



GUSTAVO VICTOR ROCHA LEBRON

**RANDOM FOREST APLICADO AO AFILAMENTO DE
ÁRVORES DE EUCALYPTUS**

**LAVRAS – MG
2023**

GUSTAVO VICTOR ROCHA LEBRON

**RANDOM FOREST APLICADO AO AFILAMENTO DE ÁRVORES DE
EUCALYPTUS**

Monografia apresentada à Universidade Federal de Lavras, como parte das exigências do Curso de Engenharia Florestal, para a obtenção do título de Bacharel.

Samuel José Silva Soares da Rocha
Orientador

**LAVRAS – MG
2023**

GUSTAVO VICTOR ROCHA LEBRON

**RANDOM FOREST APLICADO AO AFILAMENTO DE ÁRVORES DE
EUCALYPTUS**

**RANDOM FOREST APPLIED TO THE TRIMING OF EUCALYPTUS
TREES**

Monografia apresentada à Universidade Federal de Lavras, como parte das exigências do Curso de Engenharia Florestal, para a obtenção do título de Bacharel.

APROVADO em 24 de novembro de 2023

Dra. Thiza Falqueto Altoé
Dr. Kalill José Viana da Páscoa

Samuel José Silva Soares da Rocha
Orientador

**LAVRAS – MG
2023**

AGRADECIMENTOS

Gostaria de agradecer a Samuel Jose Silva Soares Da Rocha pela orientação no trabalho, pelo suporte dado a este TCC. Queria agradecer também aos membros da banca avaliadora, os Técnicos Thiza e Kalil que disponibilizaram seu tempo para participar do desfecho deste projeto bem como aos questionamentos feitos e também aos ensinamentos.

A meus pais, Jorge Lebron Ferreira e Luciana Imaculada Rocha, quero agradecer pois foram uma peça fundamental, fornecendo o apoio e sustentação necessários para me manter até o final desta jornada, e a eles sou muito grato.

RESUMO

Este estudo aborda a importância das florestas plantadas no Brasil, destacando seu papel econômico, social e ambiental. Com uma extensão de aproximadamente 9,5 milhões de hectares, o Brasil se destaca globalmente na produção de celulose, papel e outros produtos florestais. O trabalho visa aprimorar a gestão sustentável desses recursos, com foco na estimativa do afilamento das árvores de eucalipto. Para alcançar esse objetivo, utiliza o algoritmo Random Forest (R.F.) em comparação com modelos tradicionais, como polinômio de 5º grau de Schöepfer, função de forma variável de Kozak e função de forma variável de Bi. Os resultados revelam que o Random Forest e o modelo Kozak apresentam desempenho superior na previsão do afilamento, evidenciado por métricas como RMSE e MAE. A análise visual e a importância relativa das variáveis destacam a robustez do Random Forest. A conclusão ressalta a capacidade do Random Forest e do Kozak em prever com precisão o afilamento das árvores, oferecendo insights valiosos para o manejo florestal, enquanto a escolha entre modelos deve considerar o contexto específico de aplicação e conhecimento ecológico.

Palavras-chave: Modelos Estatísticos; Aprendizado de máquina; Afilamento; Modelos de Previsão; Análise de Dados.

ABSTRACT

This study addresses the importance of planted forests in Brazil, highlighting their economic, social and environmental role. With an area of approximately 9.5 million hectares, Brazil stands out globally in the production of cellulose, paper and other forestry products. The work aims to improve the sustainable management of these resources, focusing on estimating the thinning of eucalyptus trees. To achieve this objective, it uses the Random Forest (R.F.) algorithm in comparison with traditional models such as Schöepfer 5th degree polynomial, Kozak variable form function and Bi variable form function. The results reveal that Random Forest and the Kozak model have superior performance in predicting tapering, evidenced by metrics such as RMSE and MAE. Visual analysis and the relative importance of variables highlight the robustness of Random Forest. The conclusion highlights the ability of Random Forest and Kozak to accurately predict tree thinning, offering valuable insights for forest management, while the choice between models must consider the specific context of application and ecological knowledge.

Keywords: Statistical Models; Machine learning; Tapering; Prediction Models; Data Analysis.

LISTA DE FIGURAS

Figura 1: Comparação da dispersão dos resíduos ao longo do fuste.....	20
Figura 2: Análise Comparativa das Estimativas de Diâmetro na Seção i do Fuste em Quatro Modelos	21
Figura 3: Importância das variáveis dos dados no Random Forest.....	22

LISTA DE TABELAS

Tabela 1: Estatística descritiva das 157 árvores cubadas utilizadas para a modelagem do afilamento do fuste.....	15
Tabela 2: Desempenho Comparativo de Modelos: RF, KOZAK, POLI e BI.....	18

SUMÁRIO

1	INTRODUÇÃO	8
2	OBJETIVO.....	10
2.1	Objetivo geral.....	10
2.2	Objetivos específicos.....	10
3	REFERENCIAL TEÓRICO	11
3.1	Afilamento das árvores de eucalipto	11
3.2	Funções de afilamento.....	11
3.3	Aprendizado de máquina e afilamento de árvores	13
3.4	Random Forest	14
4	MATERIAIS E MÉTODOS	15
4.1	Área de estudo e dados utilizados	15
4.2	Seleção dos modelos, avaliação e inferência.....	15
5	RESULTADOS	18
6	DISCUSSÃO.....	23
7	CONCLUSÃO	24
	REFERÊNCIAS	25

1 INTRODUÇÃO

As florestas plantadas no Brasil desempenham um papel crucial no cenário econômico, social e ambiental do país, ocupando uma extensão de cerca de 9,5 milhões de hectares. Com destaque como líder mundial na produção de celulose e papel, bem como na geração de madeira serrada, carvão vegetal e outros produtos florestais, o Brasil se posiciona como um importante ator nesse setor (IBA, 2022). Além de contribuir significativamente para o desenvolvimento econômico, as florestas plantadas desempenham um papel vital na promoção de benefícios sociais e ambientais.

No âmbito econômico, as florestas plantadas geram empregos, incrementam a renda e fortalecem a arrecadação de impostos, conforme relatado pelo Instituto Brasileiro de Árvores (IBA, 2022). Do ponto de vista social, essas florestas desempenham um papel fundamental na conservação da biodiversidade, na melhoria da qualidade do ar e da água, além de atuarem como mitigadoras dos efeitos adversos das mudanças climáticas, como apontado por relatórios da Organização das Nações Unidas para Alimentação e Agricultura (FAO, 2021) e do Ministério do Meio Ambiente (MMA, 2022).

No contexto do desenvolvimento sustentável do setor florestal, a obtenção de multiprodutos a partir das florestas plantadas emerge como uma estratégia estruturante. Esta abordagem não apenas reduz os impactos ambientais associados ao setor, otimizando o uso da madeira e minimizando resíduos, mas também abre novos mercados para os produtos florestais, impulsionando a competitividade do setor (SOARES, 2003).

