



FABIANO RODRIGUES PEREIRA

**MODELOS MISTOS GENERALIZADOS E REDES NEURAIIS
ARTIFICIAIS APLICADOS NA ESTIMATIVA DA ALTURA,
AFILAMENTO E VOLUME TOTAL E PARCIAL DE
ÁRVORES DE *Pinus* spp.**

**LAVRAS – MG
2024**

FABIANO RODRIGUES PEREIRA

**MODELOS MISTOS GENERALIZADOS E REDES NEURAS ARTIFICIAIS
APLICADOS NA ESTIMATIVA DA ALTURA, AFILAMENTO E VOLUME TOTAL
E PARCIAL DE ÁRVORES DE *Pinus* spp.**

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

Prof. Dr. Natalino Calegario
Orientador

**LAVRAS – MG
2024**

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

Pereira, Fabiano Rodrigues.

Modelos mistos generalizados e redes neurais artificiais aplicados na estimativa da altura, afilamento e volume total e parcial de árvores de *Pinus* spp. / Fabiano Rodrigues Pereira. - 2024.

176 p. : il.

Orientador(a): Natalino Calegario.

Dissertação (mestrado acadêmico) - Universidade Federal de Lavras, 2024.

Bibliografia.

1. Modelagem florestal. 2. Modelos mistos. 3. Aprendizagem de máquina. I. Calegario, Natalino. II. Título.

FABIANO RODRIGUES PEREIRA

**MODELOS MISTOS GENERALIZADOS E REDES NEURAIAS ARTIFICIAIS
APLICADOS NA ESTIMATIVA DA ALTURA, AFILAMENTO E VOLUME TOTAL
E PARCIAL DE ÁRVORES DE *Pinus spp.***

**GENERALIZED MIXED MODELS AND MACHINE LEARNING APPLIED TO
ESTIMATE THE HEIGHT, TAPERING AND TOTAL AND PARTIAL VOLUME OF
*Pinus spp.***

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

APROVADA em 26 de março de 2024.

Prof. Dr. Natalino Calegario UFLA

Profa. Dra. Marcela De Castro Nunes Santos Terra UFSJ

Prof. Dr. Valdir Carlos Lima de Andrade UFT

Prof. Dr. Natalino Calegario
Orientador

**LAVRAS – MG
2024**

*Ao meu pai Isac pelo amor, constante apoio, por ter me ensinado a perseverar diante das
dificuldades e pelo exemplo de como encarar a vida.
À minha mãe Neuza pelo amor e constante apoio e incentivo pela busca de conhecimento e
por ser meu exemplo de vida.
À minha namorada Thaís pelo constante apoio e carinho nessa jornada e por sempre estar ao
meu lado.
Dedico.*

AGRADECIMENTOS

Agradeço primeiramente a Deus pela vida e todas as bênçãos proporcionadas. Aos meus pais, Isac Pereira da Silva e Neuza Rodrigues da Silva, pelo incentivo constante, investimento em minha educação e por todos os sacrifícios que fizeram para que eu pudesse chegar até aqui. Sempre amarei vocês. Agradeço imensamente à minha melhor amiga e namorada, Thaís Chaves Almeida, pelo apoio constante, torcida, companheirismo e por sempre me incentivar a continuar em busca dos meus sonhos. Aos amigos que fiz em Lavras, Sueliane Santos, Rebecca Garcia, Bruno Corrêa e Maria Kely, pelos conselhos, risadas e companheirismo. Muito obrigado pelo acolhimento. Com vocês, essa jornada se tornou muito mais leve e alegre. Ao Dr. Natalino Calegario, agradeço imensamente por ter me recebido na Pós-graduação em Engenharia Florestal, pelo conhecimento transmitido sobre estatística e mensuração florestal e pela confiança. O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 88887.669728/2022-00, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) e Universidade Federal de Lavras (UFLA). Meus agradecimentos aos órgãos citados por fomentarem o desenvolvimento desta pesquisa e o meu aperfeiçoamento profissional. Agradeço à Dra. Marcela De Castro Nunes Santos Terra pelos ensinamentos sobre pesquisa científica e disponibilidade em me ajudar e aconselhar. Ao Dr. Samuel José Silva Soares da Rocha pela confiança em meu trabalho e por proporcionar oportunidades de crescimento profissional. Ao Dr. Jose Marcio De Mello pelos ensinamentos sobre inventário florestal e pelos conselhos. Ao Dr. Lucas Rezende Gomide pelos ensinamentos e dicas valiosas de programação. À Chica por ter me recebido e auxiliado sempre que necessário. À Pós-graduação em Engenharia Florestal da Universidade Federal de Lavras e a todos os professores. Patrícia Chaves e Lucinei Almeida por me acolherem como um filho e pelo imenso apoio. À minha amiga Iana Delmare por sempre me apoiar, incentivar e torcer por mim. Valéria Guimarães pela amizade, torcida, conselhos valiosos e compreensão de sempre. Érica Rocha pela amizade de sempre e apoio. Aos meus amigos Igor Borges, Yago Santana e Camilo Santana pela torcida de sempre e parceria. Agradeço à Sabrina Cangussu por, lá no início, ter me incentivado a prestar o vestibular que me abriu as portas para a Engenharia Florestal. Thalles Loiola pela amizade, conselhos e parceria de laboratório. Severino Macô e Sérgio Bila pela amizade e trocas de conhecimentos na Pós-graduação. Ana Beatris Moura pela amizade e parceria de laboratório. Dra. Patrícia Anjos Bittencourt Barreto-Garcia por ter me incentivado na pesquisa científica desde a

graduação e por todas as oportunidades e conselhos que me inspiraram a seguir o caminho da pesquisa. E a todos amigos e colegas que, de alguma forma, me ajudaram a chegar até aqui.

“Se a educação sozinha não transforma a sociedade, sem ela tampouco a sociedade muda.”
(Paulo Freire)

RESUMO

A preocupação com o planejamento, ordenamento e a utilização da madeira exigem cada vez mais uma maior precisão. Além disso, as necessidades crescentes da predição dos multiprodutos das florestas têm corroborado para o emprego de diversas técnicas de modelagem para descrever o perfil dos fustes. Essas técnicas compreendem o uso de modelos mistos e o uso de aprendizagem de máquinas, tendo proporcionado ganhos de precisão. Dessa forma, este trabalho avaliou, além de modelos clássicos de regressão locais e modelos genéricos, a aplicação de métodos alternativos, como Modelos Mistos e Redes Neurais Artificiais em diferentes etapas da mensuração florestal (predição da altura e modelagem do afilamento do fuste). Os dados utilizados no presente estudo foram obtidos em plantios de *Pinus* spp., situados no município de Nova Ponte, Minas Gerais. Foram testadas diferentes configurações dos métodos e, a partir de medidas estatísticas de qualidade das predições, foi verificado o ganho de precisão com o uso das técnicas propostas. Para a avaliação das predições realizadas pelos diferentes métodos, em cada etapa, e seleção dos melhores modelos, foram considerados os critérios: raiz do erro quadrático médio (RMSE), erro médio absoluto percentual (MAPE), índice de eficiência do modelo (EF), desvio médio (BIAS), Critério de informação de Akaike (AIC), Critério Bayesiano de Schwarz (BIC), teste da razão de verossimilhança (MLRT), teste t pareado, teste de Wilcoxon *signed rank* pareado e análises gráficas das distribuições e histogramas dos resíduos. Todas as análises foram realizadas por meio da linguagem de programação R. A equação de Ratkowsky (1990) destacou-se como a mais adequada para explicar a interação entre altura e diâmetro das árvores de *Pinus caribaea* var. *hondurensis*. A inclusão do diâmetro dominante e altura dominante resultou em melhorias significativas nas predições da altura total das árvores em diferentes sítios e com diferentes idades. A introdução de efeitos aleatórios e calibração por unidade amostral demonstrou grande eficácia na melhoria dos modelos hipsométricos. Além disso, as alternativas de calibrações mais adequadas foram identificadas, sendo necessário nove árvores distribuídas igualmente em diferentes classes diamétricas para o modelo local na forma mista e, para o modelo que já possuía variáveis a nível de parcela, a calibração não foi necessária. As Redes Neurais Artificiais demonstraram pouca capacidade de generalização, resultando em predições viesadas para a base de dados de teste. Considerando a modelagem do afilamento do fuste, a modelagem com efeitos aleatórios e a função de autocorrelação contínua de primeira ordem proporcionaram melhorias significativas ao modelo de Kozak (2004), tornando os modelos mais coerentes com os dados, resultando em valores significamente iguais aos reais para *Pinus caribaea* var. *caribaea* e *Pinus oocarpa*. As Redes Neurais Artificiais demonstraram eficiência na predição de diâmetro ao longo do fuste e volume total para *Pinus caribaea* var. *caribaea* e *Pinus caribaea* var. *hondurensis*, gerando predições estatisticamente iguais aos valores observados de forma simultânea para as duas espécies por meio de variáveis *dummy*. Os resultados encontrados neste trabalho evidenciam a capacidade da modelagem mista e das Redes Neurais Artificiais proporcionarem ganhos significativos de precisão nas predições de atributos florestais. Além disso, demonstrou-se a necessidade de estudos mais aprofundados sobre a otimização de estratégias para a seleção da estrutura mais apropriada das Redes Neurais Artificiais e interpretação de seus parâmetros ajustados. Também, recomenda-se a adição de outros atributos a nível de parcela, a fim de melhorar a representação do afilamento dos fustes das árvores, desde que a coleta destas variáveis não comprometam a viabilidade das operações de inventário florestal.

Palavras-chave: Modelos mistos; Aprendizagem de máquina; Relação hipsométrica; Volumetria florestal; Afilamento do fuste; Modelagem florestal.

ABSTRACT

Concerning planning, ordering, and the use of wood increasingly requires greater precision. Furthermore, the growing need for predicting forest multiproduct has supported the use of various modeling techniques to describe the profile of boles. These techniques include the use of mixed effect models and the use of machine learning, having provided gains in accuracy. Therefore, this work evaluated, in addition to classical local regression models and generic models, the application of alternative methods, such as mixed models and artificial neural networks in different stages of forest measurement (height prediction and stem taper modeling). The data used in the present study were obtained from *Pinus* spp. plantations, located in the municipality of Nova Ponte, Minas Gerais. Different configurations of the methods were tested and based on statistical measurements of the quality of the predictions, the gain in precision was verified with the use of the proposed techniques. To evaluate the predictions made by the different methods, at each stage, and select the best models, the following criteria were considered: root mean square error (RMSE), mean absolute percentage error (MAPE), model efficiency index (EF), mean deviation (BIAS), Akaike Information Criterion (AIC), Schwarz Bayesian Criterion (BIC), likelihood ratio test (MLRT), paired t-test, paired Wilcoxon signed rank test and graphical analysis of residual distributions and histograms. All analyses were carried out using the R programming language. Ratkowsky's equation (1990) stood out as the most appropriate to explain the interaction between height and diameter of *Pinus caribaea* var. *hondurensis*. The inclusion of dominant diameter and dominant height resulted in significant improvements in predictions of the total height of trees at different sites and with different ages. The introduction of random effects and calibration per plot demonstrated great effectiveness in improving hypsometric models. Furthermore, the most appropriate calibration alternatives were identified, requiring nine trees distributed equally in different diameter classes for the local model in mixed form and, for the model that already had plot-level variables, calibration was not necessary. Artificial neural networks demonstrated poor generalization capacity, resulting in biased predictions for the test database. Considering the modeling of the stem tapering, the modeling with random effects and the continuous first-order autocorrelation function provided significant improvements to the Kozak (2004) model, making the models more coherent with the data, resulting in values significantly equal to the real ones for *Pinus caribaea* var. *caribaea* and *Pinus oocarpa*. Artificial Neural Networks demonstrated efficiency in predicting diameter along the stem and total volume for *Pinus caribaea* var. *caribaea* and *Pinus caribaea* var. *hondurensis*, generating predictions statistically equal to the observed values simultaneously for both species using dummy variables. The results found in this work demonstrate the ability of mixed modeling and artificial neural networks to provide significant gains in accuracy in predicting forest attributes. Furthermore, it demonstrated the need for more in-depth studies on the optimization of strategies for selecting the most appropriate structure of artificial neural networks and interpreting their adjusted parameters. Also, it is recommended to add other attributes at the plot level, to improve the representation of the tapering of tree stems, if the collection of these variables does not compromise the viability of forest inventory operations.

Keywords: Mixed models; Machine learning; Hypsometric relationship; Forest volumetry; Stem tapering; Forest modeling.

INDICADORES DE IMPACTO

A precisão otimizada das estimativas dos recursos florestais geradas a partir deste estudo pode contribuir com uma gestão mais eficiente e sustentável das florestas plantadas de *Pinus* spp. e, conseqüentemente, auxiliar na preservação do ambiente e manutenção de ecossistemas equilibrados, uma vez que a utilização adequada de florestas plantadas acarreta em uma redução da pressão sobre a exploração inadequada das florestas naturais. Estes impactos sociais positivos concretos e diretos são capazes de contribuir para as populações que dependem desses recursos. A utilização de modelos mistos e redes neurais artificiais pode representar um grande avanço tecnológico nas etapas de quantificação da biomassa florestal, possuindo um potencial de uso amplo e sendo capazes de impulsionar o uso de tecnologias avançadas, como sensores remotos e sistemas automatizados de coleta de dados, fazendo com que esses processos sejam mais eficientes e atualizados. Dessa forma, estudos detalhados sobre as melhores estratégias de modelagem e a aplicabilidade dessas técnicas são essenciais para sua difusão e aplicação prática no setor florestal, tanto público quanto privado, contribuindo como potenciais impactos tecnológicos positivos. Predições mais precisas também podem ter efeitos econômicos positivos, uma vez que contribui para uma gestão mais eficaz dos recursos florestais, podendo otimizar a produção de madeira e outros produtos derivados, colaborando para o desenvolvimento sustentável da indústria florestal. Além disso, ao implementar tecnologias avançadas, é possível alcançar a redução de custos operacionais com uma alocação mais inteligente de recursos, resultando em potenciais impactos econômicos positivos. O tema abordado no presente estudo também possui impactos culturais potenciais, podendo influenciar a cultura relacionada à gestão florestal, uma vez que, a introdução de abordagens mais avançadas e inovadoras com ênfase na aplicação prática pode influenciar a forma como as comunidades locais compreendem e impactam seus recursos naturais. Todos estes impactos se enquadram nas áreas de Meio Ambiente e Tecnologia e produção e estão alinhados com Objetivos de Desenvolvimento Sustentável (ODS) da Organização das Nações Unidas (ONU), como ODS 13 - Ação Contra a Mudança Global do Clima, ODS 15 - Vida Terrestre, ODS 8 - Trabalho Decente e Crescimento Econômico, ODS 9 - Indústria, Inovação e Infraestrutura, ODS 12 - Consumo e Produção Responsáveis, ODS 11 - Cidades e Comunidades Sustentáveis e ODS 17 - Parcerias e Meios de Implementação.

IMPACT INDICATORS

The optimized accuracy of forest resource estimates generated from this study can contribute to more efficient and sustainable management of *Pinus* spp. plantations, consequently aiding in environmental preservation and the maintenance of balanced ecosystems. Proper utilization of planted forests reduces pressure on the improper exploitation of natural forests. These concrete and direct positive social impacts can benefit populations that depend on these resources. The use of mixed models and artificial neural networks can represent a significant technological advancement in the stages of forest biomass quantification, possessing broad potential for use and capable of driving the adoption of advanced technologies, such as remote sensors and automated data collection systems, making these processes more efficient and up to date. Thus, detailed studies on the best modeling strategies and the applicability of these techniques are essential for their dissemination and practical application in the forestry sector, both public and private, contributing to potential positive technological impacts. More accurate predictions can also have positive economic effects, as they contribute to more effective management of forest resources, optimizing the production of timber and other derived products, thereby supporting the sustainable development of the forestry industry. Furthermore, implementing advanced technologies can achieve operational cost reductions through more intelligent resource allocation, resulting in potential positive economic impacts. The topic addressed in this study also has potential cultural impacts, as the introduction of more advanced and innovative approaches with an emphasis on practical application can influence how local communities understand and impact their natural resources. All these impacts fall within the areas of Environment and Technology and Production and are aligned with the United Nations (UN) Sustainable Development Goals (SDGs), such as SDG 13 - Climate Action, SDG 15 - Life on Land, SDG 8 - Decent Work and Economic Growth, SDG 9 - Industry, Innovation, and Infrastructure, SDG 12 - Responsible Consumption and Production, SDG 11 - Sustainable Cities and Communities, and SDG 17 - Partnerships for the Goals.

SUMÁRIO

PRIMEIRA PARTE	14
1 INTRODUÇÃO.....	14
2 REFERENCIAL TEÓRICO.....	17
2.1 Relação hipsométrica.....	17
2.2 Predição volumétrica	19
2.3 Afilamento do fuste	21
2.4 Modelos mistos.....	23
2.5 Redes neurais artificiais.....	25
REFERÊNCIAS	29
SEGUNDA PARTE – ARTIGOS	44
ARTIGO 1 – Modelagem mista e aprendizagem de máquina para a predição de altura total de árvores de <i>Pinus caribaea</i> var. <i>hondurensis</i>	44
1 INTRODUÇÃO.....	47
2 MATERIAL E MÉTODOS.....	50
2.1 Caracterização da área de estudo.....	50
2.2 Base de dados	50
2.3 Modelos não lineares de efeitos fixos	53
2.4 Modelos não lineares de efeitos fixos genéricos	53
2.5 Modelos não lineares de efeito misto	54
2.6 Modelos não lineares de efeitos mistos genéricos.....	55
2.7 Calibração dos modelos de efeito misto	55
2.8 Redes Neurais Artificiais.....	56
2.9 Avaliação das equações ajustadas	59
3 RESULTADOS	62
3.1 Modelos não lineares de efeitos fixos	62
3.2 Modelos não lineares de efeitos fixos genéricos	65
3.3 Modelos não lineares de efeitos mistos	68
3.4 Calibração dos modelos de efeito misto	70
3.5 Redes Neurais Artificiais.....	73
3.6 Avaliação geral	77
4 DISCUSSÃO.....	81
5 CONCLUSÃO.....	85
REFERÊNCIAS	86
ARTIGO 2 – Modelagem do afilamento do fuste de espécies de <i>Pinus</i> com modelos de efeito misto e redes neurais artificiais	94

1	INTRODUÇÃO.....	97
2	MATERIAL E MÉTODOS.....	99
2.1	Caracterização da área de estudo.....	99
2.2	Base de dados	99
2.3	Funções de afilamento	104
2.4	Modelos de efeitos mistos	105
2.5	Autocorrelação.....	107
2.6	Calibração dos modelos de efeito misto	108
2.7	Redes Neurais Artificiais.....	109
2.8	Avaliação dos modelos	111
2.9	Avaliação da predição do volume total	114
2.10	Avaliação de predição dos métodos	114
3	RESULTADOS	115
3.1	Equações de afilamento de efeitos fixos.....	115
3.2	Modelos de efeito misto	123
3.3	Redes Neurais Artificiais.....	146
3.4	Avaliação geral.....	159
4	DISCUSSÃO.....	161
5	CONCLUSÃO.....	166
	REFERÊNCIAS	168
	TERCEIRA PARTE.....	175
1	CONSIDERAÇÕES FINAIS	175

PRIMEIRA PARTE

1 INTRODUÇÃO

A demanda crescente por produtos de origem florestal, tanto no Brasil como no mundo, vem estimulando a instalação de novos plantios florestais e, conseqüentemente, levando ao surgimento de novos mercados e à expansão dos horizontes para o setor florestal (Bernardi *et al.*, 2021). Atualmente, o setor de base florestal brasileiro destaca-se por sua alta produtividade, tendo apresentado um incremento de 10,2 % ao PIB nacional no período de 2010 a 2020. Além disso, 9,94 milhões de hectares do país são constituídos de plantios renováveis, em que há a predominância de espécies dos gêneros *Eucalyptus* e *Pinus* (IBÁ, 2023).

O setor florestal possui alguns aspectos especiais diferindo-o dos demais, como a dependência com o meio ambiente que, por sua vez, proporciona sérios riscos à produção, demandando uma constante atenção por parte dos profissionais do setor florestal. Além disso, a produção em larga escala com retorno a longo prazo faz com que a responsabilidade dos tomadores de decisões seja maior (Silva *et al.*, 2005). Dessa forma, os diferentes setores da ciência florestal como silvicultura, economia e, principalmente, o manejo devem fornecer informações que auxiliem de forma segura a tomada de decisões (Buongiorno; Gillies, 2003).

