51	1,02	0,97	0,93
Clutter + SAVI	1,5600	-16,6544	0,0269	1,1067	8,49	1,02	0,97	0,94
Clutter + SR	1,5581	-16,6307	0,0269	1,1074	8,46	0,99	0,97	0,94
Clutter + RVI	1,5606	-16,6612	0,0269	1,1065	8,50	1,03	0,97	0,93
Clutter + DVI	1,5618	-16,6769	0,0269	1,1060	8,53	1,06	0,97	0,93
Clutter + NBRI	1,5566	-16,6115	0,0269	1,1080	8,46	1,01	0,97	0,94
Clutter + EVI	1,5637	-16,7003	0,0270	1,1052	8,50	1,01	0,97	0,93
Clutter + Energy	1,5798	-16,9015	0,0272	1,0988	8,58	1,04	0,97	0,93
Clutter + Entropy	1,5760	-16,8541	0,0271	1,1003	8,59	1,05	0,97	0,93
Clutter + Correlation	1,5786	-16,8872	0,0271	1,0993	8,59	1,05	0,97	0,93
Clutter + IDM	1,5670	-16,7418	0,0270	1,1039	8,55	1,06	0,97	0,93
Clutter + Inertia	1,5627	-16,6878	0,0269	1,1056	8,52	1,07	0,97	0,93
Clutter + Cshade	1,5644	-16,7092	0,0270	1,1049	8,53	1,06	0,97	0,93
Clutter + Cprominence	1,5517	-16,5494	0,0268	1,1100	8,43	1,02	0,97	0,94
Clutter + HC	1,5661	-16,7300	0,0270	1,1043	8,55	1,06	0,97	0,93

*Tabela 9 – Valores do p-valor dos coeficientes da equação do Volume.*

Entrada de IE/IT		Coeficientes estimados		
Modelo	p-valor de $\beta_0$	p-valor de $\beta_1$	p-valor de $\beta_2$	p-valor de $\beta_3$
Clutter usual	2,55E-53	1,91E-31	2,30E-54	4,33E-137
Clutter + NDVI	1,45E-54	3,79E-32	1,68E-55	4,69E-140
Clutter + GNDVI	3,14E-54	5,68E-32	3,69E-55	7,53E-139
Clutter + SAVI	1,45E-54	3,79E-32	1,68E-55	4,69E-140
Clutter + SR	5,00E-55	2,11E-32	6,13E-56	2,49E-141
Clutter + RVI	2,26E-54	4,86E-32	2,54E-55	1,46E-139
Clutter + DVI	3,46E-54	6,11E-32	3,86E-55	5,71E-139
Clutter + NBRI	4,22E-55	1,95E-32	5,00E-56	9,61E-142
Clutter + EVI	1,76E-54	4,04E-32	2,25E-55	2,89E-139
Clutter + Energy	3,40E-54	4,78E-32	6,82E-55	5,27E-136
Clutter + Entropy	1,10E-53	9,86E-32	1,70E-54	1,41E-135
Clutter + Correlation	4,67E-54	5,83E-32	8,70E-55	6,57E-136
Clutter + IDM	2,26E-53	1,68E-31	2,41E-54	2,07E-136
Clutter + Inertia	2,54E-53	1,90E-31	2,33E-54	5,08E-137
Clutter + Cshade	2,44E-53	1,81E-31	2,37E-54	8,95E-137
Clutter + Cprominence	1,70E-53	1,74E-31	1,13E-54	3,10E-139
Clutter + HC	2,40E-53	1,76E-31	2,47E-54	1,64E-136