No entanto, para garantir uma gestão eficiente desses recursos, é essencial compreender a variação do diâmetro do fuste das árvores em função da altura, para a qual as funções de afinamento desempenham um papel crucial. A conicidade do caule, a taxa de diminuição dos diâmetros com o aumento das alturas do nível do solo até a ponta da árvore, é uma expressão quantitativa da forma da árvore. Em geral, uma função de conicidade do caule inclui quatro variáveis da árvore: diâmetro à altura do peito (DAP), altura total da árvore (Ht) e diâmetros do caule dentro ou fora da casca (d) nas alturas (h) acima do solo (BURKHART; TOMÉ, 2012).

Além disso, a taxa na qual o diâmetro do caule diminui para cima depende também das espécies de árvores, idade, condições espaciais, genética das árvores, práticas silviculturais, condições do local e características climáticas (LIU et al., 2020; SAYGILI; KAHRIMAN, 2023).

Modelos matemáticos que descrevem essa variação, são fundamentais para atividades florestais como o cálculo de volume de madeira. Embora os modelos paramétricos tenham se mostrado úteis para produzir previsões confiáveis para diversas espécies, o ajuste de modelos paramétricos pode se tornar um desafio, especialmente para determinar formas de modelo apropriadas e valores de parâmetros iniciais com observações não lineares altamente variáveis. Em algumas circunstâncias (por exemplo, povoamentos significativamente danificados), a suposição subjacente às formas geométricas dos troncos das árvores pode limitar a aplicação dos modelos paramétricos para descrever formas atípicas das árvores (YANG; BURKHART; SEKI, 2023).

Reconhecendo a importância dessa ferramenta na estimativa precisa de variáveis cruciais, como volume de madeira, biomassa e crescimento das árvores, este estudo explora o uso do Random Forest (R.F.) para ajuste de afilamento, visando aprimorar a precisão e confiabilidade dessas estimativas na engenharia florestal (BURKHART; TOMÉ, 2012; FERREIRA, 2004). Este enfoque inovador não apenas fortalecerá o planejamento de manejo florestal, mas também contribuirá para a sustentabilidade a longo prazo do setor florestal no Brasil.

2 OBJETIVO

2.1 Objetivo geral

Este trabalho tem como objetivo principal avaliar a aplicação e o desempenho do algoritmo Random Forest na estimativa do afilamento de árvores de eucalipto, buscando aprimorar a precisão e a confiabilidade das estimativas utilizadas na engenharia florestal.

2.2 Objetivos específicos

- Comparar o desempenho do algoritmo Random Forest com modelos tradicionais de regressão na estimativa do afilamento de árvores de eucalipto.
- Avaliar a precisão das estimativas de afilamento geradas pelo Random Forest em comparação com modelos como o polinômio de 5º grau de Schöepfer, a função de forma variável de Kozak e a função de forma variável de Bi.
- Investigar a consistência e robustez das previsões do Random Forest em diferentes conjuntos de dados, utilizando métricas como Root Mean Square Error (RMSE), Erro Médio Absoluto (MAE) e coeficiente de correlação de Pearson (r_{xy}).
- Analisar visualmente a capacidade de cada modelo em prever a altura do fuste das árvores, considerando a relação estimada entre a altura do fuste e variáveis preditoras distintas.
- Investigar a importância relativa das variáveis utilizadas nos modelos, destacando aquelas que mais contribuem para a precisão das estimativas.

3 REFERENCIAL TEÓRICO

3.1 Afilamento das árvores de eucalipto

A estimação do afilamento de árvores de eucalipto é crucial para a engenharia florestal. Ela fornece informações valiosas para o manejo eficiente e sustentável das plantações. O afilamento, que se refere à variação na forma e diâmetro de uma árvore ao longo do seu tronco, é particularmente relevante para espécies como o eucalipto, que são amplamente cultivadas para diversos fins, incluindo a produção de papel, celulose e madeira (SILVA, 2014).

O conhecimento preciso do afilamento é fundamental para otimizar a produção e a qualidade da madeira, aspectos essenciais na indústria florestal. O eucalipto é conhecido por seu rápido crescimento e versatilidade, mas a qualidade da madeira está diretamente relacionada à sua conformação. Árvores com afilamento inadequado podem resultar em produtos finais de menor qualidade e menor valor comercial (SOUZA, 2017).

Na engenharia florestal, a estimação do afilamento é frequentemente realizada por meio de técnicas como o uso de modelos matemáticos e estatísticos. Esses modelos levam em consideração diversos fatores, como idade da árvore, espaçamento entre plantas, condições de solo e clima, para prever o diâmetro ao longo do tronco. A precisão dessas estimativas é crucial, pois influencia diretamente as decisões tomadas no manejo florestal (SOUZA, 2017).

A tomada de decisões no manejo florestal envolve escolhas estratégicas, como o momento ideal para a colheita, a alocação de recursos e o planejamento de novos plantios. Estimar o afilamento com precisão permite que os gestores florestais tomem decisões informadas sobre quando e como realizar a colheita, visando a maximização da produção de madeira de qualidade (SILVA, 2014).

Além disso, a precisão na estimação do afilamento contribui para a sustentabilidade ambiental. Um manejo florestal bem-sucedido não apenas garante a rentabilidade econômica, mas também preserva ecossistemas e promove a regeneração florestal. Ao entender a variação no diâmetro das árvores, os engenheiros florestais podem implementar práticas de manejo que minimizem o desperdício de recursos e maximizem a eficiência da produção (SILVA, 2014).

3.2 Funções de afilamento

Funções de afilamento são modelos matemáticos que descrevem a variação do diâmetro do fuste de uma árvore ao longo da sua altura. Essas funções são fundamentais para a engenharia

florestal, pois são utilizadas para estimar o volume de madeira, a área de copa e outros parâmetros florestais (SALEKIN et al, 2021).

A escolha do modelo de afilamento apropriado é um passo crucial para a estimativa precisa de parâmetros florestais. No entanto, a seleção do modelo mais adequado pode ser desafiadora, pois existem muitos modelos diferentes disponíveis e cada um deles apresenta suas próprias vantagens e desvantagens (KOZAK; MUNRO; SMITH, 1969; NUNES; GÖRGENS, 2016; SALEKIN et al., 2021; SCOLFORO et al., 2018).

Os modelos de afilamento mais simples são os modelos lineares, que assumem que o diâmetro do fuste varia linearmente ao longo da altura (LARSEN, 2017). No entanto, esses modelos geralmente são imprecisos para árvores de grande porte, pois não levam em consideração a curvatura do fuste (LARSEN, 2017).