Frente à necessidade de informações precisas, o conhecimento sobre os processos e fatores que influenciam a produção florestal é essencial para a regulação da continuidade de suprimentos de matéria prima de qualidade para as fábricas e aumento da competitividade no setor florestal (Silva *et al.*, 2021). Estes dados fundamentam o planejamento das atividades e estratégias econômicas de comercialização e são obtidos por meio do inventário florestal (Campos; Leite, 2013; Oliveira *et al.*, 2016) que, por sua vez, é uma etapa essencial para o gerenciamento dos recursos florestais (Chen *et al.*, 2019) tendo como principais variáveis para medição a altura total e o diâmetro a 1,3 m do solo (DAP) que podem então ser empregadas em modelos de predições de atributos importantes como o volume comercial e total (Clutter *et al.*, 1983; Van Laar; Akça, 2007; Ercanli, 2020).

As medidas de DAP no campo podem ser obtidas de forma rápida, prática e precisa (Sharma; Parton, 2007), enquanto que, a mensuração das alturas requer maior tempo e custo, além de não ser um método de obtenção direta (Raptis *et al.*, 2021). Desse modo, nos inventários florestais, geralmente são medidos todos os DAP's e as alturas totais de apenas algumas árvores por parcela, tornando a predição da altura uma etapa extremamente relevante (Sánchez-González *et al.*, 2007; Andrade; Leite, 2011; Floriano *et al.*, 2006; Özçelik *et al.*,

2018). A predição da altura é frequentemente realizada por meio do uso de modelos que relacionam a altura e o DAP (Peng *et al.*, 2004). Essa relação de regressão da altura sobre o diâmetro é definida como relação hipsométrica (Schimidt, 1977).

De posse das equações hipsométricas e os valores de DAP, estas informações podem ser utilizadas em equações volumétricas, permitindo assim estimar o volume total ou a um determinado diâmetro comercial de interesse (Miguel *et al.*, 2018; Figueiredo *et al.*, 2016). O volume é um dos principais objetivos do inventário florestal, pois é um importante indicador de produção. No entanto, sua obtenção de forma direta envolve um alto custo e grande consumo de tempo (Sousa *et al.*, 2013; Mendonça *et al.* 2015), pois envolve o abate da árvore, de modo que, para sua obtenção, é realizada a amostragem de algumas árvores do povoamento e, por meio de equações volumétricas, estima-se o volume das árvores restantes.

O volume do fuste também pode ser estimado por meio de funções de afilamento que, além disso, proporcionam o uso mais eficiente dos produtos florestais (Nicoletti *et al.*, 2016), possibilitando a predição do diâmetro e volume a qualquer porção do fuste. Dessa forma, essas funções permitem a quantificação de multiprodutos de uma floresta, já que as possibilidades de utilização da madeira dependem do comprimento da tora e dos diâmetros máximo e mínimo (Kozak *et al.* 1969, Clutter *et al.* 1983, Bailey 1994; Návar *et al.*, 2013). O procedimento de segmentação em multiprodutos é denominado de sortimento e é um dos instrumentos mais importantes para o planejamento das florestas de produção, possibilitando a quantificação da matéria-prima com base em suas dimensões conforme os múltiplos usos da madeira (Müller; Finger; Schneider, 2005).

A modelagem de atributos florestais é, tradicionalmente, realizada por meio de modelos de regressão utilizando o método dos Mínimos Quadrados Ordinários (Costa *et al.*, 2021). Esses modelos assumem que os dados são independentes entre si, possuem distribuição normal e homoscedasticidade e que não há erros de medições nas variáveis independentes. No entanto, essas pressuposições são frequentemente violadas, uma vez que, as bases de dados florestais geralmente possuem erros e *outliers* e as relações entre os atributos florestais são, frequentemente, complexas e não lineares (Ashraf *et al.*, 2013; Liu *et al.* 2003; Weiskittel *et al.* 2011). Além disso, os dados coletados em inventários florestais possuem estrutura hierárquica, contendo várias amostras de parcelas do mesmo sítio, várias árvores da mesma parcela e várias medidas da mesma árvore (De-Miguel *et al.*, 2013). Essas medidas repetidas dos indivíduos não são estatisticamente independentes (Soares *et al.* 1995) e as diferenças entre as parcelas e entre as árvores dentro das parcelas geram duas fontes de variabilidade significantes nos dados (Subedi; Sharma 2011).

Por sua vez, os modelos mistos são adequados para lidar com essa variação e analisar dados que possuem autocorrelação, uma vez que permitem modelar de forma explícita a estrutura de covariância. Além de contemplar variáveis fixas (observáveis), os modelos de efeitos mistos (MEM) consideram variáveis aleatórias (não observáveis), possibilitando a inclusão da variabilidade do indivíduo e o ajuste de um modelo mais coerente (Pinheiros; Bates, 2006). Desse modo, a partir desses modelos, torna-se possível a generalização de estruturas de correlações espaço-temporais e variância não constante (Gouveia *et al.*, 2015). Por consequência, nos últimos anos, a aplicação desses modelos tem se tornado cada vez mais expressiva no setor florestal, sendo empregados nos diferentes estudos de mensuração (Calegario *et al.*, 2005; Crecente-Campo *et al.*, 2010; Corral Rivas *et al.*, 2019; Sharma *et al.*, 2019; Ciceu *et al.*, 2020; Fu *et al.*, 2020; Hao *et al.*, 2020).

Além dos modelos mistos, nos últimos anos, as técnicas de aprendizagem de máquinas (AM) tem sido cada vez mais empregada em diversas áreas, incluindo a florestal, onde métodos como Redes Neurais Artificiais (RNA) (Diamantopoulou, 2005; Nunes; Görgens, 2016, Ercanlı, 2020) vem sendo cada vez mais utilizados (LIU *et al.*, 2018). Esta técnica não exige pressuposições sobre a distribuição, tipo e número das variáveis de entrada, possibilitando predições robustas e acuradas mesmo lidando com relações complexas (Tamm; Remm, 2009; Wang *et al.*, 2018).

Devido a possibilidade de se obter informações mais precisas e confiáveis e do aumento da generalização das predições por meio dos modelos de efeitos mistos e RNA, faz-se necessário estudos mais aprofundados aplicando estes métodos nas etapas da mensuração florestal, que por sua vez é fundamental para todo o planejamento das etapas subsequentes no setor florestal.

OBJETIVOS

Objetivo geral

Avaliar a precisão de modelos tradicionais de regressão, modelos mistos e Redes Neurais Artificiais nas estimativas de altura total, diâmetros ao longo do fuste e volume do fuste de árvores de *Pinus* spp.

Objetivos específicos

Avaliar métodos alternativos para as etapas de mensuração (predição de alturas totais, modelagem do afilamento do fuste e predição de volumes) avaliando modelos mistos e redes neurais artificiais.

Comparar as técnicas utilizadas, com diferentes configurações, a partir de medidas estatísticas de qualidade das predições, verificando se há ganho de precisão com o uso de modelos mistos e redes neurais artificiais.

Propor a adoção de técnicas de modelagem mista e redes neurais artificiais na área de Engenharia Florestal como alternativa às técnicas tradicionais de regressão nas modelagens hipsométrica e afilamento do fuste.

2 REFERENCIAL TEÓRICO

2.1 Relação hipsométrica

A demanda constante por estimativas precisas nas empresas florestais justifica as pesquisas que visam aumentar acurácia e precisão dos métodos empregados para estimar o estoque volumétrico das árvores. Estes dados são fundamentais para o planejamento das atividades e estratégias econômicas de comercialização (Campos; Leite, 2013). A estimativa do volume é realizada a partir de informações obtidas por meio do inventário florestal, sendo elas, o diâmetro à altura de 1,30 m do solo (DAP) e a altura total das árvores (Oliveira *et al.*, 2016).

As medidas de DAP no campo podem ser obtidas de forma rápida, fácil e precisa (Sharma; Parton, 2007). Por outro lado, a obtenção das alturas envolve um grande consumo de tempo e custo, além de ser um procedimento menos preciso, principalmente devido ao fato de não ser um método de obtenção direta e aos obstáculos físicos gerados pelo declive e densidade do sub-bosque (Raptis *et al.*, 2021). Dessa forma, nos inventários florestais, tradicionalmente medem-se todos os DAP's e as alturas totais de apenas algumas árvores em cada unidade amostral, fazendo com que a estimação da altura tenha uma grande relevância (Sánchez-González *et al.*, 2007; Andrade; Leite, 2011; Floriano *et al.*, 2006; Özçelik *et al.*, 2018).

Uma forma comum de realizar essa estimação é por meio do uso de modelos que relacionam a altura e o DAP, possibilitando assim a predição das alturas das árvores a partir de medidas diretas de diâmetro (Peng *et al.*, 2004). Altura e DAP possuem uma relação alométrica, o que os tornam fundamentais para a modelagem e a previsão do crescimento e produção das florestas (Mayer 1936; Calama; Montero 2004). Schimidt (1977) define a relação de regressão

da altura sobre o diâmetro em um povoamento a uma determinada idade como relação hipsométrica.

Diversos modelos hipsométricos podem ser encontrados na literatura (Curtis, 1967; Huang *et al.*, 1992; Huang *et al.*, 2000; Moore *et al.*, 1996; Huang 1999; Peng *et al.*, 2001; Jayaraman; Zakrzewski, 2001; Özçelik *et al.*, 2014; Ige *et al.*, 2013; Akinyemi; Smith, 2013). Geralmente esses modelos podem ser classificados como lineares e não lineares (Lei *et al.*, 2009). Devido à facilidade de ajuste e o bom estabelecimento da teoria dos modelos lineares, estes têm sido mais utilizados. No entanto, com o avanço computacional, o ajuste de modelos não lineares atualmente não é mais uma limitação.

Modelos não lineares frequentemente resultam em melhores ajustes hipsométricos quando comparados aos modelos lineares com o mesmo número de parâmetros (Huang *et al.*, 1992). Tal comportamento pode ser explicado pelo fato de que, de forma geral, a relação entre altura e diâmetro tende a uma forma assintótica, com a altura restringida a um limite superior enquanto o diâmetro não (Niklas, 1995).

Embora o diâmetro seja amplamente reconhecido como uma variável essencial na estimação da altura, a sua relação com a altura varia entre estruturas florestais, é influenciada pelas condições do local de crescimento e também não é constante para diferentes idades, mesmo para o mesmo local (Calama; Montero, 2004; Sharma; Zhang, 2004; Curtis, 1967). Esses fatores indicam que variáveis preditoras adicionais são necessárias no desenvolvimento de modelos hipsométricos capazes de abrangerem uma ampla gama de dados provenientes de diferentes locais e idades, ao invés do ajuste individual de relações hipsométricas para cada unidade amostral (Temesgen; Gadow, 2004). Dessa forma, os modelos hipsométricos podem ser classificados em dois tipos: modelos que utilizam apenas o DAP como variável explicativa da altura e modelos que incluem variáveis a nível de local (Mehtätalo *et al.*, 2015). Esses dois tipos de modelos são chamados de modelos hipsométricos local e regional ou genérico, respectivamente (Soares; Tomé, 2002; Paulo *et al.*, 2011).

Espera-se que os modelos locais possuam maior erro (Huang *et al.*, 1992; Fang; Bailey, 1998; Sharma; Breidenbach, 2015) quando comparados com os modelos que possuem variáveis além do DAP, como altura dominante, idade, índice de sítio (Lynch *et al.*, 1999; Wang; Hann 1988), densidade, área basal (Larsen; Hann 1987; Parresol, 1992; Temesgen *et al.*, 2004), índice de espaçamento relativo, diâmetro quadrático médio (Saud *et al.*, 2009; Zakrzewski; Bella, 1988), altura média, DAP e altura das árvores dominantes (Lei *et al.*, 2009; Sánchez-González *et al.*, 2007), composição de espécies, índice de competição (Temesgen *et al.*, 2007) e percentil da distribuição diamétrica (Calama; Montero, 2004). Sendo assim, a inclusão dessas variáveis

pode permitir a diferenciação da altura total de indivíduos de mesmo DAP, porém em locais com diferentes capacidades produtivas (Leite; Andrade, 2003).

Outro método alternativo para trabalhar com dados hierárquicos, representando locais com diferentes características (Ercanli, 2020) é a utilização de modelos de efeitos mistos, que vem sendo cada vez mais utilizados na área florestal, como no ajuste de relações hipsométricas (Mehtätalo, 2004; Trincado *et al.*, 2007; Crecente-Campo, 2010; Castaño-Santamaría *et al.*, 2013; Bronisz; Mehtätalo, 2020; Raptis *et al.*, 2021).

Nos últimos anos o campo da Aprendizagem de Máquinas (AM) também tem visto um grande aumento de aplicações em várias áreas da ciência, incluindo a florestal (Skudnik; Jevšenak, 2022). Dessa forma, além dos métodos estatísticos, técnicas de AM têm sido utilizadas como uma forma inovadora na predição de atributos individuais e populacionais em florestas (Hasenauer *et al.*, 2001; Diamantopoulou, 2005; Özçelik *et al.*, 2008; Iliadis *et al.*, 2013; Binoti, *et al.*, 2016; Chen *et al.*, 2020) e, dentre os diversos algoritmos existentes, as Redes Neurais Artificiais (RNA) tem sido um dos mais utilizados na modelagem florestal (Castaño-Santamaría *et al.*, 2013). Neste contexto, a estimação de altura total de árvores, utilizando RNA (Özçelik *et al.*, 2013; Castro *et al.*, 2013; Thanh *et al.*, 2019) tem mostrado ganhos de precisão significantes.

2.2 Predição volumétrica

O conhecimento do volume de madeira em um plantio florestal é extremamente importante, dado o aumento da valorização da madeira e dos produtos florestais e a demanda do uso social, ecológico e economicamente adequado dos recursos florestais. Sendo assim, a preocupação com o planejamento, ordenamento e a utilização da madeira exigem cada vez mais uma maior precisão (Miguel *et al.*, 2014).

A quantificação do volume de madeira florestal é fundamental para a implementação de planos de manejo sustentável das florestas. Essa quantificação é realizada a partir de um inventário florestal que consiste na mensuração parcial da população, ou seja, de unidades amostrais ou parcelas e, posteriormente, os resultados são extrapolados para a área total de floresta. Dessa forma, estimativas da quantidade e da distribuição da madeira disponível serão obtidas e contribuirão para o planejamento das operações florestais (Leite; Andrade, 2002).

A obtenção do volume de madeira pode ser feita de várias formas, como por meio do xilômetro (deslocamento de água), métodos de cubagem rigorosa como o de Smalian, Huber e Newton (Gomes, 1957) e métodos não destrutivos a partir de equações volumétricas de

regressão. O procedimento de cubagem utilizando o xilômetro é o único que fornece o volume verdadeiro, porém as equações volumétricas são mais utilizadas, uma vez que não exigem o abate das árvores e por serem capazes de fornecer resultados satisfatórios em relação a resíduos e, geralmente, a partir de cálculos relativamente fáceis de serem computados em comparação com outras metodologias (Martins *et al.*, 2016). Além disso, conforme Guimarães e Leite (1996) e Campos *et al.* (1985), a aplicação de equações volumétricas constituem-se em um dos métodos mais eficiente de quantificar a produção volumétrica de modo a reduzir o erro do inventário.

As equações volumétricas permitem estimar o volume total ou a um determinado diâmetro comercial de interesse, que é uma variável de difícil obtenção, a partir de variáveis de mais fácil obtenção como o DAP e a altura das árvores (Miguel *et al.*, 2018; Figueiredo *et al.*, 2016). Os ajustes podem ser lineares ou não lineares, sendo divididos em modelos de simples entrada (DAP) e de dupla entrada (DAP e altura), além de suas diversas combinações (Eder *et al.*, 2016). Além disso, dentre esses métodos, os modelos de dupla entrada, geralmente, apresentam melhores resultados (Thomas *et al.*, 2006; Azevedo *et al.*, 2011; Miguel *et al.*, 2014).

O modelo de Schumacher e Hall (Schumacher; Hall, 1933) tem sido um dos mais utilizados para obtenção de equações ajustadas no Brasil, devido às suas propriedades estatísticas, resultando em estimativas quase sempre não tendenciosas (Miguel *et al.*, 2018; Andrade, 2017; Oliveira *et al.*, 2009; Leite; Andrade, 2002; Campos; Leite, 2006). Este modelo é, frequentemente, ajustado na sua forma aproximada (linearizado por transformação logarítmica) a partir do Método dos Mínimos Quadrados Ordinários (MMQO). Porém, quando o ajuste é realizado pelo Método dos Mínimos Quadrados Não Lineares, os algoritmos Levenberg-Marquardt e Gauss-Newton têm sido os mais utilizados (Silva *et al.*, 2009).

Apesar da eficiência das equações na estimação volumétrica, elas podem apresentar algumas limitações (Martins *et al.*, 2016; Khan *et al.*, 2018; Azevedo *et al.*, 2020), como a necessidade de ajuste conforme as características do povoamento (Gorgens *et al.*, 2009; Binoti; Binoti; Leite, 2014), devido às variações de forma dos fustes. Além disso, os modelos clássicos de regressão baseiam-se nas pressuposições de que as observações são independentes entre si e que há homogeneidade da variância, o que, em determinados casos, não ocorre (Dantas *et al.*, 2020).

Os modelos mistos surgem como alternativa para a análise de dados correlacionados, pois permitem a modelagem explícita da estrutura de covariância, possibilitando assim, generalizar as estruturas de correlações espaço-temporais e a variância não constante (Gouveia

et al., 2015). Estudos sobre a estimativa do volume florestal por meio de modelos mistos têm sido cada vez mais adotados nas últimas décadas desde meados dos anos 90 (Lappi, 1991; Grégoire; Schabenberger, 1996) até os dias mais atuais (Gouveia *et al.*, 2015; Dantas *et al.*, 2020; Cerqueira *et al.*, 2020; Monteiro *et al.*, 2021), uma vez que apresentam um ganho de precisão relevante em relação aos modelos de regressão tradicionais.

Outra alternativa para a predição volumétrica, que dispensa as pressuposições estatísticas dos modelos clássicos de regressão, para representar relações não lineares entre as variáveis é a utilização das técnicas de aprendizagem de máquinas (AM) (Eder *et al.*, 2016). As técnicas de AM disponíveis são diversas e tem recebido grande interesse por parte do setor florestal por possuírem potencial para aplicações que resultem na geração de estimativas com maior acurácia e mais eficientes (Binoti *et al.*, 2016; Hamidi *et al.*, 2021), como as RNA (Binoti *et al.*, 2014; Görgens *et al.*, 2014).

2.3 Afilamento do fuste

A forma do fuste das árvores tem sido objetivo de vários estudos e métodos de expressão na mensuração florestal, de forma que a sua relação com o volume das árvores é a principal motivação. As diferentes taxas de decréscimo do diâmetro da base para o topo do fuste causam variações na forma do fuste. Essa diminuição do diâmetro, conhecida como “*taper*” ou adelgaçamento, que varia de acordo a espécie, DAP e idade das árvores e com a capacidade produtiva do local (sítio) é o principal motivo para a variação no volume. Fatores de forma, quocientes de forma, ponto de forma, tabelas *taper*, curvas e fórmulas podem expressar a forma do fuste, de modo que, em todos os métodos, o objetivo principal é usar as expressões na estimativa do volume da árvore (Jonson, 1940; Kershaw Jr, 2016).

A primeira tentativa de representar a forma das árvores foi realizada por Höjer no início do século XX. Desde então, diversas formas e tipos de modelo do perfil do fuste vem sendo testados, de modo que, no início, os modelos eram relativamente simples, mas, com o avanço computacional, modelos mais complexos foram desenvolvidos para obtenção de funções de afilamento, resultando em estimativas mais realistas (Figueiredo-Filho *et al.*, 1996).

As funções de afilamento, também chamadas de funções de *taper*, funções de adelgaçamento, modelos de perfil ou, ainda, funções de forma, são uma forma de descrever matematicamente o perfil de um fuste. A partir dessas funções, é possível realizar a predição do diâmetro a qualquer altura do fuste e, a integração dessas funções em relação a uma determinada altura fornece o volume contido nessa porção, de modo que, essas funções se

tornem essenciais na quantificação de multiprodutos de uma floresta, uma vez que as alternativas de uso da madeira baseiam-se no comprimento da tora e no diâmetro máximo e mínimo (Kozak *et al.* 1969, Clutter *et al.* 1983, Bailey 1994; Návar *et al.*, 2013).

Vários modelos de regressão têm sido desenvolvidos para a descrição do perfil do fuste, contemplando as variações de forma de cada espécie. O melhor modelo depende dos objetivos desejados para cada caso específico. A variável dependente de uma função de afilamento, comumente, é a razão entre o diâmetro (d) a uma altura conhecida (h) e o DAP. Esses modelos apresentam uma grande diversidade em relação ao grau de complexidade dos ajustes e à qualidade das informações geradas (Assis *et al.*, 2001; Figueiredo; Scolforo; Oliveira, 2006).

Dentre as diversas técnicas estatísticas de modelagem aplicadas à descrição do perfil do fuste, pode-se citar os modelos polinomiais segmentados e não segmentados, considerando formas sigmóides (Kozak *et al.*, 1969; Demaerschalk, 1973; Ormerod, 1973; Max; Burkhart, 1976; Clutter, 1980; Guimarães; Leite, 1992; Andrade; Leite, 1998).