Os modelos de afilamento mais complexos são os modelos não lineares, que podem descrever melhor a curvatura do fuste (NICOLETTI et al, 2019). Os modelos não lineares mais comuns são os modelos sigmoidais, que assumem que o diâmetro do fuste varia de acordo com uma função sigmoide (SALEKIN et al, 2021).

Outros modelos não lineares que também são utilizados para descrever o afilamento de árvores incluem os modelos parabólicos, os modelos de potências fracionárias e os modelos de potências inteiras (SALEKIN et al, 2021). O modelo de afilamento mais adequado para uma determinada aplicação depende de vários fatores, incluindo a espécie arbórea, o ambiente de crescimento e a precisão desejada das estimativas (SALEKIN et al, 2021).

Estudos com funções de afilamento têm mostrado que essas funções variam de acordo com a espécie arbórea, o ambiente de crescimento e a idade da árvore (THAMMANU et al, 2021). Em geral, as árvores de espécies de grande porte apresentam um afilamento mais acentuado, enquanto as árvores de espécies de pequeno porte apresentam um afilamento mais suave (THAMMANU et al, 2021). Além disso, os estudos têm mostrado que o afilamento de árvores pode ser afetado por fatores ambientais, como a intensidade da luz solar, a disponibilidade de água e os nutrientes do solo (THAMMANU et al, 2021). A compreensão da variação das funções de afilamento em diferentes condições é importante para o desenvolvimento de modelos mais precisos para a estimativa de parâmetros florestais (THAMMANU et al, 2021).

3.3 Aprendizado de máquina e afilamento de árvores

O aprendizado de máquina, também conhecido como *machine learning*, é uma área da inteligência artificial (IA) e da ciência da computação que se concentra no uso de dados e algoritmos para imitar a maneira como os humanos aprendem, melhorando gradualmente sua precisão (IMB, c2023).

É um método de análise de dados que automatiza a construção de modelos analíticos, sendo treinados para fazer classificações ou previsões, revelando os principais insights em projetos de mineração de dados. Esses insights subsequentemente conduzem a tomada de decisões em aplicativos e negócios, impactando de forma ideal as principais métricas de crescimento (IMB, c2023).

Nos últimos anos, o aprendizado de máquina tem se mostrado uma alternativa promissora aos métodos tradicionais de modelagem. Essa estratégia de modelagem pode fornecer ferramentas para resolver problemas difíceis, como a estimativa e previsão de características do afilamento do tronco das árvores (ÖZÇELİK; DIAMANTOPOULOU; TRINCADO, 2019). Estudos com esse propósito em diferentes regiões e tipos de florestas foram realizados.

No Brasil, estudos avaliando algoritmos de inteligência artificial para descrever o perfil do fuste de árvores em diferentes tipos florestais foram encontrados. Em florestas nativas, modelos de RNA testados apresentaram superioridade em relação a um modelo Random Forest e funções de afilamento tradicionais (NUNES; GÖRGENS, 2016). Em florestas plantadas, algoritmo genético (AG) foram testados em árvores de *Eucalyptus urograndis* em diferentes idades, e os resultados apontaram para a eficiência do método em comparação com abordagens tradicionais (LACERDA et al., 2022).

Na Polônia, estudos baseados em Redes Neurais Artificiais (RNA) para oito espécies características da região, encontraram os melhores resultados em diferentes configurações de RNA em comparação com modelagem tradicional (SOCHA; NETZEL; CYWICKA, 2020). No Chile, estimativas do afilamento usando redes também foram testadas e os resultados foram superiores aos tradicionais modelos de afilamento, como, por exemplo, o modelo de Kozak (SANDOVAL; ACUÑA, 2022). Na Turquia, as RNAs também foram testados para prever o diâmetro do caule de *Pinus sylvestris L.* e apresentaram os melhores resultados para estimar o diâmetro do caule (SAKICI; OZDEMIR, 2018).

Contudo, apesar dos resultados satisfatórios, nota-se que os trabalhos ainda são escassos, sobretudo com a utilização de técnicas baseadas em árvores de decisão. Além disso, nos estudos realizados os autores sugerem que novos trabalhos com diferentes algoritmos e abordagens

devem ser testados a fim de avaliar a capacidade dos métodos nas estimativas do afilamento das árvores.

3.4 Random Forest

O algoritmo Random Forest (RF) é um método de aprendizado de máquina supervisionado notável por sua versatilidade. Ele é capaz de lidar com uma variedade de tarefas, incluindo classificação e regressão. RF usa uma abordagem de 'dividir para conquistar' para melhorar o desempenho. RF constrói centenas de árvores de decisão (portanto, 'floresta') usando subconjuntos aleatórios de variáveis previstas e preditoras. Essas múltiplas árvores são então selecionadas com base em sua variação, a fim de verificar a previsão correta (NUNES; GÖRGENS, 2016).

O funcionamento do Random Forest pode ser descrito em três etapas principais. A primeira etapa envolve a geração de um conjunto de árvores de decisão. Cada árvore é treinada em um subconjunto aleatório dos dados de treinamento. O número de árvores geradas é um parâmetro crucial que influencia tanto a precisão quanto a complexidade do modelo (BREIMAN, 2001).

Na segunda etapa, cada árvore de decisão realiza previsões para novos dados, seguindo sua regra de decisão específica. Esta regra de decisão é uma função que mapeia um conjunto de dados de entrada para um conjunto de dados de saída. Cada árvore contribui com sua previsão única para o resultado final (BREIMAN, 2001).

A terceira etapa consiste na combinação das previsões de todas as árvores para gerar a previsão final do modelo. Normalmente, esse processo de combinação é realizado por meio de votação, onde a classe ou valor mais frequentemente previsto pelas árvores é escolhido como a previsão final. Isso proporciona uma robustez adicional ao modelo, reduzindo o impacto de variações individuais (BREIMAN, 2001).

O Random Forest apresenta várias vantagens. Ele é robusto a ruídos e outliers devido ao treinamento em subconjuntos aleatórios. Sua capacidade de lidar com conjuntos de dados de alta dimensionalidade é outra característica notável, graças à habilidade das árvores de decisão em aprender funções complexas. Além disso, o Random Forest é relativamente fácil de interpretar, pois as regras de decisão individuais das árvores podem ser analisadas (DIDATICA TECH, 2019).

4 MATERIAIS E MÉTODOS

4.1 Área de estudo e dados utilizados

Foram utilizados dados de 157 árvores de *Eucalyptus urophylla* localizados no estado da Bahia, nordeste do Brasil. O espaçamento do plantio foi de 3x2m. As árvores foram cubadas pelo método absoluto. A estatística descritiva dos dados das árvores cubadas são apresentados na Tabela 1 a seguir. Os dados foram disponibilizados por uma empresa do setor de Celulose com operações na região.

Tabela 1: Estatística descritiva das 157 árvores cubadas utilizadas para a modelagem do afilamento do fuste

Estatística	DAP (cm)	HT (m)
Mínimo	8,30	11,40
1 Quartil	16,76	18,80
Mediana	19,73	22,94
Média	19,49	23,09
Terceiro Quartil	22,44	28,10
Máximo	30,24	31,18

Fonte: Autoria própria.

4.2 Seleção dos modelos, avaliação e inferência

Os modelos testados para estimar o afilamento foram: polinômio de 5º grau de Schöepfer (1966), a função de forma variável de Kozak (2004) a função de forma variável de Bi, e o algoritmo de aprendizado de máquina Random Forest. A mais popular função de afilamento utilizada amplamente no mundo foi desenvolvida por Kozak (2004). O autor propôs uma função de afilamento ou “taper” que descreve forma continua o perfil do tronco.

$$d_i = \alpha_0 DAP^{\alpha_1} H^{\alpha_2} X \left[\beta_1 z^4 + \beta(1/\exp(DAP/H)) + \beta_3 X^{0.1} + \beta_4 (1/DAP) + \beta_5 H^q + \beta_6 X \right] + \varepsilon$$

Em que:

d_i = diâmetro comercial (cm) ou diâmetro correspondente a qualquer altura h_i , especificada ou comercial

$\beta_0, \beta_1, \dots, \beta_n$ = Parâmetros do modelo

DAP = diâmetro a 1,3 m de altura (cm)

$$X = \left(1 - \left(\frac{h_i}{H} \right)^{1/3} / (1 - p^{1/3}) \right)$$

h_i = altura comercial (m);

H = altura total (m)

ε = erro de estimativa

O modelo de Schöepfer (1966) é um exemplo clássico de uma função polinomial utilizada para descrever o afilamento do tronco de árvores. A forma geral de um polinômio de 5º grau que poderia ser aplicada ao modelo de Schöepfer é a seguinte:

$$\frac{d_i}{dap} = \beta_0 + \beta_1 \frac{h_i^2}{H^2} + \beta_2 \frac{h_i^3}{H^3} + \beta_3 \frac{h_i^4}{H^4} + \beta_4 \frac{h_i^5}{H^5}$$

Em que:

d_i = diâmetro comercial (cm) ou diâmetro correspondente a qualquer altura h_i , especificada ou comercial

$\beta_0, \beta_1, \dots, \beta_n$ = Parâmetros do modelo

DAP = diâmetro a 1,3 m de altura (cm)

ht = altura total (m)

h_i = altura comercial (m);

O modelo de Bi (2000), assim como outras funções de afilamento, é utilizado para descrever a variação do diâmetro do tronco de árvores ao longo de sua altura. A função proposta por Bi em 2000 é um modelo de afilamento não-linear que leva em consideração variáveis como o diâmetro à altura do peito (DAP), a altura total da árvore (ht), e a altura específica (hi) para estimar o diâmetro (di) em qualquer ponto ao longo do tronco.

$$d_i = dap \left[\frac{\left(\log \sin \left(\frac{\pi}{2} \frac{h_i}{ht} \right) \right)^{\beta_0 + \beta_1 \sin \left(\frac{\pi}{2} \frac{h_i}{ht} \right) + \beta_2 \sin \left(\frac{3\pi}{2} \frac{h_i}{ht} \right) + \beta_3 \sin \left(\frac{\pi}{2} \frac{h_i}{ht} \right) / \frac{h_i}{ht} + \beta_4 dap + \beta_5 \frac{h_i}{ht} \sqrt{dap} + \beta_6 \frac{h_i}{ht} \sqrt{ht}}{\left(\log \sin \left(\frac{\pi}{2} \frac{1,3}{ht} \right) \right)} \right]$$

Em que:

d_i = diâmetro comercial (cm) ou diâmetro correspondente a qualquer altura h_i , especificada ou comercial

DAP = diâmetro a 1,3 m de altura (cm)

h_i = altura comercial (m);

ht = altura total (m)

$\beta_0, \beta_1, \dots, \beta_n$ = Parâmetros do modelo

Os dados foram separados em 70% para treino e 30 % para validação. No treinamento de RF foram configurados três parâmetros essenciais: o número de árvores de regressão aleatória (ntree, testado de 20 a 100 árvores); o número de variáveis de divisão (mtry, utilizado para determinar o número de variáveis disponíveis para cada nó da árvore, sendo o número padrão de 1/3 das variáveis independentes); e o tamanho mínimo dos nós (tamanho do nó, valor = 5).

A cada ajuste/treinamento foram aplicadas as métricas de Root Mean Square Error – RMSE (Equação 4); Erro Médio Absoluto – MAE (Equação 5), coeficiente de correlação de Pearson - $r_{y\hat{y}}$ (Equação 6); Foram calculados o BIAS (Equação 7) e o Viés Relativo (%) (Equação 8).

$$RMSE = \sqrt{\sum_{r=1}^R \frac{\sum_{i=1}^n (X_i - \hat{X}_i)^2}{n}} \quad (4)$$

$$MAE = \sum_{r=1}^R \frac{|\sum_{i=1}^n (X_i - \hat{X}_i)|}{n} \quad (5)$$

$$r_{y\hat{y}} = \frac{cov(y, \hat{y})}{\sqrt{s^2(y) \times s^2(\hat{y})}} \quad (6)$$

$$Bias = \frac{(\sum_i^n y_i - \hat{y}_i)}{n} \quad (7)$$

$$rBias (\%) = \frac{Bias}{\bar{Y}} \times 100 \quad (8)$$

Onde:

n = número de observações;

X_i = variável observada da i -ésima árvore, em m;

\hat{X}_i = variável estimada da i -ésima árvore, em m.

5 RESULTADOS

A tabela apresentada apresenta uma comparação direta entre quatro modelos de previsão do afilamento de árvores testados, utilizando cinco métricas distintas: Erro Médio Quadrático (RMSE), Erro Médio Absoluto (MAE), coeficiente de correlação de Pearson (rxy), Bias (viés médio) e rBias (viés relativo) (Tabela 2).

Tabela 2: Desempenho Comparativo de Modelos: RF, KOZAK, POLI e BI

Modelo	RMSE	MAE	rxy	Bias	rBias
RF	0.68	0.4784792	0.9953076	-0.0102229	-0.0874657
Kozak	0.70	0.4917717	0.9952612	-0.0215276	-0.1823094
Poli	6.28	0.6562938	0.9923086	0.0454503	0.3857956
Bi	1.08	0.7918964	0.9902609	0.2933510	2.4898890

Fonte: Autoria própria.