Os modelos não segmentados são aqueles onde uma única função descreve a forma do fuste da base ao topo. Esses modelos possuem a vantagem de modelar o fuste a partir de uma única expressão, resultando numa maior facilidade de ajuste e aplicação (Campos; Leite, 2013). Diferentemente, nos modelos segmentados, ocorre o ajuste de submodelos, de modo que cada um descreve uma porção do fuste e são conectados com a condição de que cada função se conecte adequadamente aos pontos de junção. Com isso, esses modelos incluem uma maior dificuldade de aplicação, porém proporcionam estimativas mais acuradas em relação aos modelos não segmentados (Souza *et al.*, 2008).

As funções de afilamento ainda podem possuir o princípio da compatibilidade entre o volume total e os volumes comerciais estimados. Este princípio permite assegurar que a soma dos volumes comerciais seja sempre inferior ao volume total (Scolforo, 1997). A principal vantagem de se utilizar sistemas compatíveis é que estes fornecem resultados consistentes (Demaerschalk, 1973), uma vez que, a precisão dos modelos de afilamento é função da precisão da equação volumétrica da qual o modelo foi derivado. Ainda, uma função mais eficiente na estimação “*taper*” não é necessariamente mais precisa na estimação do volume (Lima, 1986; Fischer, 2001).

No Brasil, diferentes modelos de afilamento já foram empregados, visando uma maior precisão das estimativas (Schöpfer, 1966; Hradetzky, 1976; Scolforo *et al.*, 1998; Fischer *et al.*, 2001; Soares; Leite; Vale, 2004; Menon, 2006; Pires; Calegario, 2007; Horle *et al.*, 2010; Nicoletti *et al.*, 2016; Da Silva *et al.*, 2018; Téó *et al.*, 2018; Farias *et al.*, 2019; Bernardi *et al.*, 2021). Porém, a proposta de novas metodologias capazes de estimar o estoque volumétrico com

maior precisão ainda é necessária, considerando o custo que essa etapa significa para o inventário florestal (Binoti *et al.*, 2014).

As necessidades crescentes de estimar os multiprodutos das florestas têm corroborado no emprego de diversas técnicas de modelagem para descrever o perfil dos fustes (Souza *et al.*, 2008). Essas técnicas compreendem o uso de modelos mistos (Arias-Rodil *et al.*, 2015; Li *et al.*, 2021) e o uso de aprendizagem de máquinas (AM), como a utilização de RNA (Sakici; Ozdemir, 2018; Socha *et al.*, 2020), tendo proporcionado ganhos de precisão.

2.4 Modelos mistos

Diversos modelos estatísticos podem ser expressados como modelos lineares que possuem ambos efeitos fixos, aqueles associados com toda a população ou com níveis repetidos de fatores experimentais, e efeitos aleatórios, que são aqueles relacionados com unidades experimentais individuais estruturadas de forma aleatória em uma população (Pinheiro; Bates, 2000). Esses modelos são conhecidos como modelos mistos e permitem a possibilidade de se analisar dados correlacionados, uma vez que possuem a capacidade de modelar explicitamente a estrutura de covariância, possibilitando o ajuste de um modelo mais adequado.

A utilização dos modelos de efeitos mistos, permite a generalização de estruturas de correlações espaço-temporais e variância não constante e, além disso, também é possível descrever blocos incompletos, parcelas segmentadas e dados estruturados de forma hierárquica (Gouveia *et al.*, 2015; Pinheiro; Bates, 2000). Dessa forma, nas últimas décadas, o emprego desses modelos vem ganhando cada vez mais relevância em diversas áreas, como física, biologia e ciências sociais (Kuznetsova; Brockhoff; Christensen, 2017).

Modelos mistos podem ser definidos como modelos de regressão que ocorrem em diferentes níveis (Bell *et al.*, 2013). Exemplificando, no estudo de Fu *et al.* (2017), foi avaliada a relação da biomassa acima do solo das árvores com as variáveis independentes DAP e altura, onde estas são medidas em diversos indivíduos (nível 1) e esses indivíduos são frequentemente coletados em parcelas alocadas em diferentes regiões (nível 2). Gregoire e Schabenberger (1996), a partir dos dados de DAP e altura e medidas repetidas de volume cumulativo ao longo do fuste, desenvolveram um modelo não linear de predição em que os parâmetros influenciando a assíntota, taxa de crescimento e forma variam entre árvores dependendo do DAP e da altura.

Modelos lineares mistos

O modelo linear mistos pode ser expresso matricialmente, na forma geral (Pinheiro; Bates, 2000):

$$y = X\beta + Zb + \varepsilon, \quad b \sim N(0, \sigma_b^2), \quad \varepsilon \sim N(0, \sigma^2)$$

Em que: y = vetor com as variáveis dependentes observadas na base de dados; X = matriz de delineamento de efeitos fixos; β = vetor com parâmetros de efeito fixo; Z : matriz de incidência de efeitos aleatórios; b = vetor com parâmetros de efeito aleatório; ε = vetor com erros aleatórios; N = distribuição normal; σ_b^2 = matriz de variâncias e covariâncias dos parâmetros aleatórios, σ^2 = matriz de variâncias e covariâncias dos erros aleatórios, as quais podem assumir diferentes estruturas.

Dessa forma, tem-se:

$$y \sim N(X\beta, Z\sigma_b^2 Z' + \sigma^2)$$

A estimação dos parâmetros, geralmente, é baseada no Estimador da Máxima Verossimilhança (Hartley; Rao, 1967).

Estimação por Máxima Verossimilhança

A máxima verossimilhança tem como principal objetivo estimar os parâmetros do modelo da população que maximizem a probabilidade de obtenção da amostra que de fato foi obtida, ou seja, os parâmetros estimados devem maximizar a probabilidade da amostra. A identificação desses valores da amostra ocorre por meio da comparação dos dados observados com os dados preditos pelo modelo associado com os parâmetros do modelo e, quanto mais próximos os valores observados estão dos valores preditos, maior a probabilidade de que os dados observados provenientes são originados de uma população com parâmetros próximos àqueles utilizados para gerar os valores preditos (Finch *et al.*, 2019).

Na prática, a máxima verossimilhança é um processo iterativo onde o algoritmo busca por valores de parâmetros que irão maximizar a probabilidade de produzir valores preditos que são os mais próximos possíveis dos valores observados.

Estruturas de variância e covariância

A seleção da estrutura de covariância residual e o modelo mais adequados é de extrema importância na análise de dados com observações repetidas, uma vez que ignorada pode levar a conclusões inapropriadas (Toral *et al.*, 2006). Os erros-padrão das médias são dependentes da escolha da estrutura de covariância, de modo que, se esta for selecionada de forma inadequada, poderá afetar diretamente as estimativas dos parâmetros, os erros-padrão de efeitos fixos e aleatórios, diagnósticos e inferências. Além disso, essa seleção é influenciada pela estrutura dos dados, informações empíricas e disponibilidade computacional (Littell *et al.*, 1998; Dantas *et al.*, 2020).

Algumas das estruturas de variância e covariância comumente utilizadas são a Diagonal, Componente de variância (CV), Simétrica composta (SC), Não estruturada (NE), Autorregressiva de primeira ordem AR(1), Autorregressiva de primeira ordem heterogênea ARH(1), Toeplitz (TOEP), dentre outras (Wolfinger, 1993; Littell *et al.*, 2000; Dantas *et al.*, 2020).

Nas ciências florestais, a modelagem mista vem sendo aplicada em diferentes estudos de mensuração, como em trabalhos pioneiros envolvendo estimativas de curvas de índice de sítio (Biging, 1985), predição do desenvolvimento da altura de árvores dominantes (Lappi; Bailey, 1988), ajuste de equação de altura e volume (Lappi, 1991) e predição de volume acumulado do fuste (Grégoire; Schabenberger, 1996). Mais recentemente, a utilização de modelos de efeitos aleatórios na área florestal tem sido cada vez mais abordada na predição de volume (Gouveia *et al.*, 2015; Cerqueira *et al.*, 2020; Dantas *et al.*, 2020; Monteiro *et al.*, 2021) e biomassa (Zeng, 2015; Ou *et al.*, 2016; Chen *et al.*, 2017; Huff *et al.*, 2018), predição de altura (Raptis *et al.*, 2021; Ogana, 2021; Skudnik; Jevšenak, 2022; Xie *et al.*, 2022), modelagem do afilamento do fuste (Santos *et al.*, 2021; Shahzad *et al.*, 2021) e do crescimento em área basal (Cunha; Finger; Schneider, 2013; Di Cosmo *et al.*, 2020), em que os modelos mistos promoveram ganhos de qualidade nas estimativas.

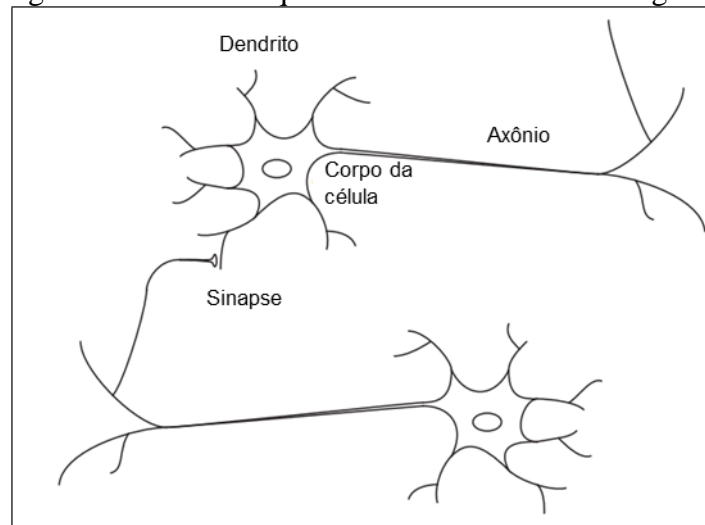
2.5 Redes neurais artificiais

Nos últimos anos as Redes neurais Artificiais (RNA) têm se tornado modelos de classificação, agrupamento, reconhecimento de padrões e predição populares e úteis em diversas áreas, incluindo a área florestal (Abiodun *et al.*, 2018; Skudnik; Jevšenak, 2022).

RNA possuem a capacidade de adquirir e manter informação baseada em conhecimento e podem ser definidas como um conjunto de unidades de processamento, representadas por neurônios artificiais interligados por várias conexões internas (sinapses artificiais), baseados em neurônios encontrados no cérebro humano (Braga *et al.*, 2007; Silva *et al.*, 2017).

O cérebro possui um grande número de elementos altamente conectados, chamados de neurônios. Estes neurônios possuem três componentes principais: os dendritos, o corpo celular e o axônio. Os dendritos são prolongamentos de fibras nervosas receptivos que levam sinais elétricos para o corpo celular. O corpo celular soma e limita estes sinais. O axônio é uma única fibra longa que leva o sinal do corpo celular para outros neurônios. O ponto de contato entre um axônio de uma célula e o dendrito de outra célula é chamado de sinapse. É a organização dos neurônios e a força das sinapses individuais, determinada por um processo químico complexo, que define a função da rede neural (Figura 1) (Hagan *et al.*, 2014).

Figura 1. Desenho esquemático de neurônios biológicos.



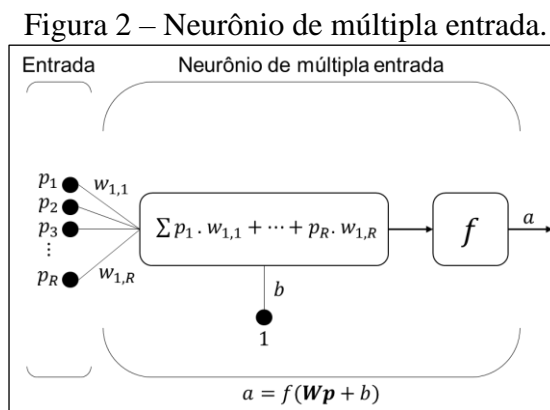
Fonte: Adaptado de Hagan et al. (2014).

Redes neurais artificiais não são tão complexas quanto o cérebro humano. No entanto, há duas semelhanças que merecem destaque entre as redes neurais artificial e biológica: 1) os blocos de construção de ambas as redes são dispositivos computacionais simples (sendo os neurônios artificiais muito mais simples) e 2) as conexões entre neurônios determinam a função da rede.

O trabalho de Warren McCulloch e Walter Pitts, nos anos 40, é reconhecido como pioneiro nos estudos sobre redes neurais. Os autores mostraram que redes de neurônios

artificiais poderiam, a princípio, computar qualquer função aritmética ou lógica (Mcculloch; Pitts, 1943).

Em uma descrição simplificada do modelo matemático do neurônio, pode-se considerar um neurônio de múltiplas entradas (Figura 2). Cada um dos valores de entrada p_1, p_2, \dots, p_R (dendritos) são multiplicados pelos seus respectivos pesos w_1, w_2, \dots, w_R , representando as sinapses, resultando em um vetor de valores $p_1w_1, p_2w_2, \dots, p_Rw_R$ (\mathbf{Wp}), que serão enviados para a célula de soma. O outro termo de entrada, 1, é multiplicado por um valor chamado *bias* (b) e então passado para a célula de soma. O resultado do somatório (n), geralmente chamado de saída da rede, servirá como entrada de uma função de ativação, que irá resultar na saída do neurônio (a).



Fonte: Adaptado de Hagan *et al.* (2014).

Geralmente, apenas um neurônio, mesmo com várias entradas, pode não ser o suficiente para obtenção de bons resultados. Dessa forma, pode ser necessário múltiplos neurônios operando paralelamente, o que podemos chamar de camadas (Hagan *et al.*, 2014), justificando assim a denominação da rede neural artificial, em que suas camadas se comunicam por pesos, captando a influência de cada variável de entrada sobre as mudanças na variável de saída e atribuindo um valor para cada variável. Portanto, uma vez que os pesos de um neurônio são ajustados, o conhecimento obtido estará disponível para ser utilizado na obtenção de saídas de interesse com base em entradas específicas. A camada em que a sua saída é a saída da rede é chamada de camada de saída e as outras camadas são chamadas de camadas ocultas.

A acurácia de predição de uma RNA depende do número de camadas utilizadas e, principalmente, do tipo de função de ativação utilizada. No entanto, não há uma especificação exata do número de camadas a serem utilizadas para melhores resultados. Além disso, também não há, na literatura, uma recomendação específica do tipo de função de ativação a ser utilizada (Sharma *et al.*, 2020).

As funções de ativação permitem a introdução da não-linearidade nas RNAs, possibilitando uma melhor adaptação à distribuição dos dados. As funções de ativação mais frequentemente utilizadas são a sigmoide e a tangente hiperbólica (Heaton, 2011). Além disso, as RNA se destacam pela sua capacidade de aprender e de generalizar o conhecimento obtido para um conjunto de dados desconhecidos, tolerância a falhas e ruídos, capacidade de modelar diversas variáveis e suas relações complexas e possibilidade de inclusão de variáveis categóricas (Haykin, 2001; Dantas *et al.*, 2018).

Uma RNA se inicia com uma fase de treinamento, por meio da qual as conexões entre unidades de processamento são ajustadas a partir de um sistema de aprendizagem supervisionado em que os ajustes podem ser realizados com diferentes algoritmos, como *Backpropagation* ou *Resilient Backpropagation*, que visam a minimização do erro entre os valores preditos de saída e os valores reais observados até que um erro aceitável, predefinido, seja alcançado (Haykin, 2009; Schmidhuber, 2015; Heaton, 2008; Mohamad-Saleh; Hoyle, 2008; De Oliveira Neto *et al.*, 2021).

Os modelos de RNA estão se tornando uma ferramenta de predição cada vez mais popular, uma vez que não é preciso obedecer a pressuposições sobre a forma de uma função de ajuste, de modo que, a RNA é treinada para encontrar a relação existentes entre as variáveis. Geralmente, as RNA são adequadas para problemas de modelagem em que: 1) há uma grande quantidade de dados e depende da interação de múltiplos parâmetros, 2) a área em estudo possui grande quantidade de dados históricos, 3) os dados disponíveis estão incompletos e 4) a função que determina a solução é desconhecida ou demanda grande esforço para ser encontrada (Bailey; Thompson, 1990; Leduc *et al.*, 2001; Diamantopoulou, 2006).

Maioria dessas condições são encontradas nos problemas de modelagem florestal e, dessa forma, diversos pesquisadores têm desenvolvido trabalhos com a aplicação de RNA na resolução desses problemas (Guan; Gertner, 1991; Blackard; Dean, 1999; Zhang *et al.*, 2000; Liu *et al.*, 2003; Özçelik *et al.*, 2010; Diamantopoulou *et al.*, 2015; LIMA *et al.*, 2021) como em ajustes hipsométricos (Diamantopoulou, 2012; Soares *et al.*, 2021; Casas *et al.*, 2022), predição volumétrica (Özçelik *et al.*, 2010; Gorgens *et al.*, 2014; Abreu *et al.*, 2018) e ajuste de funções de afilamento do fuste (Sakici; Ozdemir, 2018; Socha *et al.*, 2020).

REFERÊNCIAS

- ABIODUN, O. I. *et al.* **State-of-the-art in artificial neural network applications: A survey.** *Heliyon*, v. 4, n. 11, p. e00938, 2018.
- ABREU, J. C.; SOARES, C. P. B.; LEITE, H. G. Assessing alternatives to estimate the stem volume of a seasonal semi-deciduous forest. **Floresta**, v. 47, n. 4, p. 375-382, 2018.
- AKAIKE, H. Maximum likelihood identification of Gaussian autoregressive moving average models. **Biometrika**, v. 60, n. 2, p. 255-265, 1973.
- ANDRADE, V.C.L.; LEITE, H.G. Um método para quantificar multiprodutos de árvores individuais na unidade estere. **Revista Árvore**, v. 22, p. 299-306, 1998.
- ANDRADE, V.C.L. Modelos volumétricos de dupla entrada para aplicar em povoamentos florestais brasileiros. **Floresta e Ambiente**, v. 24, p. e00135415, 2017.
- ANDRADE, V.C.L.; LEITE, H. G. Modelagem da relação hipsométrica empregando dados amostrados na cubagem de árvores e em parcelas de inventário. **Revista Árvore**, v. 35, p. 157-164, 2011.
- ARIAS-RODIL, M. *et al.* Modelling and localizing a stem taper function for *Pinus radiata* in Spain. **Canadian Journal of Forest Research**, v. 45, n. 6, p. 647-658, 2015.
- ASHRAF, M. I. *et al.* Integrating biophysical controls in forest growth and yield predictions with artificial intelligence technology. **Canadian Journal of Forest Research**, v. 43, n. 12, p. 1162-1171, 2013.
- ASSIS, A. L. *et al.* Comparação de modelos polinomiais segmentados e não segmentados na estimativa de diâmetros e volumes ao longo do fuste de *Pinus taeda*. **Cerne**, v. 7, n. 1, p. 20-40, 2001.
- AZEVEDO, G. B. *et al.* Multi-volume modeling of *Eucalyptus* trees using regression and artificial neural networks. **Plos one**, v. 15, n. 9, p. e0238703, 2020.
- AZEVEDO, T. L. *et al.* Equações hipsométricas e volumétricas para um povoamento de *Eucalyptus* sp. localizado na FLONA do Ibura, Sergipe. **Revista Brasileira de Ciências Agrárias (Agrária)**, v.6, n.1, p.105-112, 2011.
- BAILEY, D.; THOMPSON, D. How to develop neural-network applications. **AI expert**, v. 5, n. 6, p. 38-47, 1990.
- BAILEY, R. L. A compatible volume-taper model based on the Schumacher and Hall generalized constant form factor volume equation. **Forest science**, v. 40, n. 2, p. 303-313, 1994.
- BELL, B. A. *et al.* A multilevel model primer using SAS PROC MIXED. **In: SAS global forum.** Columbia, SC, USA: University of South Carolina, p. 1-19, 2013.