A partir dos dados, podemos observar que o modelo Random Forest (RF) e o modelo Kozak têm um desempenho muito semelhante em termos de RMSE e MAE, ambos abaixo de 1, o que indica um alto grau de precisão. O RMSE muito baixo de 0.68 para RF, ligeiramente melhor que o 0.70 do Kozak, junto com um rxy extremamente alto próximo de 1 para ambos, sugere que esses modelos têm uma capacidade notável de prever o afilamento das árvores aqui avaliadas.

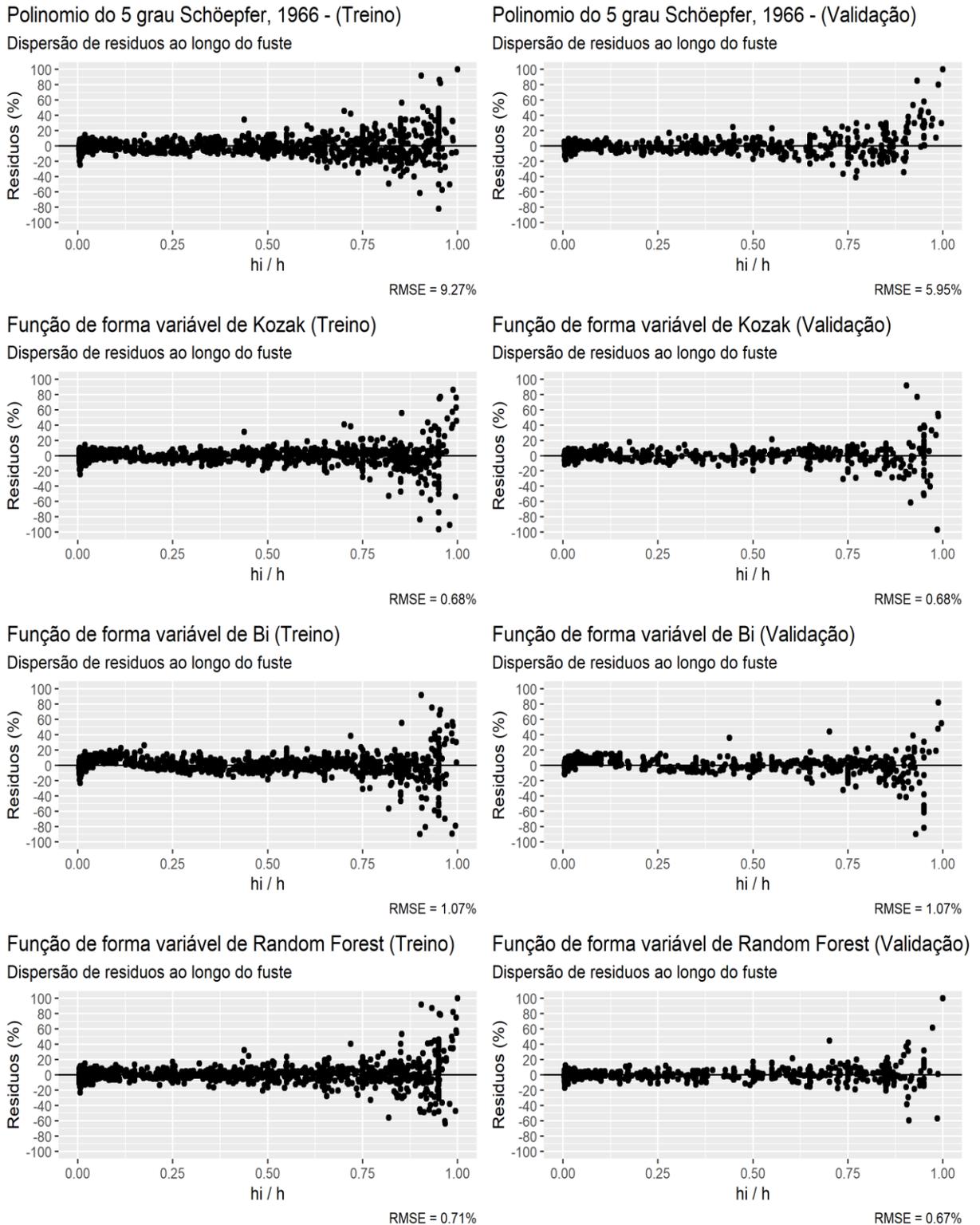
O Bias negativo tanto para RF quanto para Kozak indica que estes modelos, em média, superestimam ligeiramente o crescimento das árvores. No entanto, ambos têm valores de rBias próximos a zero, o que sugere que os erros de viés são relativamente insignificantes em comparação com a magnitude das estimativas, o que é positivo.

Por outro lado, o modelo Polinomial mostra um RMSE significativamente mais alto de 6.28, o que sugere que, embora sua correlação (rxy) ainda seja alta, suas previsões estão, em média, mais distantes dos valores reais. O Bias próximo de zero indica que não há uma tendência clara de subestimação ou superestimação em geral, mas o RMSE alto é motivo de preocupação.

O modelo Bi tem um RMSE mais alto do que RF e Kozak, mas ainda significativamente melhor do que o modelo Polinomial. Contudo, o Bias positivo e o rBias positivo são os maiores entre os quatro modelos, indicando que, em média, o modelo Bi tende a superestimar o crescimento das árvores mais do que qualquer outro modelo.

Os gráficos de resíduos, que são as diferenças percentuais entre os valores observados e os estimados pelo modelo, corroboram os nossos resultados (Figura 1). Todos os modelos parecem ter uma distribuição de resíduos centrada em torno de zero sem padrões claros de viés, o que é um indicativo de bons ajustes. No entanto, o modelo de Kozak e o modelo Random Forest parecem ter um desempenho particularmente bom, com baixos RMSE e uma distribuição de resíduos que não mostra padrões sistemáticos ou tendências, indicando que os erros de predição são principalmente aleatórios, que é o desejável em um modelo de previsão.