- BERNARDI, L. K. *et al.* Diferentes modelos para o afilamento do tronco de *Eucalyptus* sp. para o cenário florestal brasileiro. **Ciência Florestal**, Santa Maria, v. 31, n. 3, p. 1364-1382, 2021.
- BERNARDI, L. K. *et al.* Diferentes modelos para o afilamento do tronco de *Eucalyptus* sp. para o cenário florestal brasileiro. **Ciência Florestal**, v. 31, p. 1364-1382, 2021.
- BIGING, G. S. Improved estimates of site index curves using a varying-parameter model. **Forest Science**, Lawrence, v. 31, p. 411-423, 1985.
- BINOTI, D. H. B.; BINOTI, M. L. M. S.; LEITE, H. G. Configuração de redes neurais artificiais para estimação do volume de árvores. **Revista Ciência da Madeira**, v. 5, n. 1, p. 58-67, 2014.
- BINOTI, D. H. B. *et al.* Support vector machine to estimate volume of eucalypt trees. **Revista Árvore**, v. 40, p. 689-693, 2016.
- BINOTI, M. L. M. S. *et al.* Redes neurais artificiais para estimação do volume de árvores. **Revista Árvore**, v. 38, n. 2, p. 283-288, 2014.
- BINOTI, M. L. M. S. *et al.* Prognose em nível de povoamento de clones de eucalipto empregando redes neurais artificiais. **Cerne**, v. 21, n. 1, p. 97-105, 2015.
- BISHOP, Christopher M. *et al.* **Neural networks for pattern recognition**. Oxford university press, 1995.
- BLACKARD, J. A.; DEAN, D. J. Comparative accuracies of artificial neural networks and discriminant analysis in predicting forest cover types from cartographic variables. **Computers and electronics in agriculture**, v. 24, n. 3, p. 131-151, 1999.
- BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDEMIR, T. B. **Redes Neurais Artificiais: Teoria e Aplicações**. Rio de Janeiro: Editora LTC, 2007. 226p.
- BRONISZ, K.; MEHTÄTALO, L. Mixed-effects generalized height–diameter model for young silver birch stands on post-agricultural lands. **Forest Ecology and Management**, v. 460, p. 117901, 2020.
- BUONGIORNO, J.; GILLESS, J. K. **Decision Methods for Forest Resource Management**. Elsevier, 2003. 458p.
- CALAMA, R.; MONTERO, G. Interregional nonlinear height diameter model with random coefficients for stone pine in Spain. **Canadian Journal of Forest Research**, v. 34, n. 1, p. 150-163, 2004.
- CALEGARIO, N. *et al.* Modeling dominant height growth based on nonlinear mixed-effects model: a clonal Eucalyptus plantation case study. **Forest Ecology and Management**, v. 204, n. 1, p. 11-21, 2005.
- CAMPOS, J. C. C.; LEITE, H. G. **Mensuração florestal: perguntas e respostas**. 2.ed. Viçosa, MG: Universidade Federal de Viçosa, 2006, 470p.

CAMPOS, J. C. C.; LEITE, H. G. **Mensuração Florestal: perguntas e respostas**. 4. ed. Viçosa, MG: UFV, 2013, 605 p.

CAMPOS, J. C. C.; TREVIZOL-JUNIOR, T. L.; PAULA-NETO, F. Ainda sobre a seleção de equações de volume. **Revista Árvore**, v. 9, n. 2, p. 115-126, 1985.

CANTY, A.; RIPLEY, B. Package 'boot'. **Bootstrap Functions**. CRAN R Proj, 2017.

CASAS, G. G. *et al.* Configuration of the Deep Neural Network Hyperparameters for the Hypsometric Modeling of the *Guazuma crinita* Mart. in the Peruvian Amazon. **Forests**, v. 13, n. 5, p. 697, 2022.

CASTAÑO-SANTAMARÍA, J. *et al.* Tree height prediction approaches for uneven-aged beech forests in northwestern Spain. **Forest Ecology and Management**, v. 307, p. 63-73, 2013.

CERQUEIRA, C. L. *et al.* Modelagem do volume de eucalipto em sistema de integração Lavoura-Pecuária-Floresta. **Advances in Forestry Science**, v. 7, n. 4, p. 1213-1221, 2020.

CHEN, D. *et al.* Biomass modeling of larch (*Larix* spp.) plantations in China based on the mixed model, dummy variable model, and Bayesian hierarchical model. **Forests**, v. 8, n. 8, p. 268, 2017.

CHEN, L. *et al.* Multi-sensor prediction of stand volume by a hybrid model of support vector machine for regression kriging. **Forests**, v. 11, n. 3, p. 296, 2020.

CHEN, S. *et al.* Applicability of personal laser scanning in forestry inventory. **PLoS One**, v. 14, n. 2, p. e0211392, 2019.

CICEU, A. *et al.* A generalized nonlinear mixed-effects height–diameter model for Norway spruce in mixed-uneven aged stands. **Forest Ecology and Management**, v. 477, p. 118507, 2020.

CLUTTER, J. L. Development of taper functions from variable top merchantable volume equations. **Forest Science**, v. 26, p.117-120, 1980.

CLUTTER, J. L. *et al.* **Timber management: A quantitative approach**. John Wiley & Sons, Inc., 1983.

CORDEIRO, M. A. *et al.* Volumetric estimates in eucalyptus stands using support vector machines and artificial neural networks. **Madera y bosques**, v. 28, n. 1, 2022.

CORRAL RIVAS, S.; SILVA ANTUNA, A. M.; QUIÑONEZ BARRAZA, G. A generalized nonlinear height-diameter model with mixed-effects for seven pinus species in Durango, Mexico. **Revista mexicana de ciencias forestales**, v. 10, n. 53, p. 86-117, 2019.

COSTA, E. A. *et al.* Understanding bark thickness variations for *Araucaria angustifolia* in southern Brazil. **Journal of Forestry Research**, v. 32, n. 3, p. 1077-1087, 2021.

CRECENTE-CAMPO, F. *et al.* A generalized nonlinear mixed-effects height–diameter model for *Eucalyptus globulus* L. in northwestern Spain. **Forest Ecology and Management**, v. 259, n. 5, p. 943-952, 2010.

CUNHA, T. A.; FINGER, C. A. G.; SCHNEIDER, P. R. Modelo linear misto para o incremento em área basal de árvores individuais de cedro (*Cedrela odorata* L.) na Amazônia ocidental, Brasil. **Ciência Florestal**, v. 23, n. 3, p. 461-470, 2013.

CURTIS, R.O. Height-diameter and height-diameter-age equations for second-growth Douglas-fir. **For. Sci.** 13, 365-375, 1967.

DA SILVA, S. A. *et al.* Funções de afilamento para *Pinus elliottii* engelm. proveniente de condução de regeneração natural. **Biofix Scientific Journal**, v. 3, n. 1, p. 130-136, 2018.

DANTAS, D. *et al.* Multilevel nonlinear mixed-effects model and machine learning for predicting the volume of *Eucalyptus* spp. trees. **Cerne**, v. 26, n. 1, p. 48-57, 2020.

DANTAS, D. *et al.* Comparação entre redes neurais artificiais e comitê de redes na estimação do volume individual de árvores de *Eucalyptus*/Artificial neural networks committee in the estimation of the individual volume of *Eucalyptus* trees. **Caderno de Ciências Agrárias**, v. 10, n. 1, p. 01-06, 2018.

DE OLIVEIRA NETO, R. R. *et al.* Estimation of *Eucalyptus* productivity using efficient artificial neural network. **European Journal of Forest Research**, v. 141, n. 1, p. 129-151, 2022.

DEMAERSCHALK, J. P. Integrated systems for the estimation of tree taper and volume. **Canadian Journal of Forest Research**, v. 3, n. 1, p. 90-94, 1973.

DE-MIGUEL, S.; GUZMÁN, G.; PUKKALA, T. A comparison of fixed-and mixed-effects modeling in tree growth and yield prediction of an indigenous neotropical species (*Centrolobium tomentosum*) in a plantation system. **Forest Ecology and Management**, v. 291, p. 249-258, 2013.

DI COSMO, L. *et al.* An individual-tree linear mixed-effects model for predicting the basal area increment of major forest species in Southern Europe. **Forest Systems**, v. 29, n. 3, p. e019-e019, 2020.

DIAMANTOPOULOU, M. J. Artificial neural networks as an alternative tool in pine bark volume estimation. **Computers and electronics in agriculture**, v. 48, n. 3, p. 235-244, 2005.

DIAMANTOPOULOU, M. J. Assessing a reliable modeling approach of features of trees through neural network models for sustainable forests. **Sustainable Computing: Informatics and Systems**, v. 2, n. 4, p. 190-197, 2012.

DIAMANTOPOULOU, M. J. *et al.* Estimation of Weibull function parameters for modelling tree diameter distribution using least squares and artificial neural networks methods. **Biosystems Engineering**, v. 133, p. 33-45, 2015.

DIAMANTOPOULOU, M. J. Predicting fir trees stem diameters using artificial neural network models. **Southern African Forestry Journal**, v. 205, n. 1, p. 39-44, 2005.

DIAMANTOPOULOU, M. J. **Tree-bole volume estimation on standing pine trees using cascade correlation artificial neural network models**. Agricultural Engineering International: CIGR Journal, 2006.

EDER, P. M. *et al.* Artificial intelligence tools in predicting the volume of trees within a forest stand. **African Journal of Agricultural Research**, v. 11, n. 21, p. 1914-1923, 2016.

EFRON, B. **The jackknife, the bootstrap and other resampling plans**. Philadelphia: Society for industrial and applied mathematics, 1982.

ERCANLI, İ. Artificial intelligence with deep learning algorithms to model relationships between total tree height and diameter at breast height. **Forest Systems**, v. 29, n. 2, p. e013-e013, 2020.

ERCANLI, İ. Innovative deep learning artificial intelligence applications for predicting relationships between individual tree height and diameter at breast height. **Forest Ecosystems**, v. 7, n. 1, p. 1-18, 2020.

FANG, Z.; BAILEY, R. L. Height–diameter models for tropical forests on Hainan Island in southern China. **Forest ecology and management**, v. 110, n. 1-3, p. 315-327, 1998.

FARIAS, K. J.; NICOLETTI, M. F.; DE AGUIAR, N. S. Quantificação do volume total por meio de modelos volumétricos e de afilamento em **Eucalyptus grandis** W. Hill ex Maiden. **Revista de Ciências Agroveterinárias**, v. 18, n. 1, p. 33-40, 2019.

FIGUEIREDO, E. O.; SCOLFORO, J. R. S.; OLIVEIRA, A. D. de. Seleção de modelos polinomiais para representar o perfil e volume do fuste de *Tectona grandis* L.f. **Acta Amazonica**, v. 36, n. 4, p. 465-482, 2006.

FIGUEIREDO, E. O. *et al.* LIDAR-based estimation of bole biomass for precision management of an Amazonian forest: Comparisons of ground-based and remotely sensed estimates. **Remote Sensing of Environment**, v. 187, p. 281-293, 2016.

FIGUEIREDO-FILHO, A.; BORDERS, B. E.; HITCH, K. L. Taper equations for *Pinus taeda* plantations in Southern Brazil. **Forest Ecology and Management**, v. 83, n. 1-2, p. 39-46, 1996.

FINCH, W. H.; BOLIN, J. E.; KELLEY, K. **Multilevel modeling using R**. Crc Press, 2019.

FISCHER, F. *et al.* Exatidão dos modelos polinomiais não-segmentados e das razões entre volumes para representar o perfil do tronco de *Pinus taeda*. **Ciência Florestal**, v. 11, p. 167-188, 2001.

FLORIANO, E. P. *et al.* Fitting and selecting traditional models for tree's height time series data. **Ciência Florestal**, v. 16, n. 2, p. 177-199, 2006.

FU, L. *et al.* Prediction of individual tree diameter using a nonlinear mixed-effects modeling approach and airborne LiDAR Data. **Remote sensing**, v. 12, n. 7, p. 1066, 2020.

FU, L.; SUN, W.; WANG, G. A climate-sensitive aboveground biomass model for three larch species in northeastern and northern China. **Trees**, v. 31, n. 2, p. 557-573, 2017.

GOMES, A.M.A. *Medições dos Arvoredos*. Lisboa: Livraria Sá Da Costa, 1957. 413p.

GORGENS, E. B. *et al.* Estimação do volume de árvores utilizando redes neurais artificiais. **Revista Árvore**, v. 33, n. 6, p. 1141-1147, 2009.

GORGENS, E. B. *et al.* Influência da arquitetura na estimativa de volume de árvores individuais por meio de redes neurais artificiais. **Revista Árvore**, v. 38, n. 2, p. 289-295, 2014.

GOUVEIA, J. F. *et al.* Modelos volumétricos mistos em clones de *Eucalyptus* no polo gesseiro do Araripe, Pernambuco. **Floresta**, v. 45, n. 3, p. 587-598, 2015.

GRÉGOIRE, T.G.; SCHABENBERGER, O. A non-linear mixed-effects model to predict cumulative bole volume of standing trees. **Journal of Applied Statistics**. v. 23, n. 2–3, p. 257–272, 1996.

GREGOIRE, T. G.; SCHABENBERGER, O. Nonlinear mixed-effects modeling of cumulative bole volume with spatially correlated within-tree data. **Journal of Agricultural, Biological, and Environmental Statistics**, p. 107-119, 1996.

GUAN, B. T.; GERTNER, G. Modeling red pine tree survival with an artificial neural network. **Forest Science**, v. 37, n. 5, p. 1429-1440, 1991.

GUIMARÃES, D. P.; LEITE, H. G. Influência do número de árvores na determinação de equação volumétrica para *Eucalyptus grandis*. **Scientia forestalis**, v. 50, p. 37-42, 1996.

GUIMARÃES, D.P.; LEITE, H.G. Um novo modelo para descrever o perfil do tronco. **Revista Árvore**, v.16, p.170-180, 1992.

GÜNTHER, F.; FRITSCH, S. Neuralnet: training of neural networks. **R J.**, v. 2, n. 1, p. 30, 2010.

HAGAN, M. T. *et al.* **Neural Network Design**. eBook, Ed. 2., 2014, 1012p.

HALL, K. B. *et al.* A growth and yield model for *Eucalyptus benthamii* in the southeastern United States. **Forest Science**, v. 66, n. 1, p. 25-37, 2020.

HAMIDI, S. K. *et al.* Analysis of plot-level volume increment models developed from machine learning methods applied to an uneven-aged mixed forest. **Annals of Forest Science**, v. 78, n. 1, p. 1-16, 2021.

HAO, Y. *et al.* Individual tree diameter estimation in small-scale forest inventory using UAV laser scanning. **Remote Sensing**, v. 13, n. 1, p. 24, 2020.

HARTLEY, H. O.; RAO, J. N. K. Maximum-likelihood estimation for the mixed analysis of variance model. **Biometrika**, v. 54, n. 1-2, p. 93-108, 1967.

HASENAUER, H.; MERKL, D.; WEINGARTNER, M. Estimating tree mortality of Norway spruce stands with neural networks. **Advances in Environmental Research**, v. 5, n. 4, p. 405-414, 2001.

HAYKIN, S. O. Neural networks and learning machines: international version, ed. **Upper Saddle River [etc.] Pearson Education**, 2009.

HAYKIN, S. **Redes neurais: princípios e prática**. Bookman Editora, 2001.

HEATON, Jeff. **Introduction to neural networks with Java**. Heaton Research, Inc., ed. 2, 2008, 417p.

HEATON, Jeff. **Introduction to the Math of Neural Networks (Beta-1)**. Heaton Research Inc, St. Louis, MO, USA, 2011, 112p.

HORLE, D. C. *et al.* Modelagem não linear do perfil longitudinal de fustes de *Pinus oocarpa*. **Cerne**, v. 16, p. 177-184, 2010.

HRADETZKY, J. **Analyse und interpretation statistischer abränger keiten. (Biometrische Beiträge zu aktuellen forschungs projekten)**. Baden: Württemberg Mitteilungen der FVA, 1976. 146 p.

HUANG S, PRICE D, TITUS S. J. Development of ecoregion-based height–diameter models for white spruce in boreal forests. **Forest ecology and management**, v. 129, n. 1-3, p. 125–141, 2000.

HUANG, S.; TITUS, S. J.; WIENS, D. P. Douglas P. Comparison of nonlinear height–diameter functions for major Alberta tree species. **Canadian Journal of Forest Research**, v. 22, n. 9, p. 1297-1304, 1992.

HUANG, S.; TITUS, S. J. An age-independent individual tree height prediction model for boreal spruce–aspen stands in Alberta. **Canadian Journal of Forest Research**, v. 24, n. 7, p. 1295-1301, 1994.

HUFF, S. *et al.* Quantifying aboveground biomass for common shrubs in northeastern California using nonlinear mixed effect models. **Forest ecology and management**, v. 424, p. 154-163, 2018.

HUSCH, B.; MILLER, C.I.; BEERS, T.W. **Forest mensuration**. New York: John Wiley & Sons, 1982. 402p.

IGE, P. O.; AKINYEMI, G. O.; SMITH, A. S. Nonlinear growth functions for modeling tree height–diameter relationships for *Gmelina arborea* (Roxb.) in south-west Nigeria. **Forest Science and Technology**, v. 9, n. 1, p. 20-24, 2013.

ILIADIS, L. *et al.* Predicting Douglas-fir wood density by artificial neural networks (ANN) based on progeny testing information. **Holzforschung**, v. 67, n. 7, p. 771-777, 2013.

INDÚSTRIA BRASILEIRA DE ÁRVORES. **Relatório anual IBÁ 2023**. São Paulo, 2023.

JAYARAMAN, K.; ZAKRZEWSKI, W. T. Practical approaches to calibrating height–diameter relationships for natural sugar maple stands in Ontario. **Forest Ecology and Management**, v. 148, n. 1-3, p. 169-177, 2001.

JONSON, T. **Taxatoriska Undersökningar Över Skogsträdens Form**. Commonwealth Forestry Bureau, 1940.

KERSHAW JR, J. A. *et al.* **Forest mensuration**. John Wiley & Sons, 2016.

KHAN, M. N. I. *et al.* Allometric relationships of stem volume and stand level carbon stocks at varying stand density in *Swietenia macrophylla* King plantations, Bangladesh. **Forest ecology and management**, v. 430, p. 639-648, 2018.

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

KUZNETSOVA, A.; BROCKHOFF, P. B.; CHRISTENSEN, R. H. B. lmerTest: package: tests in linear mixed effects models. **Journal of statistical software**, v. 82, n. 1, p. 1-26, 2017.

LAPPI, J. Calibration of height and volume equations with random parameters. **Forest Science**, v. 37, n. 3, p. 781 - 801, 1991.

LAPPI, J.; BAILEY, R. L. A height prediction model with random stand and tree parameters: An alternative to traditional site index methods. **Forest. Science**, v. 38, n. 2, p.409-429, 1988.

LARSEN, D.R.; HANN, D.W. **Height-diameter equations for seventeen tree species in southwest Oregon**. Forest Research Laboratory, Oregon State University, Corvallis. Res. Pap. 49., 1987, 161 p.

MEHTÄTALO, L.; DE-MIGUEL, S.; GREGOIRE, T. G. Modeling height-diameter curves for prediction. **Canadian Journal of Forest Research**, v. 45, n. 7, p. 826-837, 2015.

LEDUC, D. J. **Predicting diameter distributions of longleaf pine plantations: a comparison between artificial neural networks and other accepted methodologies**. US Department of Agriculture, Forest Service, Southern Research Station, 2001.

LEI, X. *et al.* Individual height–diameter models for young black spruce (*Picea mariana*) and jack pine (*Pinus banksiana*) plantations in New Brunswick, Canada. **The Forestry Chronicle**, v. 85, n. 1, p. 43-56, 2009.

LEITE, H. G.; ANDRADE, V. C. L. Importância das variáveis altura dominante e altura total em equações hipsométricas e volumétricas. **Revista Árvore**, v. 27, n. 3, p. 301-310, 2003.

LEITE, H. G.; ANDRADE, V. C. L. Um método para condução de inventários florestais sem o uso de equações volumétricas. **Revista Árvore**, v. 26, p. 321-328, 2002.

- LI, D. *et al.* Analysis of taper functions for *Larix olgensis* using mixed models and TLS. **Forests**, v. 12, n. 2, p. 196, 2021.
- LIMA, R. B. *et al.* Estimating tree volume of dry tropical forest in the Brazilian Semi-Arid Region: a comparison between regression and artificial neural networks. **Journal of Sustainable Forestry**, v. 40, n. 3, p. 281-299, 2021.
- LITTELL, R.C.; HENRY, P.R.; AMMERMAN, C.B. Statistical analysis of repeated measures data using SAS procedures. **Journal of Animal Science**, v. 76, n. 4, p. 1216-1231, 1998.
- LITTELL, R. C.; PENDERGAST, J.; NATARAJAN, R. Modelling covariance structure in the analysis of repeated measures data. **Statistics in medicine**, v. 19, n. 13, p. 1793-1819, 2000.
- LIU, C. *et al.* Comparison of neural networks and statistical methods in classification of ecological habitats using FIA data. **Forest Science**, v. 49, n. 4, p. 619-631, 2003.
- LIU, Z. *et al.* Application of machine-learning methods in forest ecology: recent progress and future challenges. **Environmental Reviews**, v. 26, n. 4, p. 339-350, 2018.
- LYNCH, T. B.; MURPHY, P. A. A compatible height prediction and projection system for individual trees in natural, even-aged shortleaf pine stands. **Forest Science**, v. 41, n. 1, p. 194-209, 1995.
- MARTINS, A. P. M. *et al.* Eficiência de métodos para estimativa do volume total de *Corymbia citriodora*. **Scientia Agraria Paranaensis**, v. 15, n. 2, p. 178-184, 2016.
- MAX, T. A.; BURKHART, H. E. Segmented polynomial regressions applied to taper equations. **Forest Science**, v.22, p.283-289, 1976.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, v. 5, n. 4, p. 115-133, 1943.
- MEHTÄTALO, L. A longitudinal height–diameter model for Norway spruce in Finland. **Canadian Journal of Forest Research**, v. 34, n. 1, p. 131-140, 2004.
- MENDONÇA, A. D.; CARVALHO, S. P. C; CALEGARIO, N. Modelos hipsométricos generalizados mistos na predição da altura de *Eucalyptus* sp. **Cerne**, v. 21 n. 1, p. 107-115, 2015.
- MENON, M. U. Ajuste dos modelos de Kozak e do sistema Burkhart e Cao para plantações de *Pinus oocarpa* Adjustment to the Models of Kozak and the Burkhart e Cao System for the *Pinus oocarpa* Plantation. **Ambiência**, v. 2, n. 2, p. 197-205, 2006.
- MEYER, D. *et al.* Package ‘e1071’. **The R Journal**, 2019.
- MIGUEL, E. P. *et al.* Alternative methods of scaling *Eucalyptus urophylla* trees in forest stands: compatibility and accuracy of volume equations. **iForest-Biogeosciences and Forestry**, v. 11, n. 2, p. 275, 2018.

MIGUEL, E. P. *et al.* Modelagem na predição do volume individual em plantio de *Eucalyptus urograndis*. **Rev. Bras. Biom**, v. 32, n. 4, p. 584-598, 2014.

MOHAMAD-SALEH, J. *et al.* Improved neural network performance using principal component analysis on Matlab. **International journal of the computer, the internet and Management**, v. 16, n. 2, p. 1-8, 2008.

MONTEIRO, B. C. *et al.* Uso de modelos mistos para estimativa do volume de árvores individuais em tipologias florestais no estado do Amapá. **Biota Amazônia**, v. 11, n. 2, p. 7-10, 2021.

MOORE, J. A.; ZHANG, L.; STUCK, D. Height-diameter equations for ten tree species in the Inland Northwest. **Western Journal of Applied Forestry**, v. 11, n. 4, p. 132-137, 1996.

MÜLLER, I.; FINGER, C. A. G.; SCHNEIDER, P. R. Forma de tronco e sortimentos de madeira de *Eucalyptus grandis* (Hill ex. Maiden), na Região Sudeste do Estado do Rio Grande do Sul. **Ciência Florestal**, v. 15, n. 3, p. 293-305, 2005.

NÁVAR, J.; DE JESÚS RODRÍGUEZ-FLORES, F.; DOMÍNGUEZ-CALLEROS, Pedro A. Taper functions and merchantable timber for temperate forests of northern Mexico. **Annals of Forest Research**, v. 56, n. 1, p. 165-178, 2013.

NICOLETTI, M.F.; MACHADO, S.A.; CARVALHO, S.P.C.; FIGUEIREDO FILHO, A.; SANQUETTA, C.R. Descrição do perfil do fuste de árvores por meio das funções de afilamento. **Scientia Agraria Paranaensis**, v. 15, n. 4, p. 365-372, 2016.

NIETO, P. J. G. *et al.* Support vector machines and neural networks used to evaluate paper manufactured using *Eucalyptus globulus*. **Applied Mathematical Modelling**, v. 36, n. 12, p. 6137-6145, 2012.

NIKLAS, K. J. Size-dependent allometry of tree height, diameter and trunk-taper. **Annals of botany**, v. 75, n. 3, p. 217-227, 1995.

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, 2016.

OGANA, F. N. A mixed-effects height-diameter model for *Gmelina arborea* Roxb stands in Southwest Nigeria. **Journal of Forest Research**, v. 27, n. 1, p. 1-7, 2021.

OLIVEIRA, M. L. R. *et al.* Estimação do volume de árvores de clones de eucalipto pelo método da similaridade de perfis. **Revista Árvore**, v. 33, p. 133-141, 2009.

OLIVEIRA, R. R. *et al.* Acurácia de relações hipsométricas para diferentes estratégias de validação em *Eucalyptus urograndis*. **Revista Verde de Agroecologia e Desenvolvimento Sustentável**, v. 11, n. 5, p. 123-127, 2016.

ORMEROD, D.W. A simple bole model. **Forestry Chronicle**, v.49, p.136-138, 1973.

OU, G. *et al.* Incorporating topographic factors in nonlinear mixed-effects models for aboveground biomass of natural Simao pine in Yunnan, China. **Journal of Forestry Research**, v. 27, n. 1, p. 119-131, 2016.

ÖZÇELİK, R. *et al.* Predicting tree height from tree diameter and dominant height using mixed-effects and quantile regression models for two species in Turkey. **Forest ecology and management**, v. 419, p. 240-248, 2018.

ÖZÇELİK, R. *et al.* Comparative study of standard and modern methods for estimating tree bole volume of three species in Turkey. **Forest Products Journal**, v. 58, n. 6, p. 73, 2008.

ÖZÇELİK, R. *et al.* Development of ecoregion-based height-diameter models for 3 economically important tree species of southern Turkey. **Turkish Journal of Agriculture and Forestry**, v. 38, n. 3, p. 399-412, 2014.

ÖZÇELİK, R. *et al.* Estimating Crimean juniper tree height using nonlinear regression and artificial neural network models. **Forest ecology and management**, v. 306, p. 52-60, 2013.

ÖZÇELİK, R. *et al.* Estimating tree bole volume using artificial neural network models for four species in Turkey. **Journal of environmental management**, v. 91, n. 3, p. 742-753, 2010.

PARRESOL, B. R. Baldcypress height–diameter equations and their prediction confidence intervals. **Canadian journal of forest research**, v. 22, n. 9, p. 1429-1434, 1992.

SOARES, P.; TOMÉ, M. Height–diameter equation for first rotation eucalypt plantations in Portugal. **Forest Ecology and Management**, v. 166, n. 1–3, p. 99-109, 2002.

PAULO, J.A., TOMÉ, J.; TOMÉ, M. Nonlinear fixed and random generalized height–diameter models for Portuguese cork oak stands. **Annals of Forest Science**, n. 68, p. 295–309, 2011.

PENG, C. *et al.* Developing and evaluating tree height-diameter models at three geographic scales for black spruce in Ontario. **Northern Journal of Applied Forestry**, v. 21, n. 2, p. 83-92, 2004.

PENG, C.; ZHANG, L.; LIU, J. Developing and validating nonlinear height–diameter models for major tree species of Ontario's boreal forests. **Northern Journal of Applied Forestry**, v. 18, n. 3, p. 87-94, 2001.

PINHEIRO, J.; BATES, D. **Mixed-effects models in S and S-PLUS**. Springer science & business media, 2006.

PINHEIRO, J.; BATES, D. **Mixed-effects models in S and S-PLUS**. Springer science & business media, 2000.

PIRES, L. M.; CALEGARIO, N. Ajuste de modelos estocásticos lineares e não-lineares para a descrição do perfil longitudinal de árvores. **Revista Árvore**, v. 31, p. 845-852, 2007.

RAPTIS, D. I. *et al.* Mixed-effects height–diameter models for black pine (*Pinus nigra* Arn.) forest management. **Trees**, v. 35, n. 4, p. 1167-1183, 2021.

RESENDE, M. D. V.; SILVA, F. F.; AZEVEDO, C. F. **Estatística matemática, biométrica e computacional: Modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão aleatória, seleção genômica, QTL-GWAS, estatística espacial e temporal, competição, sobrevivência.** Viçosa: Suprema gráfica e Editora Ltda, 2014, 882p.

RICHARDS, F. J. A flexible growth function for empirical use. **Journal of experimental Botany**, v. 10, n. 2, p. 290-301, 1959.

CASTAÑO-SANTAMARÍA, Javier *et al.* Tree height prediction approaches for uneven-aged beech forests in northwestern Spain. **Forest Ecology and Management**, v. 307, p. 63-73, 2013.

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, 2018.

SÁNCHEZ-GONZÁLEZ, M.; CAÑELLAS, I.; MONTERO, G. Generalized height-diameter and crown diameter prediction models for cork oak forests in Spain. **Forest Systems**, v. 16, n. 1, p. 76-88, 2007.

SANTOS, G. M. *et al.* Modelagem mista generalizada para estimar afilamento do fuste de árvores de *Pinus taeda* em diferentes espaçamentos de plantio. **Advances in Forestry Science**, v. 8, n. 1, p. 1261-1269, 2021.

SAUD, P. *et al.* Using quadratic mean diameter and relative spacing index to enhance height–diameter and crown ratio models fitted to longitudinal data. **Forestry**, v. 89, n. 2, p. 215-229, 2016.

SCHIMIDT, P. B. Determinação indireta da relação hipsométrica para povoamentos de *Pinus taeda* L. **Floresta**, v. 8, n. 1, p. 24-27, 1977.

SCHMIDHUBER, J. Deep learning in neural networks: An overview. **Neural networks**, v. 61, p. 85-117, 2015.

SCHÖEPFER, W. **Automatisierung des massen, sorten und wertberechnung stenender waldbestände schriftenreihe Bad.** Berlin: Wurt-Forstl, 1966.

SCHUMACHER, F. X.; HALL, F.S. Logarithmic expression of timber-tree volume. **J Agric Res**, n. 47, p. 719–734, 1933.

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, 2018.

SCOLFORO, J. R. S. *et al.* Acuracidade de equações de afilamento para representar o perfil do fuste de *Pinus elliottii*. **Cerne**, v. 4, n. 1, p. 100-122, 1998.

SCOLFORO, J. R. S. **Biometria florestal 2: técnicas de regressão aplicada para estimar: volume, biomassa, relação hipsométrica e múltiplos produtos de madeira**. UFLA/FAEPE, Lavras. 292p, 1997.

SHAHZAD, M. K. *et al.* Stem taper functions for *Betula platyphylla* in the Daxing'an Mountains, northeast China. **Journal of Forestry Research**, v. 32, n. 2, p. 529-541, 2021.

SHARMA, M.; PARTON, J. Height-diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach. **Forest Ecology and Management**, n. 249, p. 187–198, 2007.

SHARMA, M.; YIN ZHANG, S. Height–diameter models using stand characteristics for *Pinus banksiana* and *Picea mariana*. **Scandinavian Journal of Forest Research**, v. 19, n. 5, p. 442-451, 2004.

SHARMA, R. P. *et al.* A nonlinear mixed-effects height-to-diameter ratio model for several tree species based on Czech national forest inventory data. **Forests**, v. 10, n. 1, p. 70, 2019.

SHARMA, R. P.; BREIDENBACH, J. Modeling height-diameter relationships for Norway spruce, Scots pine, and downy birch using Norwegian national forest inventory data. **Forest Science and Technology**, v. 11, n. 1, p. 44-53, 2015.

SHARMA, S.; SHARMA, S.; ATHAIYA, A. Activation functions in neural networks. **International Journal of Engineering Applied Sciences and Technology**, v. 4, n. 12, p. 310-316, 2020.

SILVA, I. N. *et al.* Artificial neural networks. **Springer**, Cham, p. 21-28, 2017.

SILVA, J. P. M. *et al.* Prognosis of forest production using machine learning techniques. **Information Processing in Agriculture**, 2021.

SILVA, M. L.; JACOVINE, L. A. G.; VALVERDE, S. R. **Economia Florestal**, Viçosa: UFV, 2005. 178p.

SILVA, M. L. M. *et al.* Ajuste do modelo de Schumacher e Hall e aplicação de redes neurais artificiais para estimar volume de árvores de eucalipto. **Revista Árvore**, v. 33, p. 1133-1139, 2009.

SKUDNIK, M.; JEVŠENAK, J. Artificial neural networks as an alternative method to nonlinear mixed-effects models for tree height predictions. **Forest Ecology and Management**, v. 507, p. 1-9, 2022.

SMALIAN, H. L. **Beitrag zur holzmeßkunst (Contribution to the art of measuring wood)**. Verlag Löffler, Stralsund, Germany, 1837, 87p.

SOARES, F. A. A. M. N. *et al.* Recursive diameter prediction and volume calculation of eucalyptus trees using Multilayer Perceptron Networks. **Computers and electronics in agriculture**, v. 78, n. 1, p. 19-27, 2011.

- SOARES, G. M. *et al.* Artificial Neural Networks (ANN) For Height Estimation in A Mixed-Species Plantation of *Eucalyptus globulus* Labill and *Acacia mearnsii* De Wild. **Revista Árvore**, v. 45, 2021.
- SOARES, P. *et al.* Evaluating a growth model for forest management using continuous forest inventory data. **Forest Ecology and Management**, v. 71, n. 3, p. 251-265, 1995.
- SOARES, T. S.; LEITE, H. G.; VALE, A. B. Um modelo alternativo para a quantificação de multiprodutos em árvores individuais. **Revista Árvore**, v. 28, p. 875-882, 2004.
- 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, 2020.
- SOUSA, G. T. O.; AZEVEDO, G. B.; BARRETO, P. A. B.; CONCEIÇÃO JÚNIOR, V. Relações hipsométricas para *Eucalyptus urophylla* conduzidos sob regime de alto fuste e talhadia no Sudoeste da Bahia. **Revista Scientia Plena**, v. 9, n. 4, 2013.
- SOUZA, C. A. M. de. *et al.* Avaliação de modelos de afilamento segmentados na estimação da altura e volume comercial de *Eucalyptus* sp. **Revista Árvore**, v. 32, n. 3, p. 453-463, 2008.
- SOUZA, C. A. M. *et al.* Avaliação de modelos de afilamento não-segmentados na estimação da altura e volume comercial de *Eucalyptus* sp. **Ciência Florestal**, v. 18, p. 387-399, 2008.
- SUBEDI, N.; SHARMA, M. Individual-tree diameter growth models for black spruce and jack pine plantations in northern Ontario. **Forest Ecology and Management**, v. 261, n. 11, p. 2140-2148, 2011.
- TAMM, T.; REMM, K. Estimating the parameters of forest inventory using machine learning and the reduction of remote sensing features. **International Journal of Applied Earth Observation and Geoinformation**, v. 11, n. 4, p. 290-297, 2009.
- TAN, P. N.; STEINBACH, M.; KUMAR, V.; **Introdução ao data mining: mineração de dados**. Rio de Janeiro: Ciência Moderna, 2009, 900p.
- TEMESGEN, H. *et al.* Generalized height–diameter models—an application for major tree species in complex stands of interior British Columbia. **European Journal of Forest Research**, v. 123, n. 1, p. 45-51, 2004.
- TEMESGEN, H.; HANN, D. W.; MONLEON, V. J. Regional height–diameter equations for major tree species of southwest Oregon. **Western Journal of Applied Forestry**, v. 22, n. 3, p. 213-219, 2007.
- TÉO, S. J.; MIGUEL, E. P.; NASCIMENTO, R. G. M. Modelos de afilamento e forma dos fustes de eucalipto, em diferentes qualidades de sítio. **Nativa**, v. 6, p. 832-840, 2018.
- THANH, T. N.; TIEN, T. D.; SHEN, H. L. Height-diameter relationship for *Pinus koraiensis* in Mengjiagang Forest Farm of Northeast China using nonlinear regressions and artificial neural network models. **Journal of Forest Science**, v. 65, n. 4, p. 134-143, 2019.

- THOMAS, C. *et al.* Comparação de equações volumétricas ajustadas com dados de cubagem e análise de tronco. **Ciência Florestal**, v. 16, p. 319-327, 2006.
- TORAL, F. L. B.; ALENCAR, M. M.; FREITAS, A. R. Arranjos para efeitos fixos e estruturas de (co) variâncias residuais para análises de medidas repetidas do peso de bovinos da raça Canchim. **Revista Brasileira de Zootecnia**, v. 35, p. 1951-1958, 2006.
- TRINCADO, G.; VANDERSCHAAF, C. L.; BURKHART, H. E. Regional mixed-effects height–diameter models for loblolly pine (*Pinus taeda* L.) plantations. **European Journal of Forest Research**, v. 126, n. 2, p. 253-262, 2007.
- VALENÇA, M. **Fundamentos das redes neurais: exemplos em Java**. Olinda, Pernambuco: Editora Livro Rápido, 2007, 382p.
- VAN LAAR, A.; AKÇA, A. **Forest mensuration**. Springer Science & Business Media, 2007.
- VANCLAY, J. K. **Modelling Forest growth and yield: applications to mixed tropical forests**. CAB International, 1994. 330 p.
- CASTRO, R. V. O. *et al.* Individual growth model for *Eucalyptus* stands in Brazil using artificial neural network. **International Scholarly Research Notices**, 2013.
- WANG, C. H.; HANN, D. W. **Height-diameter equations for sixteen tree species in the central western Willamette valley of Oregon**. Forest Research Laboratory Oregon State University, Corvallis, 1988.
- WANG, M.; SUN, R.; XIAO, Z. Estimation of forest canopy height and aboveground biomass from spaceborne LiDAR and Landsat imageries in Maryland. **Remote Sensing**, v. 10, n. 2, p. 344, 2018.
- WEISKITTEL, A. R. *et al.* **Forest growth and yield modeling**. John Wiley & Sons, 2011, 344p.
- WICKHAM, H. **ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics**. 2016.
- WOLFINGER, R. Covariance structure selection in general mixed models. **Communications in statistics-Simulation and computation**, v. 22, n. 4, p. 1079-1106, 1993.
- XIE, L. *et al.* Evaluation of the mixed-effects model and quantile regression approaches for predicting tree height in larch (*Larix olgensis*) plantations in northeastern China. **Canadian Journal of Forest Research**, v. 52, n. 3, p. 309-319, 2022.
- ZAKRZEWSKI, W. T.; BELLA, I. E. Two new height models for volume estimation of lodgepole pine stands. **Canadian Journal of Forest Research**, v. 18, n. 2, p. 195-201, 1988.
- ZENG, W. S. Using nonlinear mixed model and dummy variable model approaches to develop origin-based individual tree biomass equations. **Trees**, v. 29, n. 1, p. 275-283, 2015.
- ZHANG, Q. B. *et al.* Modeling tree-ring growth responses to climatic variables using artificial neural networks. **Forest Science**, v. 46, n. 2, p. 229-239, 2000.

SEGUNDA PARTE – ARTIGOS

ARTIGO 1 – Modelagem mista e aprendizagem de máquina para a predição de altura total de árvores de *Pinus caribaea* var. *hondurensis*

ARTIGO FORMATADO DE ACORDO COM AS NORMAS DA NBR 6022 (ABNT, 2018)

RESUMO

O presente estudo teve como objetivo comparar, a partir de medidas estatísticas de qualidade, o desempenho de modelos não lineares de efeitos fixos e mistos e redes neurais artificiais (RNA) locais e genéricos na predição da altura total de árvores de *Pinus caribaea* var. *hondurensis*. Os dados utilizados no presente estudo foram coletados em plantios de *Pinus caribaea* var. *hondurensis* situados no município de Nova Ponte, Minas Gerais. Foram ajustados cinco modelos hipsométricos não lineares de efeitos fixos e, a partir do melhor, buscou-se desenvolver um novo modelo genérico adicionando variáveis a nível de UA. Em seguida, para descrever a relação hipsométrica, por meio dos modelos mistos, foi incorporada, aos modelos local e genérico selecionados, a variabilidade entre parcelas. Além disso, foi realizada a calibração, a partir da base de teste, para estimação dos novos parâmetros aleatórios e seleção da melhor estratégia de calibração. Também foram treinados modelos de redes neurais artificiais para estimar as alturas totais das árvores adotando-se diferentes tipos e configurações. O desempenho dos diferentes tipos de modelos testados foi avaliado com base na raiz do erro quadrático médio (RMSE), erro médio absoluto percentual (MAPE), desvio médio (BIAS), índice de eficiência do modelo (EF) e Critério de informação de Akaike (AIC). Para comparação dos modelos mistos entre si, foi utilizado o critério de Informação de Akaike (AIC), o critério Bayesiano de Schwarz (BIC), o logaritmo da verossimilhança maximizada (LogLik) e o teste da razão de verossimilhança (MLRT). Também, gráficos de distribuição dos resíduos foram construídos para auxiliar na avaliação dos modelos avaliados. Por fim, o teste t pareado foi aplicado, comparando os valores observados com os valores preditos pelas metodologias. Dentre as equações não lineares avaliadas nesse estudo, a equação de Ratkowsky (1990) foi a mais adequada. A inclusão dos atributos altura e diâmetro dominantes resultou em melhorias das predições, sendo recomendada a inclusão dessas variáveis na modelagem da relação hipsométrica de *Pinus caribaea* var. *hondurensis*. A inclusão de efeitos aleatórios, combinada com a calibração utilizando nove árvores distribuídas em diferentes classes diamétricas por UA, foi capaz de aumentar significativamente o desempenho do modelo local. A utilização dessa técnica implica em uma redução do número de alturas mensuradas por UA. A adição dos efeitos aleatórios ao modelo genérico foi capaz de gerar predições satisfatórias sem a necessidade de calibração quando aplicada à base de teste, sendo uma estratégia viável e conveniente para o aumento da precisão das predições de altura total. Os modelos de redes neurais artificiais não foram capazes de superar os modelos ONLS genéricos nem a modelagem mista e, também, resultaram em uma generalização de predições com viés, não sendo assim recomendadas para a base de dados em estudo. Dessa forma, verificou-se a necessidade de estudo mais aprofundados sobre a otimização de estratégias para a escolha da estrutura mais adequada de RNA e interpretação de seus parâmetros.