Figura 1: Comparação da dispersão dos resíduos ao longo do fuste

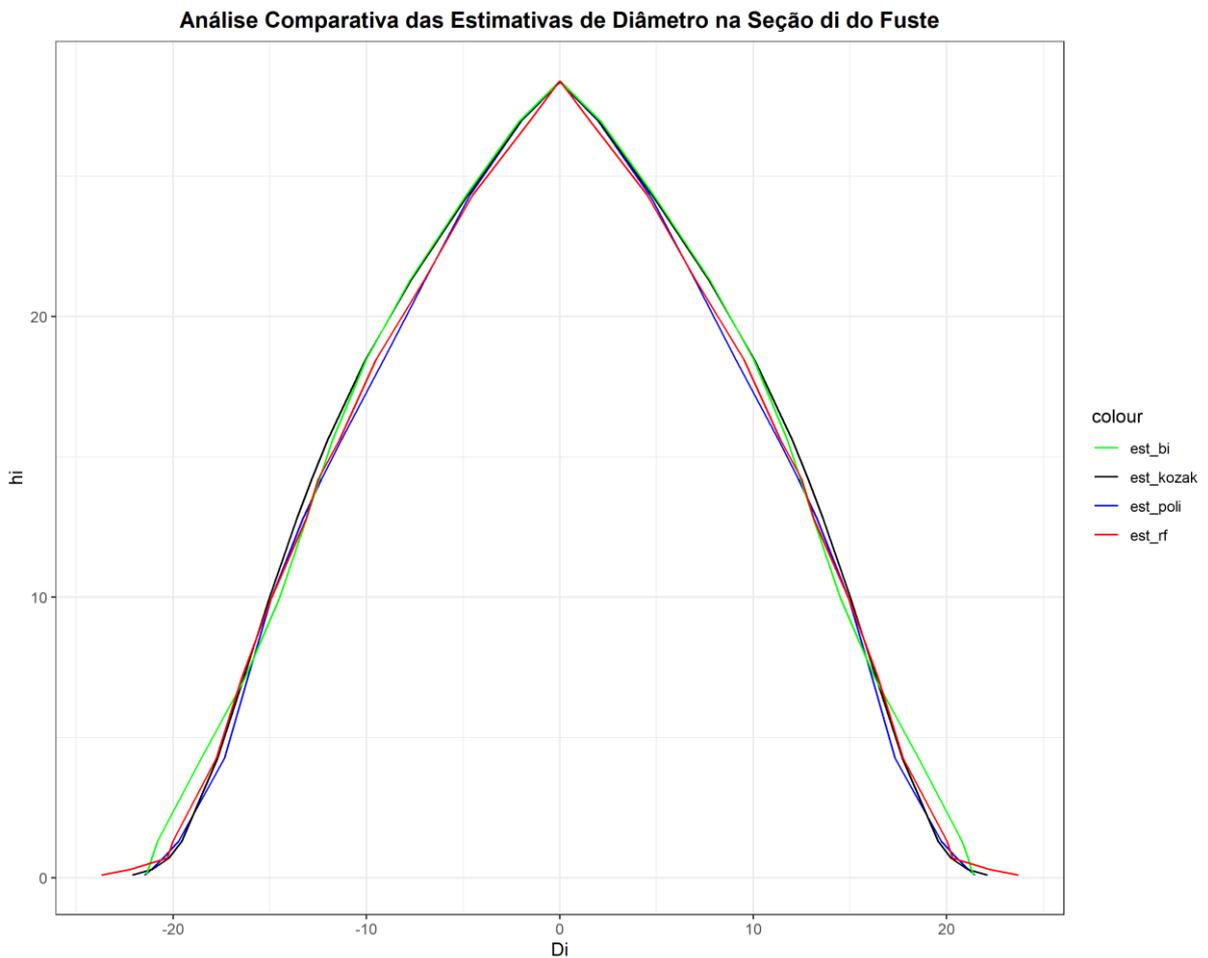


Fonte: Autoria própria

A Figura 2 apresentada é um gráfico de linhas gerado por meio do pacote ggplot2 no software R, representando a relação estimada entre a altura do fuste das árvores (hi) e variáveis predictoras distintas para quatro diferentes modelos de crescimento florestal. Foi selecionada uma árvore com características específicas, denotadas pelos valores DAP = 20.8 e HT = 28.4.

As linhas no gráfico são simétricas com relação ao eixo y, refletindo as relações positivas e negativas das variáveis preditoras com a altura. As cores das linhas foram definidas para representar cada modelo: azul para o modelo Polinomial (est_poli), preto para o modelo Kozak (est_kozak), verde para o modelo Bi (est_bi), e vermelho para o modelo Random Forest (est_rf). Este gráfico oferece uma comparação visual direta da capacidade de cada modelo em prever a altura do fuste, com a simetria das linhas destacando a uniformidade da influência das variáveis preditoras, tanto positivas quanto negativas, nas estimativas de altura produzidas por cada modelo.

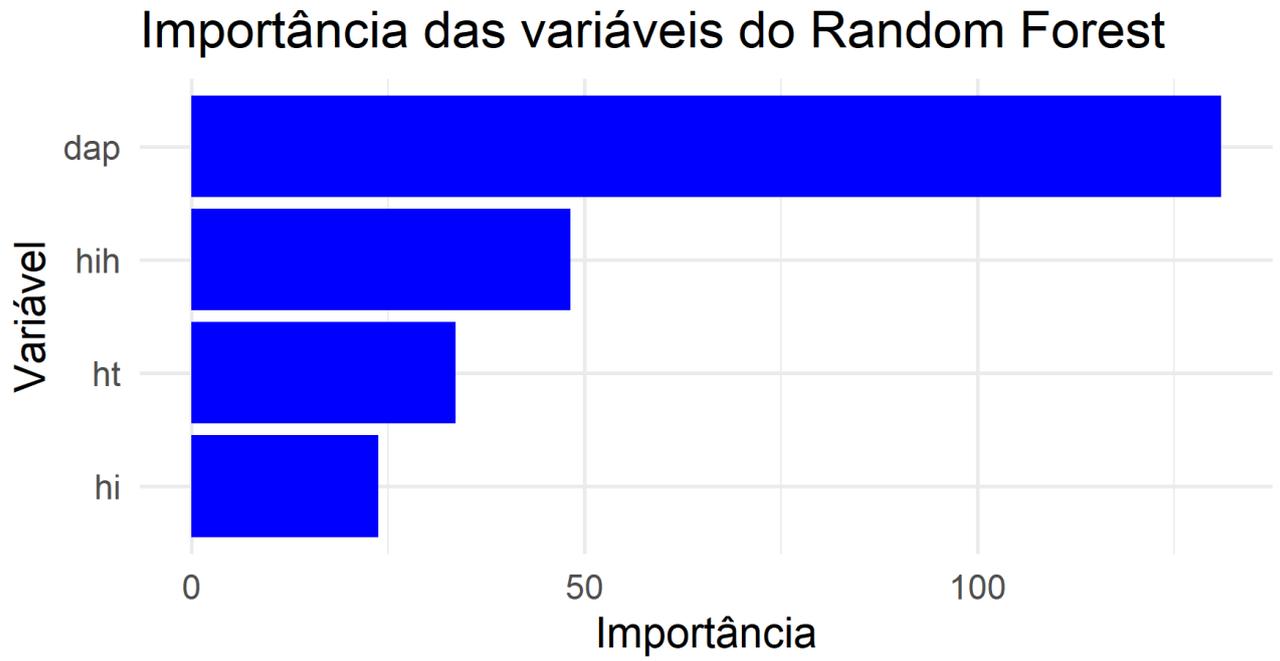
Figura 2: Análise Comparativa das Estimativas de Diâmetro na Seção i do Fuste em Quatro Modelos



Fonte: Autoria própria

A Figura 3 apresenta um relativo entre a importância das variáveis dos dados no Random Forest. De acordo com o apresentado, a variação da variável DAP é a que mais gera influência no resultado do R.F.