Palavras-chave: Efeito aleatório; Redes neurais artificiais; Ratkowsky; Relação hipsométrica; Modelagem florestal.

ABSTRACT

The present study aimed to compare, based on statistical quality measures, the performance of nonlinear fixed and mixed effects models and local and generic artificial neural networks (ANN) in predicting the total height of *Pinus caribaea* var. *hondurensis*. The data used in the present study were collected in *Pinus caribaea* var. *hondurensis* located in the municipality of Nova Ponte, Minas Gerais. Five nonlinear fixed effects models were adjusted and, based on the best, we sought to develop a new generic model by adding variables at the plot level. Then, to describe the hypsometric relationship, through mixed effect modeling, the variability between plots was incorporated into the selected local and generic models. Furthermore, calibration was carried out, based on the test base, to estimate the new random parameters and select the best calibration strategy. Artificial neural network models were also trained to estimate the total heights of the trees. Different types and configurations of neural networks were tested. The performance of the different types of models tested was evaluated based on root mean square error (RMSE), mean absolute percentage error (MAPE), mean deviation (BIAS), model efficiency index (EF), and Akaike Information Criterion (AIC). To compare the mixed models with each other, the Akaike Information Criterion (AIC), the Schwarz Bayesian Criterion (BIC), the maximized logarithm of likelihood (LogLik), and the likelihood ratio test (MLRT) were used. Also, residual distribution graphs were constructed to assist in the evaluation of the evaluated models. Finally, the paired t-test was applied, comparing the observed values with the predicted values. Among the nonlinear equations evaluated in this study, Ratkowsky (1990) was the most appropriate. The inclusion of dominant height and dominant diameter resulted in improved predictions, and the inclusion of these variables in modeling the hypsometric relationship of *Pinus caribaea* var. *hondurensis*. The inclusion of random effects, combined with calibration using nine trees distributed in different diameter classes per plot, was able to significantly increase the performance of the local model. The use of this technique implies a reduction in the number of heights measured per plot. The addition of random effects to the generic model was able to generate satisfactory predictions without the need for calibration when applied to the test base, being a viable and convenient strategy for increasing the accuracy of total height predictions. Artificial neural network models were not able to overcome generic fixed effects models or mixed modeling and resulted in a generalization of biased predictions, thus not being recommended for the database under study. Therefore, there was a need for further study on the optimization of strategies for choosing the most appropriate RNA structure and interpreting its parameters.

Keywords: Random effect; Artificial neural networks; Ratkowsky; Hypsometric relationship; Forest modeling.

1 INTRODUÇÃO

O setor de base florestal brasileiro destaca-se por sua alta produtividade e é composto de 9,94 milhões de hectares de plantios renováveis, onde há a predominância de espécies dos gêneros *Eucalyptus* e *Pinus* (IBÁ, 2023). Este setor possui dependência do meio ambiente, o que, por sua vez, acarreta em sérios riscos à produção, demandando um constante monitoramento. Dessa forma, as diferentes áreas da ciência florestal como silvicultura, economia e, principalmente, o manejo buscam fornecer informações que auxiliem de forma segura a tomada de decisões (Buongiorno; Gilles, 2003).

Dada a demanda por informações confiáveis, o conhecimento sobre os processos e fatores que influenciam a produção florestal é fundamental regular a continuidade de suprimentos de matéria-prima de qualidade para a indústria e aumentar o poder competitivo no setor florestal (Silva *et al.*, 2021). Este tipo de informação, capaz de nortear o planejamento das atividades e estratégias econômicas de comercialização, são obtidos por meio de levantamentos, como o inventário florestal (Campos; Leite, 2013; Oliveira *et al.*, 2016).

O inventário florestal é uma etapa essencial para o gerenciamento dos recursos florestais (Chen *et al.*, 2019), onde são mensurados os atributos diâmetro a 1,3 m de altura (DAP) e altura total (H), que são essenciais em diversas etapas como estimativa do índice de sítio, previsão da produtividade, projeção do crescimento, mensuração do estoque de biomassa e carbono e outros parâmetros importantes (Curtis 1967, Adame *et al.* 2008; Karatepe *et al.*, 2022; Dong *et al.*, 2020; Trim *et al.*, 2020; Yang; Burkhart, 2020; Soares; Tomé, 2002; Mensah *et al.*, 2018).

As medidas de DAP podem ser obtidas de forma rápida, prática e precisa no campo (Sharma; Parton, 2007), enquanto que, a mensuração das alturas exige maior tempo e custo, não é um método de obtenção direta e é altamente afetada pela dificuldade de visualização completa da árvore em povoamentos adensados, o que pode gerar erros consideráveis de medição (Castaño-Santamaría *et al.*, 2013; Raptis *et al.*, 2021).

Desta forma, a previsão da altura se torna uma etapa extremamente relevante, sendo frequentemente realizada por meio do uso de modelos que relacionam a altura e o DAP (Peng *et al.*, 2004). No entanto, embora apenas o diâmetro seja frequentemente utilizado na previsão da altura, há evidências de que variáveis adicionais, como altura dominante, idade, índice de sítio (Lynch; Murphy, 1995; Wang; Hann, 1988), densidade de plantio, área basal (Larsen; Hann, 1987; Parresol, 1992; Temesgen *et al.*, 2004), diâmetro quadrático médio (Saud *et al.*, 2009), podem contribuir de forma relevante na melhoria de modelos hipsométricos, podendo estes serem aplicados para uma ampla gama de dados oriundos de diferentes locais e idades, ao

contrário do ajuste individual de relações hipsométricas por unidade amostral (UA) (Temesgen *et al.*, 2004).

Na literatura podem ser encontrados diferentes metodologias para descrever a relação hipsométrica para diferentes espécies florestais (Lin *et al.*, 2022; Özçelik *et al.*, 2013; Karatepe *et al.*, 2022; Raptis *et al.*, 2021), porém não há um consenso sobre qual método é mais adequado. Os modelos de regressão, pelo método dos mínimos quadrados ordinários (OLS), foram amplamente utilizados no passado, devido à facilidade de ajuste e o bom estabelecimento da teoria dos modelos lineares (Curtis, 1967). No entanto, os atributos florestais estão, geralmente, agrupados em estruturas com mais um nível, como as árvores dentro de diferentes UAs e as UAs dentro de diferentes regiões, comportamento este que a técnica OLS não é capaz de explicar (Pinheiro; Bates, 2000).

Uma forma de lidar com dados estruturados de forma hierárquica é a utilização de modelos de efeitos mistos. Estes modelos consideram as diferenças entre grupos, podendo melhorar a qualidade das previsões (Zhang *et al.*, 2015; Timilsina; Staudhammer, 2013). Esta técnica vem sendo cada vez mais utilizada na área florestal, como no ajuste de relações hipsométricas (Mehtätalo, 2004; Trincado *et al.*, 2007; Crecente-Campo, 2010; Castaño-Santamaría *et al.*, 2013; Bronisz; Mehtätalo, 2020; Raptis *et al.*, 2021).

Além dos modelos mistos, os modelos de Rede neurais artificiais (RNA) estão se tornando uma ferramenta de previsão cada vez mais utilizada, por não depender de pressuposições sobre a forma de uma função de ajuste, de forma que, a RNA é treinada para encontrar a relação existentes entre as variáveis de entrada e saída, sendo adequadas para problemas de modelagem em que: 1) há uma grande quantidade de dados e depende da interação de múltiplos parâmetros, 2) a área em estudo possui grande quantidade de dados históricos, 3) os dados disponíveis estão incompletos e 4) a função que determina a solução é desconhecida ou demanda grande esforço para ser encontrada (Bailey; Thompson, 1990; Leduc *et al.*, 2001; Diamantopoulou, 2006). Maioria dessas condições são encontradas nos problemas de modelagem florestal, de modo que, pesquisadores têm desenvolvido trabalhos com a aplicação de RNA na resolução destes problemas (Özçelik *et al.*, 2010; Diamantopoulou *et al.*, 2015; Lima *et al.*, 2021), como em ajustes hipsométricos (Diamantopoulou, 2012; Soares *et al.*, 2021; Casas *et al.*, 2022).

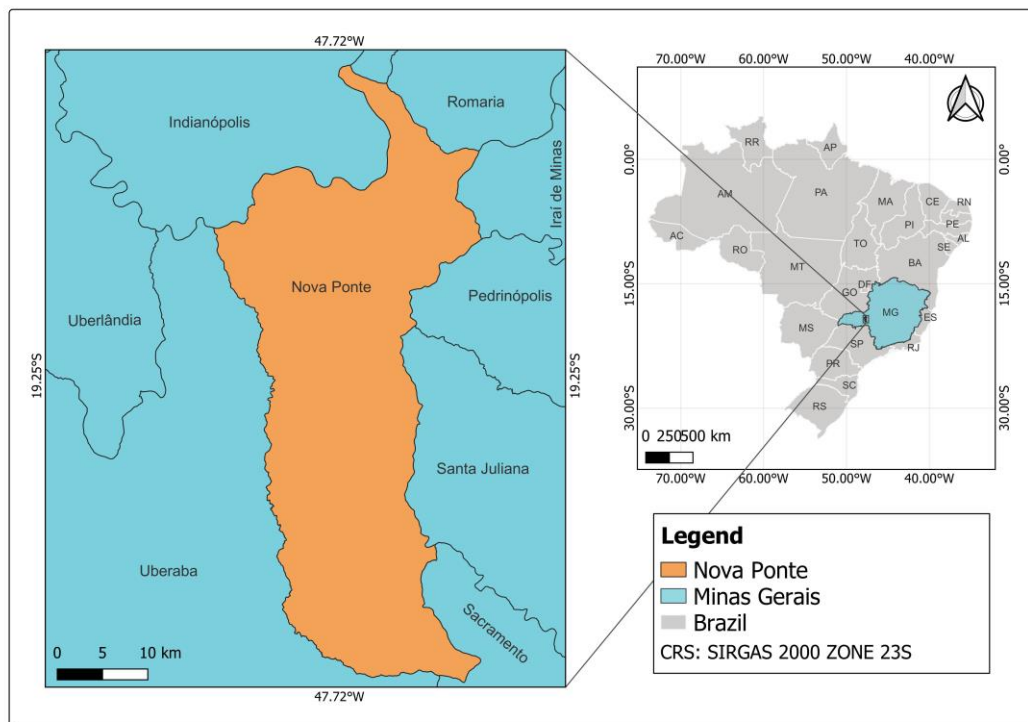
Deste modo, o presente estudo teve como objetivo comparar, a partir de medidas estatísticas de qualidade, o desempenho de modelos não lineares de efeitos fixos e mistos e RNA locais e genéricos na previsão da altura total de árvores de *Pinus caribaea* var. *hondurensis*.

2 MATERIAL E MÉTODOS

2.1 Caracterização da área de estudo

Os dados utilizados no presente estudo foram coletados em uma área com, aproximadamente, 243 ha de plantios de *Pinus caribaea* var. *hondurensis*, estabelecidos em espaçamento de 2,5 x 2 m, pertencentes à empresa Caxuana, situada no município de Nova Ponte, Minas Gerais (Figura 1). O local possui precipitação média de 1.520 mm por ano, temperatura anual média de 22,2 °C, 980 m de altitude média e umidade relativa média igual a 51%. O relevo predominante do local é do tipo plano a suave ondulado e as classes de solo são dos tipos LATOSSOLO Vermelho escuro e LATOSSOLO Vermelho-amarelo (Horle *et al.*, 2010).

Figura 1 – Mapa de localização do município de Nova Ponte, MG.



Fonte:Do autor (2024).

2.2 Base de dados

Foram utilizados dados de 602 unidades amostrais (U.A.), com área variável de 630 a 642 m² e, em cada U.A., mensurou-se todos os DAP de todas as árvores e as H das árvores existentes nas duas primeiras linhas (média de 30 árvores). Além disso, as árvores dominantes e codominantes também foram medidas seguindo o conceito de Assmann (1970).

De posse dos dados, foram estimados a área basal por hectare (AB), o diâmetro quadrático médio (Dq) e altura e diâmetro médios das árvores dominantes e codominantes (Hdom e Ddom) para cada U.A (Tabela 1). Além disso, a base de dados foi separada em base de treino (80% = 479 UAs) e teste (20% = 123 UAs) para ajuste e avaliação dos modelos, respectivamente.

Tabela 1 – Tabela descritiva dos dados.

	TRAIN (plot= 479; n= 14505)					TEST (plot=123; n= 3784)				
	Mean	Max.	Min.	Std. Dev.	CV (%)	Mean	Max.	Min.	Std. Dev.	CV (%)
D (cm)	15,78	36,00	6,05	4,99	31,62	15,95	32,00	6,05	4,97	31,18
H (m)	10,93	22,60	2,50	3,14	28,70	11,26	21,00	3,20	3,18	28,23
Age (years)	6,54	10,50	4,42	1,44	22,04	6,59	10,50	4,42	1,49	22,55
BA (m ² /ha)	31,10	70,96	8,50	10,72	34,46	32,19	58,07	12,61	9,60	29,81
Dq (cm)	16,24	27,02	8,50	3,19	19,66	16,42	24,85	10,14	3,10	18,91
Hdom (m)	12,93	21,57	5,67	2,76	21,32	13,26	19,86	6,79	2,82	21,25
Ddom (cm)	20,63	31,06	10,23	3,98	19,32	20,86	29,28	13,32	3,92	18,79

2.3 Modelos não lineares de efeitos fixos

Foram ajustados cinco modelos hipsométricos (Tabela 2) não lineares de efeitos fixos, com dois e três parâmetros (Tabela 2), utilizando o método dos quadrados mínimos ordinário não linear (ONLS – *ordinary nonlinear least squares*), com auxílio da função “nls” utilizando a linguagem R. Todos os cinco modelos não lineares possuem um comportamento assintótico e foram ajustados à base de dados para determinação do modelo mais adequado para os plantios de *Pinus caribaea* var. *hondurensis* em estudo.

Tabela 2 – Modelos não lineares avaliados.

Model	Function form	References
Ratkowsky	$H_{ij} = \beta_0 \left(1 - \exp(-(D_{ij} + \beta_1)\beta_2)\right) + 1.3 + \epsilon_{ij}$	Citado por Ratkowsky (1990)
Power	$H_{ij} = \beta_0 D_{ij}^{\beta_1} + 1.3 + \epsilon_{ij}$	Stage (1975)
Curtis	$H_{ij} = \frac{\beta_0 D_{ij}}{(1 + D)^{\beta_1}} + 1.3 + \epsilon_{ij}$	Curtis (1967)
Micmen	$H_{ij} = \frac{\beta_0 D_{ij}}{\beta_1 + D_{ij}} + 1.3 + \epsilon_{ij}$	Bates and Watts (1980); Huang <i>et al.</i> (1992)
Mitscherlich	$H_{ij} = \beta_0 \left(1 - \exp(-\beta_1 D_{ij})\right) + 1.3 + \epsilon_{ij}$	Levin e Nitsan (1964); Shen <i>et al.</i> (2020)

Onde β_0 , β_1 e β_2 são parâmetros a serem estimados; D_{ij} é o diâmetro à altura de 1,3 m do solo da árvore j na parcela i e H_{ij} é a altura total da árvore j na parcela i ; ϵ_{ij} é o erro aleatório, assumindo se independente entre as observações e $N(0, \sigma_\epsilon^2)$.

2.4 Modelos não lineares de efeitos fixos genéricos

Neste estudo, buscou-se desenvolver um novo modelo genérico utilizando como base o modelo selecionado na etapa anterior que recebeu a adição de variáveis a nível de UA como covariáveis, a fim se obter um modelo represente de forma mais adequada a variabilidade entre as UA. Nesta etapa, foram testadas diferentes relações dos parâmetros da equação com as variáveis a nível de UA e diferentes combinações destas variáveis por equação.

Uma das etapas mais importantes durante o desenvolvimento de um novo modelo hipsométrico genérico foi a seleção das variáveis adicionais a nível de UA, desta forma foram utilizadas variáveis que possuem potencial de influência na relação hipsométrica, como índice de sítio (IS), área basal (AB), diâmetro quadrático médio (Dq), idade (I), altura dominante (Hdom), diâmetro dominante (Ddom). Ao final desta etapa, o modelo de melhor desempenho foi selecionado, com base em estatísticas a serem mencionadas posteriormente.

2.5 Modelos não lineares de efeito misto

Para descrever a relação hipsométrica, por meio dos modelos mistos, foi incorporada, ao modelo selecionado anteriormente, a variabilidade entre parcelas. A altura da j -ésima árvore na i -ésima parcela foi modelada como:

$$H_{ij} = f(\boldsymbol{\beta}_{ij}, DAP_{ij}) + \varepsilon_{ij}, \quad (1)$$

$$i = 1, \dots, M, j = 1, \dots, n_i, \quad (2)$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \quad (3)$$

Onde $\boldsymbol{\beta}_{ij}$ é um vetor de p parcelas e possíveis parâmetros específicos por árvore, f é uma função (função do modelo selecionado) que relaciona a altura de uma árvore com seu diâmetro e ε_{ij} é o erro entre parcelas que foi considerado independente e com distribuição normal.

O k -ésimo elemento do vetor de parâmetros $\boldsymbol{\beta}_{ij}$ foi então modelado como uma função linear de efeitos fixo e aleatório:

$$\beta_{ij} = \mathbf{x}'_{ijk} \boldsymbol{\beta}_k + \mathbf{z}'_{ijk} \mathbf{b}_{ik}, \quad (4)$$

$$k = 0, \dots, p - 1, \quad (5)$$

$$\mathbf{b}_{ik} \sim N(\mathbf{0}, \sigma_k^2) \quad (6)$$

Onde $\boldsymbol{\beta}_k$ é um vetor de efeitos fixos e \mathbf{b}_{ik} é um vetor de efeitos aleatórios associados com a i -ésima parcela. Assume-se que \mathbf{b}_{ik} tem distribuição normal, com média 0 e matriz de covariância σ_k^2 e é independente de ε_{ij} . \mathbf{x}'_{ijk} e \mathbf{z}'_{ijk} possivelmente com a j -ésima árvore.

Na forma matricial, temos:

$$\mathbf{H}_i = f(\boldsymbol{\beta}_i, DAP_i) + \boldsymbol{\varepsilon}_i \quad (7)$$

$$\boldsymbol{\beta}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \quad (8)$$

$$\mathbf{b}_i \sim N(\mathbf{0}, \sigma_k^2) \quad (9)$$

Combinando essas duas equações, temos:

$$\mathbf{H}_i = f(\boldsymbol{\beta}, \mathbf{b}_i, \mathbf{X}_i, \mathbf{Z}_i) + \boldsymbol{\varepsilon}_i \quad (10)$$

Onde \mathbf{H}_i é o vetor de alturas das parcelas i , $\boldsymbol{\beta}_i$ é o vetor de parâmetros a nível de parcela, DAP_i é o vetor de diâmetros das parcelas i , $\boldsymbol{\varepsilon}_i$ é o vetor de erros entre parcelas, $\boldsymbol{\beta}$ é o vetor de parâmetros fixos que não dependem das parcelas, \mathbf{b}_i é o vetor de efeitos aleatórios a nível de parcela e \mathbf{X}_i e \mathbf{Z}_i são as matrizes das variáveis independentes.

Dessa forma, foi adicionado o efeito aleatório, resultando em equações com apenas a inclusão do DAP e com variáveis a nível de parcela inclusas. A inclusão dos efeitos aleatórios foi realizada apenas no nível de UA. Assim como no estudo de Gómez-García *et al.* (2014), a

autocorrelação temporal, resultante de medidas repetidas ao longo do desenvolvimento das curvas hipsométricas, não foi considerada, pois a quantidade de ocasiões de medições por UA (1,64) era pequena.

2.6 Modelos não lineares de efeitos mistos genéricos

A fim de aumentar ainda mais a precisão das predições dos modelos genéricos, incorporou-se, em pelo menos um parâmetro do modelo genérico selecionado, o efeito aleatório a nível de UA, de modo que resultasse em uma equação genérica de efeitos mistos.

2.7 Calibração dos modelos de efeito misto

Uma vantagem dos modelos de efeitos mistos é que, havendo a disponibilidade de, no mínimo, uma observação dos dados, estes podem ser utilizados para a predição do vetor de efeitos aleatórios (Özçelik *et al.*, 2013; Vonesh; Chinchilli, 1997). Este método pode ser chamado de calibração ou localização e, na área florestal, esta técnica pode ser empregada para a estimação dos parâmetros aleatórios de uma nova UA que não estava presente anteriormente na base de ajuste do modelo (Hall; Bailey 2001).