Figura 3: Importância das variáveis dos dados no Random Forest



Fonte: Autoria própria

6 DISCUSSÃO

A previsão do afilamento de árvores desempenha um papel crucial em diversas aplicações, especialmente na geração de multiprodutos das florestas (SOARES, 2003). Neste estudo, a avaliação comparativa de quatro modelos distintos - Random Forest (RF), Kozak, Polinomial e BI (2000) - oferece insights valiosos para a escolha criteriosa de modelos em estudos de crescimento florestal.

Os resultados destacam a superioridade dos modelos RF e Kozak em termos de precisão e confiabilidade. A menção da correlação de Pearson próxima entre RF e Kozak sugere não apenas eficiência em capturar tendências, mas também ressalta a capacidade do Kozak de manter uma forte relação linear, apesar de sua menor precisão em termos absolutos.

Ao analisar os modelos Polinomial e BI (2000), é interessante observar que, apesar de apresentarem valores mais altos em RMSE e MAE, indicando um ajuste aparentemente pior aos dados, eles exibem consistência em termos de correlação e viés. Isso aponta para uma confiabilidade na direção da previsão, mesmo que as previsões não estejam tão próximas dos valores reais. De acordo com Perrier (2022), a presença de padrões nos resíduos desses modelos sugere a necessidade de revisão ou melhorias, destacando a importância de avaliar a distribuição dos resíduos para compreender completamente a adequação do modelo.

A análise dos gráficos de resíduos reforça a superioridade do RF, com uma uniformidade na dispersão indicando confiabilidade na previsão da forma do fuste em diferentes alturas. Enquanto o Kozak apresenta um bom ajuste geral, a variabilidade na dispersão dos resíduos sugere possíveis fragilidades em alturas específicas do fuste.

O reconhecimento dos padrões de dispersão nos modelos Polinomial e BI (2000) oferece um entendimento valioso dos pontos fortes e fracos na estimativa. A avaliação da distribuição de resíduos destaca a complexidade da escolha do modelo, o RMSE% é utilizado para medir o desempenho global do ajuste, e quanto menores as estimativas mais confiáveis é o modelo ajustado (CAMPOS; LEITE, 2013), porém é importante destacar que um baixo RMSE não garante a superioridade em todas as circunstâncias. Modelos com RMSE ligeiramente mais alto podem ser mais adequados em determinadas aplicações, dependendo da interpretação da distribuição dos resíduos.

7 CONCLUSÃO

Em conclusão, o nosso estudo revela que o Random Forest e o Kozak são capazes de prever o afilamento das árvores com um alto grau de acurácia, e a manutenção da precisão nos dados de validação aponta para a sua robustez em diferentes conjuntos de dados.

O polinômio de 5º grau de Schöepfer, embora apresente um RMSE maior no treino, melhora significativamente na validação, o que pode indicar uma boa capacidade do modelo de se ajustar a novos dados, apesar de sua simplicidade relativa em comparação com os outros modelos. No entanto, o desempenho superior do Random Forest e do Kozak sugere que a complexidade adicional desses modelos pode capturar melhor a complexidade inerente ao afilamento das árvores.

Para o manejo florestal, a seleção de um modelo deve ser informada não apenas por essas métricas de desempenho, mas também pelo conhecimento ecológico e pelo contexto específico em que as previsões serão aplicadas. A pesquisa futura deve continuar a explorar a aplicação de modelos avançados, como o Random Forest, em conjuntos de dados mais amplos e variados para melhor entender e prever o afilamento das árvores em uma diversidade de ambientes florestais.

REFERÊNCIAS

- ALMEIDA, A. R., et al. **Aplicações de aprendizado de máquina em engenharia florestal: uma revisão de literatura.** *Revista Árvore*, 43(2), 339-350, 2019.
- ALVES, A. P., et al. **Classification of Pinus taeda species using support vector machines.** *Floresta e Ambiente*, 24(2), 220-226, 2017.
- ARAÚJO, C. M., et al. **Detection of leaf spot disease in Eucalyptus trees using support vector machines.** *Floresta e Ambiente*, 25(4), 503-510, 2018.
- BINOTI, M. L. M., & SILVA, M. A. **Redes neurais artificiais para estimativa de volume de árvores de eucalipto.** *Revista Árvore*, 39(3), 685-696, 2015.
- BREIMAN, L. **Random Forests.** *Machine Learning*, 2001.
- CAMPOS, J. C. C., & LEITE, H. G. **Mensuração florestal: perguntas e respostas.** 4. ed. Viçosa: UFV, p. 605, 2013.
- CHAVES, R. G., et al. **Uso de redes neurais artificiais para estimativa de volume e biomassa de árvores de Pinus caribaea em povoamentos plantados no estado do Ceará.** *Pesquisa Agropecuária Brasileira*, 52(10), 1311-1320, 2017.
- DIDATICA.TECH. **O que é e como funciona o algoritmo RandomForest.** Didatica Tech 2019. Disponível em: <https://didatica.tech/o-que-e-e-como-funciona-o-algoritmo-randomforest/>. Acesso em: 25 de nov. de 2023.
- FACELI, J. A., et al. **Configuração de algoritmos de aprendizado de máquina na modelagem florestal.** *Revista Árvore*, 35(5), 1083-1091, 2011.
- FERREIRA, J. A., et al. **Redes neurais artificiais para estimativa de volume de árvores de Pinus taeda em povoamentos plantados no estado de São Paulo.** *Floresta e Ambiente*, 26(4), 665-672, 2019.
- FERREIRA, M. Z. **Estudo de funções de afilamento para representar o perfil e o volume do fuste de Pinus taeda L.** Dissertação de Mestrado em Engenharia Florestal, Universidade Federal de Lavras, Lavras, MG, 2004.
- FOELKEL, C. **Funções de afilamento e fatores de forma para espécies de grande valor comercial da Amazônia Matogrossense.** *Revista Árvore*, 26(2), 213-222, 2002.
- FRIEDMAN, J., HASTIE, T., & TIBSHIRANI, R. **The elements of statistical learning.** New York, NY: Springer, 2001.
- HASTIE, T., TIBSHIRANI, R., & FRIEDMAN, J. H. **The elements of statistical learning (2nd ed.).** New York: Springer, 2009.
- IBM. **O que é machine learning?** IBM, c2023. Disponível em: <https://www.ibm.com/br-pt/topics/machine-learning>. Acesso em: 28 nov. 2023

- JAMES, G., WITTEN, D., HASTIE, T., & TIBSHIRANI, R. **An introduction to statistical learning (2nd ed.)**. New York: Springer, 2013.
- KOZAK, A.; MUNRO, D. D.; SMITH, J. H. G. Taper Functions and their Application in Forest Inventory. **The Forestry Chronicle**, v. 45, n. 4, p. 278–283, ago. 1969.
- KOZAK, J. **A neural network for solving classification and regression problems**. *IEEE Transactions on Neural Networks*, 3(5), 953-965, 1992.
- LACERDA, T. H. S. et al. Feature selection by genetic algorithm in nonlinear taper model. **Canadian Journal of Forest Research**, v. 52, n. 5, p. 769–779, maio 2022.
- LARSEN, D. R. **Simple taper: Taper equations for the field forester**. Proceedings of the 20th Central Hardwood Forest Conference, 2017.
- LIU, Y. et al. Incorporating climate effects in Larix gmelinii improves stem taper models in the Greater Khingan Mountains of Inner Mongolia, northeast China. **Forest Ecology and Management**, v. 464, p. 118065, 15 maio 2020.
- MARTINS, M. F., et al. **Avaliação do potencial de geração de multiprodutos da floresta plantada de eucalipto em Minas Gerais**. *Scientia Agraria*, 23(2), 1-10, 2022.
- MURPHY, K. P. **Machine learning: A probabilistic perspective**. Cambridge, MA: MIT Press, 2012.
- NUNES, M. H.; GÖRGENS, E. B. Artificial Intelligence Procedures for Tree Taper Estimation within a Complex Vegetation Mosaic in Brazil. **PLOS ONE**, v. 11, n. 5, p. e0154738, 17 maio 2016.
- ÖZÇELİK, R.; DIAMANTOPOULOU, M. J.; TRINCADO, G. Evaluation of potential modeling approaches for Scots pine stem diameter prediction in north-eastern Turkey. **Computers and Electronics in Agriculture**, v. 162, p. 773–782, 1 jul. 2019.
- PERRIER, A. **Design effective statistical models to understand your data**. Openclassrooms, 2022. Disponível em; <https://openclassrooms.com/en/courses/5873596-design-effective-statistical-models-to-understand-your-data/6233046-build-and-interpret-a-polynomial-regression-model>. Acesso em: 23 nov. 2023
- QUINLAN, J. R. **C4.5: Programs for machine learning**. San Francisco, CA: Morgan Kaufmann Publishers, 1993.
- SAKICI, O. E.; OZDEMIR, G. STEM TAPER ESTIMATIONS WITH ARTIFICIAL NEURAL NETWORKS FOR MIXED ORIENTAL BEECH AND KAZDAĞI FIR STANDS IN KARABÜK REGION, TURKEY. **CERNE**, v. 24, p. 439–451, dez. 2018.
- SALEKIN, S. et al. **Global Tree Taper Modelling: A Review of Applications, Methods, Functions, and Their Parameters**. *Forests*, 2021. Disponível em: <https://www.mdpi.com/1999-4907/12/7/913/htm>. Acesso em: 25 de nov. de 2023.
- SALEKIN, S. et al. Global Tree Taper Modelling: A Review of Applications, Methods, Functions, and Their Parameters. **Forests**, v. 12, n. 7, p. 913, jul. 2021.

SANDOVAL, S.; ACUÑA, E. Stem Taper Estimation Using Artificial Neural Networks for Nothofagus Trees in Natural Forest. **Forests**, v. 13, n. 12, p. 2143, dez. 2022.

SANTOS, M. A., et al. **Funções de afilamento para povoamento inequiâneo de Eucalipto em Tangará da Serra – MT**. In: Barbosa, R. C., et al. (Orgs.). Engenharia Florestal: desafios, limites e potencialidades. Viçosa, MG: Editora UFV, 2020.

SAYGILI, B.; KAHRIMAN, A. Modeling compatible taper and stem volume of pure Scots pine stands in Northeastern Turkey. **iForest - Biogeosciences and Forestry**, v. 16, n. 1, p. 38, 2023.

SCOLFORO, H. F. et al. Comparison of taper functions applied to eucalypts of varying genetics in Brazil: application and evaluation of the penalized mixed spline approach. **Canadian Journal of Forest Research**, v. 48, n. 5, p. 568–580, maio 2018.

SILVA, A. C., SANTOS, J. R., & SILVA, M. A. **Avaliação de modelos para previsão do afilamento de árvores**. *Floresta*, 53(2), 341-352, 2023.

SILVA, L. C., et al. **Prediction of growth of Pinus caribaea trees using support vector machines**. *Floresta e Ambiente*, 26(4), 673-680, 2019.

SILVA, M. A., et al. **Aprendizado de máquina para a identificação de pragas e doenças florestais: uma revisão de literatura**. *Revista Árvore*, 46(1), 151-162, 2022.

SILVA, P. V. da. **Estimação do volume e afilamento do fuste de árvores de eucalipto utilizando redes neurais artificiais**. Departamento de Engenharia Florestal da Universidade Federal de Viçosa, 2014.

SOARES, T. S.; VALE, A. B. do; LEITE, H. G.; MACHADO, C. C. **Otimização de multiprodutos em povoamentos florestais**. Departamento de Engenharia Florestal da Universidade Federal de Viçosa, 2003.

SOCHA, J.; NETZEL, P.; CYWICKA, D. Stem Taper Approximation by Artificial Neural Network and a Regression Set Models. **Forests**, v. 11, n. 1, p. 79, jan. 2020.

SOUZA, L. A. **Estimação de altura, volume e afilamento de árvores de eucalipto utilizando máquina de vetor de suporte**. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal do Espírito Santo, Jerônimo Monteiro, 2017.

TEIXEIRA, M. A., et al. **Multiprodutos da floresta plantada: uma revisão bibliográfica**. *Revista Árvore*, 42(2), 307-316, 2018.

THAMMANU, S. et al. **The influence of environmental factors on species composition and distribution in a community forest in Northern Thailand**. *Journal of Forestry Research*, 2021.

YANG, S.-I.; BURKHART, H. E.; SEKI, M. Evaluating semi- and nonparametric regression algorithms in quantifying stem taper and volume with alternative test data selection strategies. *Forestry: An International Journal of Forest Research*, v. 96, n. 4, p. 465–480, 1 out. 2023.

ZHANG, Y., LIU, Y., & ZHU, H. **Forest health risk prediction using random forest with Landsat 8 data.** *Remote Sensing*, 14(19), 2872, 2022.

ZHU, S., WANG, X., & LI, X. **Predicting forest stand volume using random forest with multi-source data.** *Forest Ecology and Management*, 486, 118059, 2021.