No presente trabalho, a calibração foi realizada com a utilização de, no mínimo, os valores dos atributos de uma árvore presente na nova UA para estimação dos novos parâmetros aleatórios. Foi utilizado o algoritmo de Lindstrom-Bates por meio do método FOCE para o processo de calibração, onde os valores do vetor de parâmetros aleatório \mathbf{b}_i , associado à UA, pode ser estimado a partir da equação abaixo (Camacho *et al.*, 2022; Vonesh; Chinchilli, 1997).

$$\mathbf{b}_i = \widehat{\mathbf{D}}\widehat{\mathbf{Z}}_i^T(\widehat{\mathbf{R}}_i + \widehat{\mathbf{Z}}_i\widehat{\mathbf{D}}\widehat{\mathbf{Z}}_i^T)^{-1}\widehat{\mathbf{e}}_i \quad (11)$$

Onde $\widehat{\mathbf{D}}$ é a matriz de variância e covariância $q \times q$ associada com os parâmetros aleatórios (q é o número de parâmetros aleatórios no modelo) de todas as UA, $\widehat{\mathbf{R}}_i$ é a matriz de variância e covariância $m \times m$ para a variabilidade dentro das UA, $\widehat{\mathbf{e}}_i$ é o vetor de resíduos $m \times 1$ (onde é calculado pela diferença entre as alturas observadas de cada árvore dentro da subamostra e os valores preditos a partir do modelo contendo apenas os efeitos fixos), $\widehat{\mathbf{Z}}_i$ é a matriz $m \times q$ das derivativas parciais dos parâmetros aleatórios avaliados em $\widehat{\mathbf{b}}_i$. Uma vez que $\widehat{\mathbf{b}}_i$ é predito, os valores das alturas podem ser calculados a partir do modelo calibrado.

Para os modelos de efeitos mistos local e genérico selecionados, as predições calibradas foram obtidas e avaliadas para diferentes esquemas de amostragem e diferentes tamanhos das subamostras dentro de cada UA da base de teste (20% = 123 UA). As alternativas de calibração

foram: 1) Altura total de 1-9 maiores árvores por UA. 2) Altura total de 1-9 menores árvores por UA. 3) Altura total de 1-9 árvores médias por UA. 4) Altura total de 3, 6 e 9 árvores por UA, conforme a distribuição diamétrica, estratificada em 3 classes, da UA. 5) Sem calibração, utilizando apenas os parâmetros de efeito fixo para a predição e 6) Calibração completa, utilizando todas as árvores da base de teste, para fins de comparação com as demais calibrações.

2.8 Redes Neurais Artificiais

Modelos de redes neurais artificiais foram desenvolvidos para estimar as alturas totais das árvores neste estudo. Há diversas variáveis como o número de camadas, algoritmo de aprendizagem, funções de ativação, número de neurônios nas camadas ocultas e o tamanho das bases de dados utilizadas para treinamento, validação e teste para definir qual a melhor arquitetura da rede neural utilizar (Sakici; Ozdemir, 2018).

O treinamento ocorre com mais facilidade quando os dados utilizados estão normalizados para um pequeno intervalo próximo de zero (Jayalakshmi; Santhakumaran, 2011). A normalização dos dados é capaz de acelerar o processo de treinamento e reduzir viés. Existem alternativas de normalização de dados distintas, sendo uma delas a normalização min-max, que converte os dados para uma escala desejada, considerando os valores mínimos e máximos de cada variável, de modo a manter a distribuição original dos dados (Lantz, 2019; Valença, 2007). No presente estudo, a normalização min-max foi aplicada para todas as variáveis de entrada e saída, utilizando a equação 12 (Soares *et al.*, 2011). Desse modo, todas as variáveis passaram para a escala de 0 a 1. Como os dados normalizados foram utilizados no treinamento das redes, as saídas das redes precisam ser desnormalizadas para que se possa obter os valores originais de saída. A desnormalização dos valores de saída foi obtida por meio da equação 13.

$$x' = \frac{(x - x_{min}) * (b - a)}{(x_{max} - x_{min})} + a \quad (12)$$

$$x = \frac{(x' - a)(x_{max} - x_{min})}{(b - a)} + x_{min} \quad (13)$$

Em que: x' = valor normalizado; x = valor original; x_{min} = valor mínimo da variável; x_{max} = valor máximo da variável; a = limite inferior do intervalo de normalização (0); b = limite superior do intervalo de normalização (1).

Para o treinamento das redes neurais, o pacote *neuralnet* (Günther; Fritsch, 2010) foi utilizado, em que se considerou as RNA do tipo *Multilayer Perceptron* (MLP) (Bishop, 1995).

Além disso, foram consideradas uma camada de entrada, uma camada intermediária (camada oculta) e uma camada de saída, sendo o valor predito. A quantidade de neurônios nas camadas intermediárias foi definido a partir da equação 14 (Júnior *et al.* 2021; Sahin *et al.*, 2023).

$$2n^{0,5} + n_2 \leq n_1 \leq 2n + 1 \quad (14)$$

Em que n é o número de variáveis de entrada na RNA, n_1 é o número de neurônios na camada oculta e n_2 é o número de neurônios na camada de saída.

Com base no número da definição da quantidade mínima e máxima do número de neurônios na camada oculta, testou-se todas as configurações para uma camada oculta. Também foram testadas as funções de ativação logística (16) e tangente hiperbólica (17), nas camadas ocultas, e a função linear na camada de saída (18).

$$f(s) = \frac{1}{1 + e^{(-s)}} \quad (15)$$

$$f(s) = \frac{1 - e^{(-2s)}}{1 + e^{(-2s)}} \quad (16)$$

$$f(s) = s \quad (17)$$

Em que s é o valor a ser fornecido para a função de ativação.

Três tipos de RNA foram ajustados: 1) ANN_local, que utiliza apenas o DAP como variável de entrada, 2) ANN_generalized, após testes prévios com variáveis a nível de UA, além do DAP e Hdom, utiliza o Ddom ou Dq como variáveis de entrada e 3) ANN_v, que, além do DAP, foi utilizada uma medida de variabilidade por UA, a variável transformada $(D - D_{\text{resíduo padrão}}) / D_{\text{médio}}$, como proposto por Özçelik *et al.* (2013). Sendo então, a ANN_local comparável com a equação ajustada pelo método ONLS e ANN_generalized e ANN_v comparáveis com as equações genérica e mistas, respectivamente. Todas as combinações testadas podem ser observadas na tabela 3.

Tabela 3 – Configurações de RNA avaliadas.

Model	Neurons	Layers	Inputs	Output	Activation function		Total combinations
					Hidden layer	Output layer	
ANN_local	3	1	DAP	H	Logistic, Tanh	Logistic, Tanh, Linear	6
ANN_generalized	5 to 7	1	DAP, Hdom, Dq	H	Logistic, Tanh	Logistic, Tanh, Linear	18
ANN_v	4 to 5	1	DAP, DAP_V	H	Logistic, Tanh	Logistic, Tanh, Linear	12
Total							36

Onde $DAP_V = (D - D_{\text{resíduo padrão}}) / D_{\text{médio}}$.

O algoritmo utilizado no ajuste dos parâmetros da RNA foi o *Resilient Backpropagation with weigh backtracking*, que é baseado no algoritmo *Backpropagation*, ajustando os pesos da rede neural com o objetivo de minimizar o erro, onde diferentes taxas de aprendizado são utilizadas para cada peso, como é abordado em mais detalhes em Riedmiller (1994).

O critério de parada das redes neurais, durante o treinamento, foi atingido pelo número máximo de 100.000 iterações ou quando todas as derivativas parciais absolutas da função de erro quadrático médio (E) em relação aos pesos (w) ($\partial E/\partial w$) fossem menores que 1%. Uma vez que uma dessas condições foi atendida, a etapa de treinamento foi concluída e a melhor configuração de parâmetros foi selecionada, baseando-se no menor erro quadrático médio. Porém, os pesos iniciais de cada treinamento são valores retirados aleatoriamente de uma distribuição normal padrão e, para assegurar a seleção de uma rede mais adequada, cada rede foi treinada três vezes e a de menor erro quadrático médio foi considerada.

Após o ajuste de todas as possíveis estruturas das redes neurais, todas foram comparadas entre si, considerando os desempenhos para as bases de treinamento e teste simultaneamente.

Para a melhor RNA de cada tipo, foi extraída sua equação não linear para predição das alturas totais das árvores, como sugerido por Dantas *et al.* (2023). Para isso, um sistema de equações, com os coeficientes sendo os pesos obtidos para cada neurônio durante o treinamento, foi desenvolvido. Dessa forma, os sistemas podem ser utilizados para predição de altura de novas árvores que não foram fornecidas durante o treinamento da RNA. O modelo 18 define a relação entre a camada oculta e a variável dependente, sendo β_0 o *bias* e β_1 – β_n os pesos referentes a cada neurônio e n o número de neurônios na camada oculta. Os modelos 19 e 20 se referem à possível função de ativação utilizada em cada neurônio da camada oculta, podendo ser, neste estudo, a logística ou a tangente hiperbólica. Por fim, o modelo 21 representa o resultado da relação entre k variáveis preditoras e seus respectivos neurônios da camada oculta, resultando em uma equação para cada neurônio.

$$Y' = \beta_0 + \beta_1 * z_1 + \beta_2 * z_2 + \dots + \beta_n * z_n \quad (18)$$

$$z_n = \frac{1}{1 + e^{(-w_i)}} \quad (19)$$

$$z_n = \frac{1 - e^{(-2w_n)}}{1 + e^{(-2w_n)}} \quad (20)$$

$$w_n = \beta_{0,n} + \beta_{1,n} * X_1' + \beta_{2,n} * X_2' \dots \beta_{k,n} * X_k' \quad (21)$$

Onde Y' é a variável dependente normalizada (sendo a altura da árvore normalizada neste trabalho), β_0 é o *bias*, β_n é o coeficiente associado com o neurônio n , $\beta_{k,n}$ é o coeficiente entre a variável de entrada k e o neurônio n , z_n é a resposta do n -ésimo neurônio da camada oculta após transformação pela função de ativação, w_i é a soma dos produtos entre as variáveis de entrada e seus respectivos pesos do n -ésimo neurônio.

2.9 Avaliação das equações ajustadas

O desempenho dos diferentes tipos de modelos testados foi avaliado com base na raiz do erro quadrático médio (RMSE – Root Mean Squared Error), erro médio absoluto percentual (MAPE – Mean Absolute Percentage Error), desvio médio (BIAS), índice de eficiência do modelo (EF - model efficiency) e Critério de informação de Akaike (AIC) (equações 22-26), conforme estudado por Raptis *et al.* (2021), Camacho *et al.* (2022) e Ogana e Ercanli (2022).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (h_{obs} - h_{est})^2}{n - p}} \quad (22)$$

$$MAPE (\%) = \frac{\sum_{i=1}^n \frac{|h_{obs} - h_{est}|}{h_{obs}}}{\bar{h}_{obs}} \times 100 \quad (23)$$

$$BIAS = \frac{\sum_{i=1}^n (h_{obs} - h_{est})}{n} \quad (24)$$

$$EF = 1 - \frac{(n - 1) \sum_{i=1}^n (h_{obs} - h_{est})^2}{(n - p) \sum_{i=1}^n (h_{obs} - \bar{h}_{obs})^2} \quad (25)$$

$$AIC = -2l(\hat{\theta}) + 2p \quad (26)$$

Em que: h_{obs} = altura observada; h_{est} = altura estimada; \bar{h}_{obs} = altura média observada; n = número de observações; $l(\hat{\theta})$ = logaritmo da verossimilhança maximizada; p = número de parâmetros ou pesos (para as redes neurais).

Para comparação dos modelos em relação à qualidade das predições, tanto para a base de treino quanto para a de teste, o método de ranking relativo (27) (Poudel; Cao, 2013; Sahin *et al.*, 2023) foi utilizado para a avaliação simultânea das cinco métricas utilizadas.

$$R_i = 1 + \frac{(m - 1) - (S_i - S_{min})}{S_{max} - S_{min}} \quad (28)$$

Onde R_i = valor do *ranking* relativo do i -ésimo modelo avaliado ($i = 1, 2, 3 \dots m$).

S_i valor da métrica estatística do modelo i . S_{min} e S_{max} = valores mínimo e máximo de S_i , respectivamente. m = quantidade de modelos sendo comparados.

Este método permite que a melhor e pior técnicas avaliadas possuam um ranking relativo de 1 e m , respectivamente. O *rank* das demais técnicas avaliadas foi expressado por um valor real entre 1 e m e, uma vez que, além da posição no *ranking*, o tamanho de S_i também é considerado, este método fornece mais informações que o método tradicional de *ranking* que utiliza apenas valores inteiros referentes às posições das métricas. Este método considera como melhor, a métrica de menor valor, dessa forma, para as métricas em que o melhor resultado é o maior valor, como o EF, os valores foram multiplicados por -1 e, para o BIAS, em que o melhor valor é o mais próximo de zero, foi utilizado o módulo desses valores. Após o ajuste dos modelos de efeito fixo, o modelo com melhor desempenho foi selecionado para ser utilizado como modelo base nas etapas seguintes.

Para comparação dos modelos mistos entre si, foi utilizado o critério de Informação de Akaike (AIC) (22), o critério Bayesiano de Schwarz (BIC) (29), o logaritmo da verossimilhança maximizada (LogLik) e o teste da razão de verossimilhança (MLRT). O AIC é obtido a partir do valor da verossimilhança e na quantidade de parâmetros, além de derivar-se da informação ou distância de Kullback-Leibler, que verifica a diferença entre dois modelos (Akaike, 1973; Floriano *et al.*, 2006), com base nesta métrica, o melhor modelo apresentará o menor valor de AIC. O Critério Bayesiano de Schwarz (BIC) é uma métrica que auxilia na seleção de modelos e, geralmente, é empregada juntamente com o AIC e descreve a associação entre a variável dependente e as variáveis independentes, aumentando as chances de se escolher o modelo mais consistente dentre os avaliados. Avaliando este critério, deve-se optar pelo modelo que apresentar o menor valor de *BIC*.

$$BIC = -2l(\hat{\theta}) + 2p \ln(n) \quad (29)$$

Em que: \ln = logaritmo neperiano; n = número de observações.

O MLRT teve como objetivo avaliar o resultado da inclusão do efeito aleatório nos modelos (Resende *et al.*, 2014) e foi feito verificando a significância da diferença entre os valores de *deviance* ($-2l(\hat{\theta})$) dos modelos com e sem a incorporação da variável aleatória,

comparando o valor calculado com o valor tabelado, a partir do teste qui-quadrado (χ^2), a 5% de significância.

Por fim, para validar a eficácia dos métodos selecionados, o teste t pareado foi aplicado, comparando os valores observados com os valores preditos pelas metodologias propostas, utilizando a base de teste. Este teste considera a hipótese nula de que “não há diferença significativa” entre os valores preditos e observados contra a hipótese alternativa de que há “diferença significativa” entre esses valores (Sahin *et al.*, 2023). Dessa forma, aceitando a hipótese nula significa que as predições resultantes dos modelos podem ser consideradas satisfatórias. Além disso, gráficos de distribuição dos resíduos foram construídos para auxiliar na avaliação dos modelos avaliados.

3 RESULTADOS

3.1 Modelos não lineares de efeitos fixos

As cinco equações não lineares locais (utilizando apenas o DAP como variável independente) foram ajustadas pelo método ONLS a partir da base de treino, enquanto a capacidade preditiva foi avaliada utilizando a base de teste. Os parâmetros e as estatísticas de desempenho das equações são apresentadas nas tabelas 4 e 5.

Tabela 4 – Estatística paramétrica das equações não lineares.

Parameter	Value	Standard error	t-value	p-value
Ratkowsky				
b0	103,0434	32,5716	3,1636	0,0016
b1	2,4869	0,3412	7,2896	0,0000
b2	0,0054	0,0019	2,8488	0,0044
Power				
b0	1,0086	0,0155	64,9733	0,0000
b1	0,8202	0,0053	153,6254	0,0000
Curtis				
b0	1,0524	0,0176	59,9167	0,0000
b1	0,1907	0,0057	33,5779	0,0000
Micmen				
b0	59,9594	1,8596	32,2428	0,0000
b1	81,3297	3,0794	26,4113	0,0000
Mitscherlich				
b0	34,2844	0,9433	36,3435	0,0000
b1	0,0212	0,0007	29,8160	0,0000

Onde foram considerados significantes os parâmetros com valor $p \leq 0,05$.

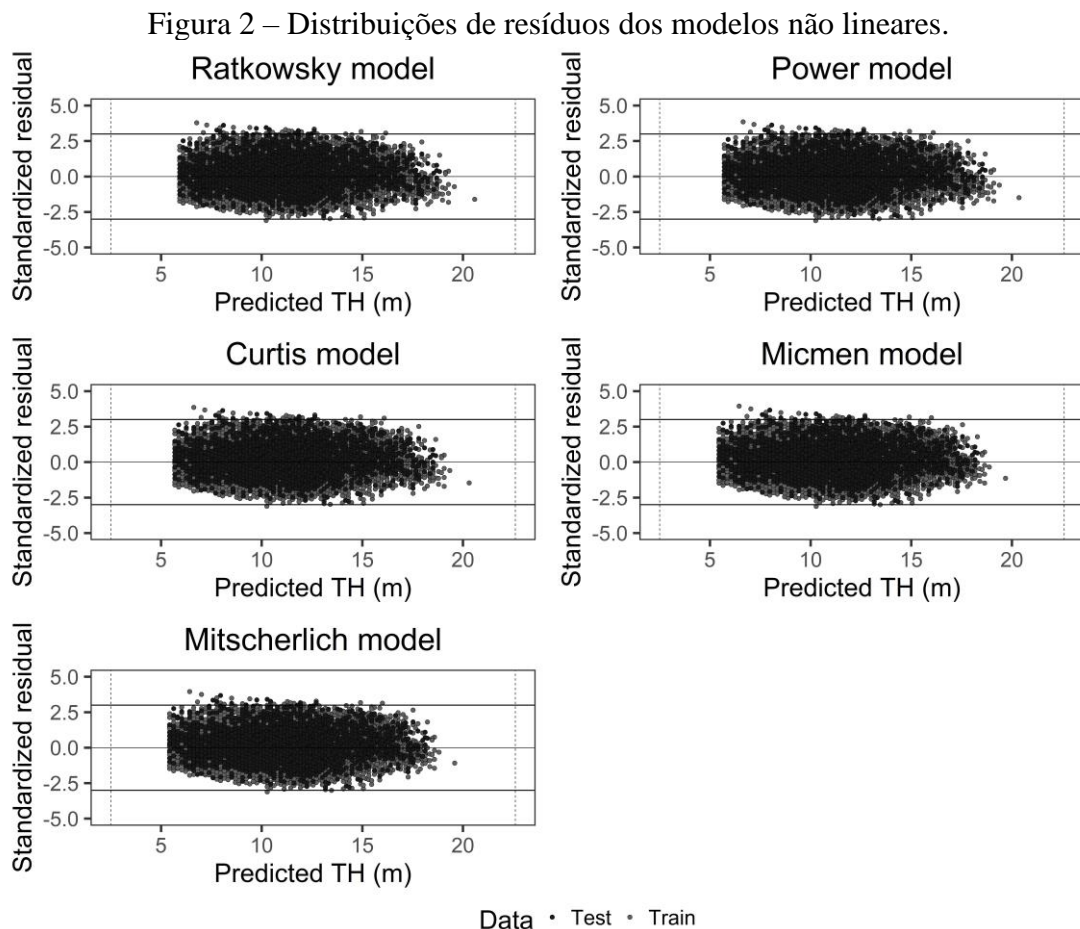
Tabela 5 – Desempenho estatístico das equações não lineares.

Model	TRAIN (plot= 479; n= 14505)					TEST (plot=123; n= 3784)					Overall rank
	RMSE	MAPE	EF	BIAS	AIC	RMSE	MAPE	EF	BIAS	AIC	
Ratkowsky	1,90	13,79	0,6329	0,0000	59783	1,96	13,66	0,6 211	0,2473	15821	1,00
Power	1,90	13,80	0,6326	0,0024	59794	1,96	13,69	0,6 202	0,2486	15829	1,78
Curtis	1,90	13,81	0,6326	0,0029	59797	1,96	13,70	0,6 200	0,2490	15830	1,94
Micmen	1,90	13,84	0,6311	0,0117	59853	1,97	13,77	0,6 177	0,2555	15854	4,56
Mitscherlich	1,91	13,84	0,6309	0,0138	59862	1,97	13,77	0,6 173	0,2573	15857	5,00

De modo geral, modelos utilizados no presente estudo foram capazes de explicar a variação da altura total com EF de 0,63 para a base de treino e 0,62 para a base de teste. Além disso, estes modelos resultaram em erros médios absolutos de 13,79% a 13,84% e 13,66% a 13,77% para as bases de treino e teste, respectivamente.

A equação Ratkowsky resultou nos menores valores de RMSE, MAPE, BIAS e AIC (1,90; 13,79; 0,0000 e 59783, respectivamente) e maior valor de EF (0,6329) para os dados de treinamento. Já para a base de teste, a equação resultou nos menores valores de RMSE, MAPE, BIAS e AIC (1,96; 13,66; 0,2473 e 15821, respectivamente) e no maior valor de EF (0,6211). Além disso, de acordo os valores do *ranking*, os modelos Power e Curtis, mesmo tendo um desempenho inferior, resultaram em métricas estatísticas semelhantes ao modelo Ratkowsky.

Analisando as distribuições de resíduos dos modelos não lineares (Figura 2), percebe-se que todas as equações resultaram em distribuições de comportamento semelhante. Além disso, maior parte dos valores de resíduo padronizado (>99%) estão entre o intervalo de -3 a 3, indicando um bom comportamento das funções na predição da altura.



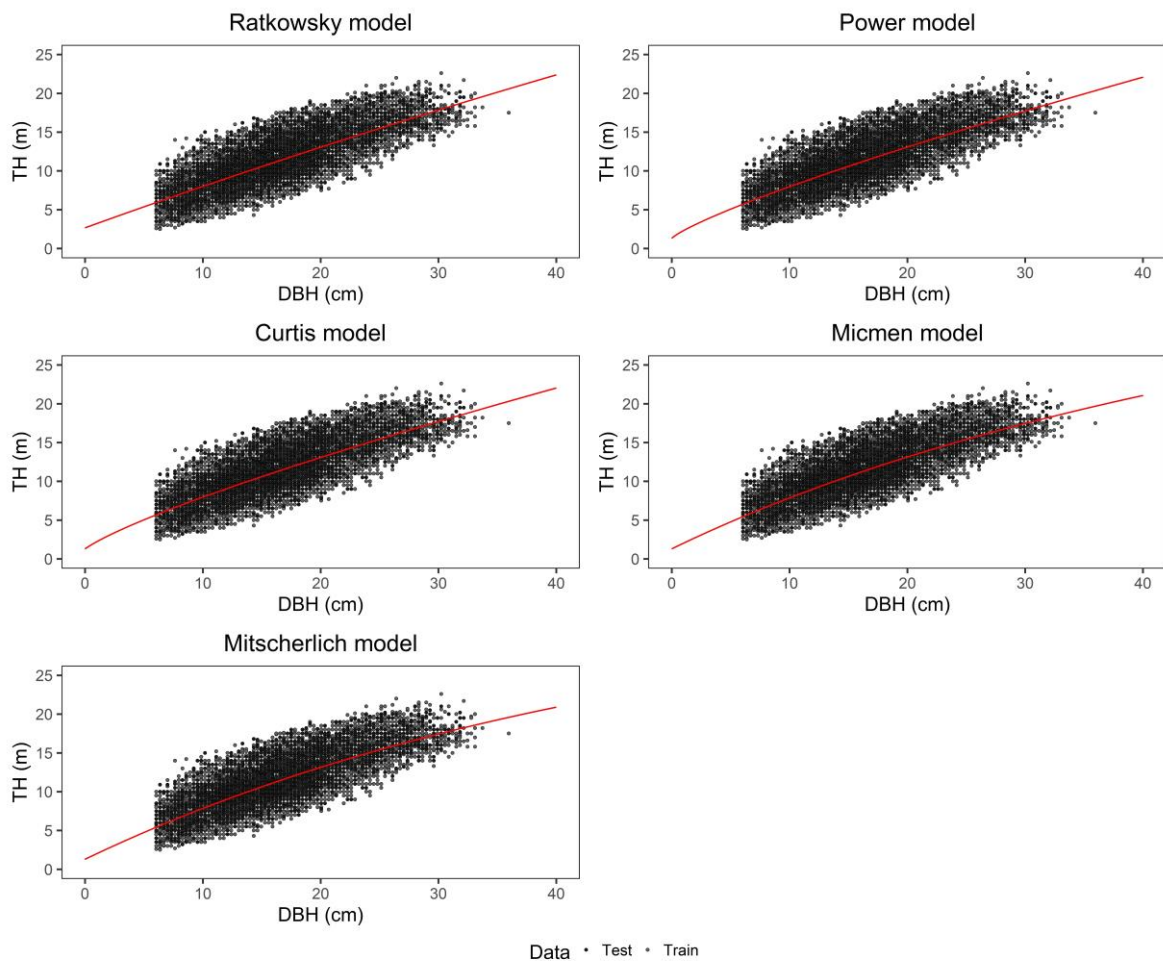
Onde as linhas tracejadas referem-se à amplitude dos valores observados.

Fonte: Do autor (2024).

Ainda, aos gráficos de distribuição de resíduos, também foram adicionados, por meio de linhas tracejadas, os valores mínimo e máximo observados na base de dados, sendo possível observar que as previsões ficaram restringidas a um intervalo menor, em relação aos dados reais. Resumindo, a equação Ratkowsky foi a que resultou no menor valor de *ranking*, indicando ser a equação mais adequada para a base de dados em estudo. Dessa forma, o modelo Ratkowsky foi selecionado como o melhor modelo não linear local de estrutura fixa para estimar a altura total no presente estudo, sendo assim considerado como o modelo de referência para as etapas posteriores.

As curvas geradas pelos diferentes modelos (Figura 3) evidenciam um comportamento assintótico, típico da relação hipsométrica, passando pelo centro dos valores observados.

Figura 3 – Curvas dos modelos.



Fonte: Do autor (2024).

3.2 Modelos não lineares de efeitos fixos genéricos

Uma vez que o modelo Ratkowsky foi considerado como o melhor modelo não linear, foram adicionadas variáveis a nível de parcela, a fim de melhorar o desempenho do modelo, tornando-o um modelo genérico. Para isso, as variáveis índice de sítio, idade, área basal, número de fustes, Dq, diâmetro e altura dominantes foram adicionadas e combinadas de diferentes formas ao modelo, resultando em diferentes modelos genéricos de altura. A partir de todas as combinações geradas, selecionou-se as 5 melhores equações genéricas (Tabela 6), utilizando as mesmas métricas utilizadas para a seleção do modelo não linear local.

Tabela 6 – Modelos genéricos não lineares.

Model	Function form
Generalized model 1	$H = (\beta_0 + H_{dom})(1 - \exp(-(D + \beta_1)\beta_2)) + 1.3 + \epsilon$
Generalized model 2	$H = (\beta_0 H_{dom}) \left(1 - \exp(-(D + (\beta_1 H_{dom}))\beta_2)\right) + 1.3 + \epsilon$
Generalized model 3	$H = (H_{dom}^{\beta_0})(1 - \exp(-(D + \beta_1)\beta_2)) + 1.3 + \epsilon$
Generalized model 4	$H = (\beta_0 H_{dom}) \left(1 - \exp(-(D + (\beta_1 D_{dom}))\beta_2)\right) + 1.3 + \epsilon$
Generalized model 5	$H = (\beta_0 H_{dom}) \left(1 - \exp(-(D + (\beta_1 Dq))\beta_2)\right) + 1.3 + \epsilon$

Após a adição das variáveis a nível de UA, a fim de se obter um modelo que representasse de forma mais adequada a variação entre as UA, Hdom, Ddom e Dq foram as variáveis que mais influenciaram o desempenho da equação, chegando a atingir uma redução de, no mínimo, 40,2% e 43,9% no valor de RMSE para base de treino e teste, respectivamente.

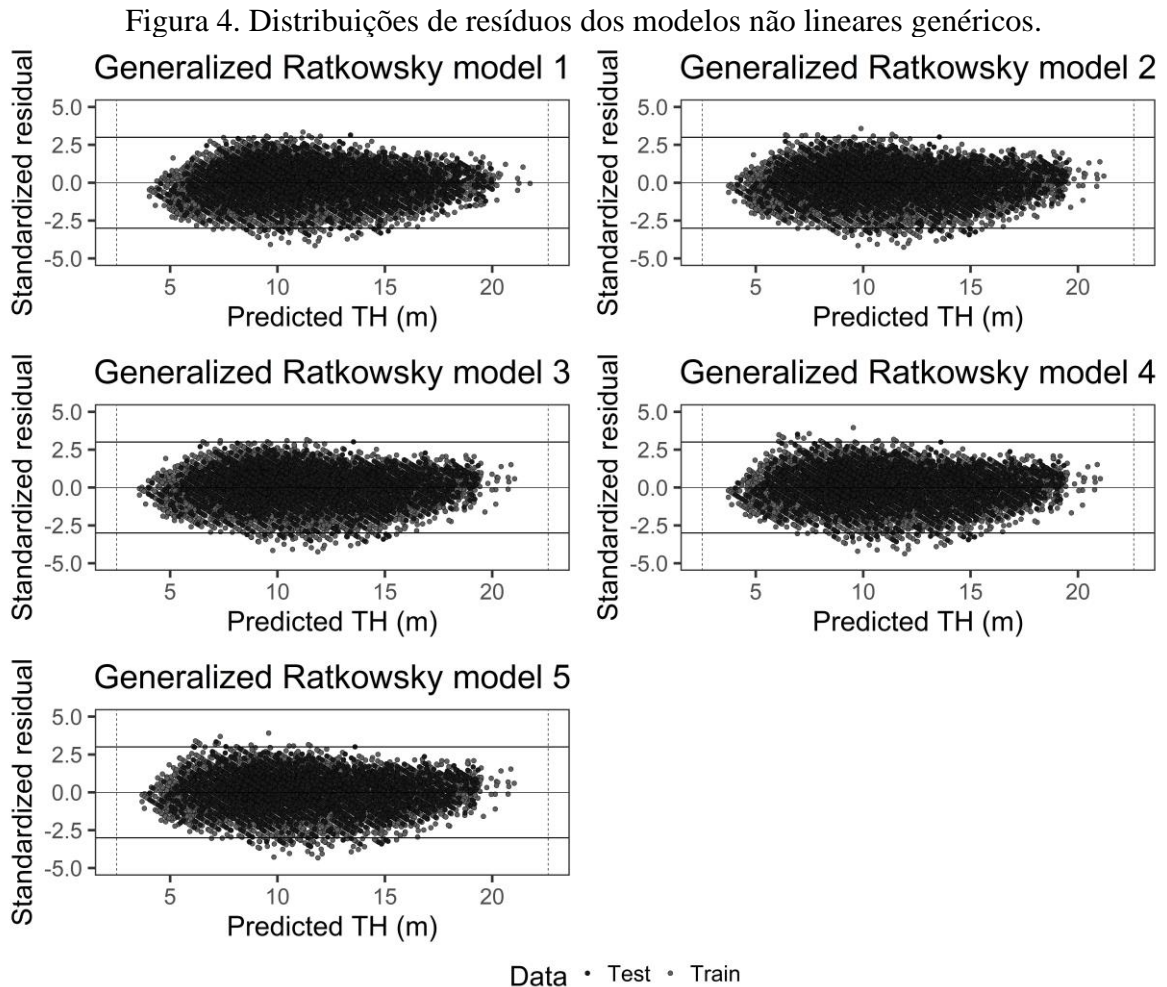
Como pode ser observado na Tabela 7, o melhor modelo genérico (Generalized model 4) foi resultado da inclusão de Hdom e Ddom ao modelo Ratkowsky local. Dentre as 5 melhores equações genéricas selecionadas, a equação 4 foi a que resultou nos menores valores de RMSE, MAPE e AIC (1,14; 8,20 e 44852, respectivamente) e no maior valor de EF (0,8689) para a base de treino. Já para a base de teste, a equação obteve desempenho superior para as mesmas métricas (RMSE = 1,10, MAPE = 7,73, EF = 0,8806 e AIC = 11451). Deste modo, a equação ajustada proveniente do Generalized model 4 foi selecionada como a melhor equação genérica.

Tabela 7 – Desempenho estatístico dos modelos genéricos testados.

Model	TRAIN (plot= 479; n= 14505)					TEST (plot=123; n= 3784)					Overall rank
	RMSE	MAPE	EF	BIAS	AIC	RMSE	MAPE	EF	BIAS	AIC	
Generalized model 1	1,15	8,34	0,8650	-0,0002	45271	1,11	7,82	0,8783	0,0672	11524	3,14
Generalized model 2	1,15	8,35	0,8643	0,0173	45344	1,11	7,84	0,8777	0,0766	11543	4,23
Generalized model 3	1,16	8,38	0,8636	0,0277	45427	1,11	7,86	0,8773	0,0807	11553	5,00
Generalized model 4*	1,14	8,20	0,8689	0,0158	44852	1,10	7,73	0,8806	0,0721	11451	1,00
Generalized model 5	1,14	8,21	0,8687	0,0159	44871	1,10	7,74	0,8804	0,0719	11457	1,17

O modelo seguido por * indica o de melhor desempenho.

Avaliando as distribuições de resíduos dos modelos genéricos (Figura 4), é possível observar uma distribuição uniforme, com os valores de desvio padronizado mais próximos de zero. Além disso, também é possível perceber que os dados preditos estão distribuídos de uma forma mais ampla dentro do intervalo de valores mínimo e máximo observados, ou seja, as predições estão representando de forma mais coerente os valores reais.



Fonte: Do autor (2024).

3.3 Modelos não lineares de efeitos mistos

De posse dos modelos local e genérico selecionados, ainda foi incorporado a estes efeitos aleatórios a nível de UA, na tentativa de aumentar a acurácia dos modelos. A inclusão do efeito aleatório aos modelos foi condicionada à capacidade de convergência do modelo e os resultados do teste MLRT das estruturas que convergiram podem ser observados na Tabela 8.

Tabela 8 – Resultados do teste MLRT para as diferentes combinações do modelo Ratkowsky na estrutura mista.

Type	Model	Random Parameters	LogLik	BIC	AIC	Test	MLRT	p-value
Basic model	1	b1, b2	-21889,94	43846,96	43793,88	1 vs 6	15996,74	0,0000
	2	b0, b2	-21879,11	43825,29	43772,21	2 vs 6	16018,41	0,0000
	3	b0	-22221,07	44490,05	44452,14	3 vs 6	15334,48	0,0000
	4	b1	-21965,23	43978,36	43940,45	4 vs 6	15846,16	0,0000
	5	b2	-22020,94	44089,80	44051,89	5 vs 6	15734,73	0,0000
	6	None	-29888,31	59814,95	59784,62	-	-	-
Generalized model	1	b0, b1	-21292,12	42651,32	42598,24	1 vs 4	2261,56	0,0000
	2	b0	-21709,36	43466,63	43428,72	2 vs 4	1427,08	0,0000
	3	b1	-21307,19	42662,29	42624,38	3 vs 4	2231,42	0,0000
	4	None	-22422,90	44884,13	44853,80	-	-	-

Onde valores $p \leq 0,05$ indicam a diferença significativa do modelo em relação ao modelo base.

O modelo básico local resultou em 5 estruturas mistas diferentes, onde os valores de AIC variaram de 43771,62 a 44452,12, BIC de 43824,70 a 44490,04 e LogLik de -22221,06 a -21878,81. As estruturas mistas apresentaram diferenças significativas entre si, com base no critério usado por Burnham e Anderson (2002) e, por meio do teste MLRT, foi verificado que a adição dos parâmetros de efeito aleatório resultou em uma melhora significativa dos modelos em todas as estruturas. Além disso, para as métricas estatísticas avaliadas (AIC, BIC e LogLik), a estrutura básica em que interação do efeito aleatório ocorreu com os parâmetros β_0 e β_2 foi a que apresentou o melhor desempenho, resultando na seguinte expressão: $H = (\beta_0 + b_0) \left(1 - \exp\left(- (D + \beta_1)(\beta_2 + b_2)\right) \right) + 1.3 + \epsilon$. Sendo assim, esta estrutura foi considerada significativamente superior às demais.

Já para o modelo genérico, três estruturas mistas testadas convergiram, onde os valores de AIC variaram de 42598,25 a 43428,72, BIC de 42651,32 a 43466,63 e LogLik de -21709,36 a -21292,12. O valor de AIC para as 3 estruturas apresentaram diferenças significativas entre si, de acordo o critério utilizado por Burnham e Anderson (2002), sendo a interação do efeito aleatório com os parâmetros β_0 e β_1 a que resultou em um desempenho superior, de acordo o teste MLRT, resultando na seguinte expressão: $H = ((\beta_0 + b_0)H_{dom}) \left(1 - \exp\left(- \left(D + ((\beta_1 + b_1)D_{dom}) \right) \beta_2 \right) \right) + 1.3 + \epsilon$. Desta forma, os parâmetros do modelo genérico de efeitos fixos, da estrutura mista do modelo básico e da estrutura mista do modelo genérico podem ser observados na Tabela 9.

Tabela 9 – Estatísticas paramétricas para as equações de estruturas fixa e mista ajustadas.

Parameter estimates	Values	Standard error	t value	p value
Generalized Model				
β_0	0,9414	0,0031	302,25	<0,001
β_1	-0,1162	0,0039	-29,84	<0,001
β_2	0,1248	0,0018	69,15	<0,001
Basic Mixed-effects				
β_0	18,3530	0,2678	68,53	<0,001
β_1	-11,2781	0,4089	-27,58	<0,001
β_2	2,4629	0,3039	8,10	<0,001
σ^2_{b0}	3,1190	-	-	-
σ^2_{b2}	1,9221	-	-	-
$\sigma^2_{b0 b2}$	0,1050	-	-	-
Generalized Mixed-effects				
β_0	1,0033	0,0055	181,70	<0,001
β_1	-0,0688	0,0068	-10,05	<0,001
β_2	0,0986	0,0020	50,09	<0,001
σ^2_{b0}	0,0273	-	-	-
σ^2_{b1}	0,0808	-	-	-
$\sigma^2_{b0 b1}$	-0,3740	-	-	-

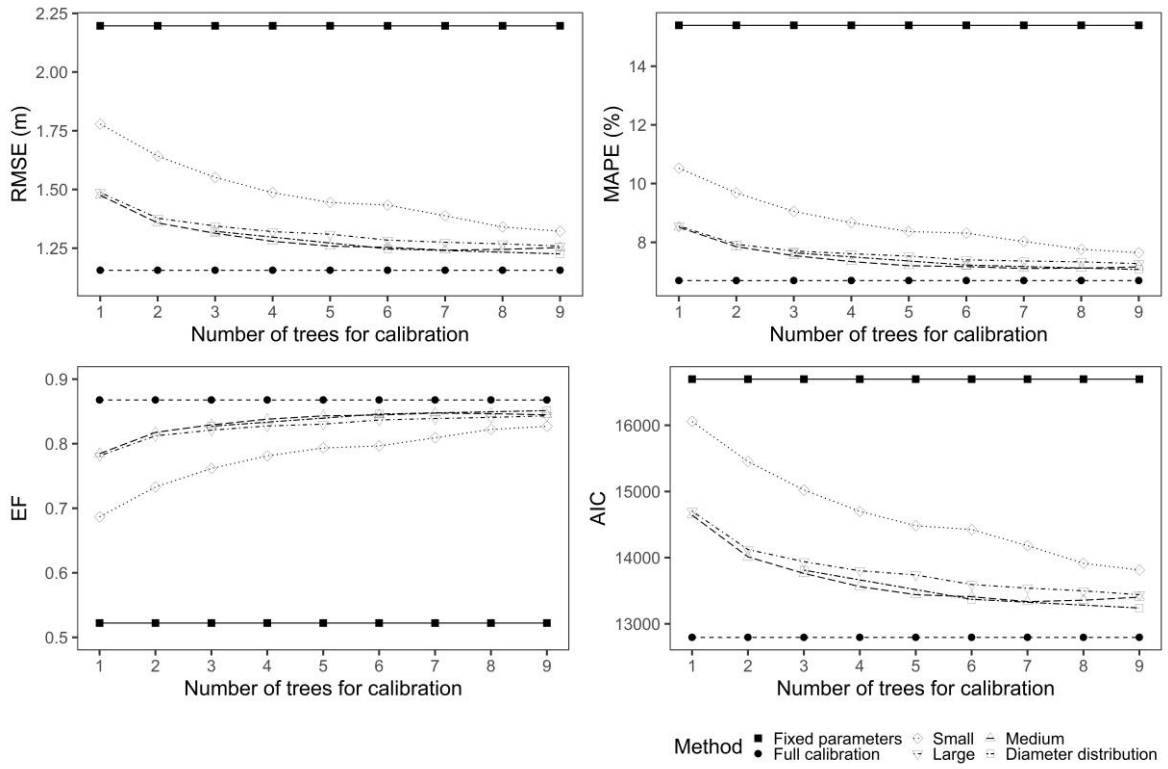
Onde foram considerados significantes os parâmetros com valor $p \leq 0,05$.

A partir da Tabela 9 é possível perceber que todos os parâmetros das equações ajustadas foram significativos.

3.4 Calibração dos modelos de efeito misto

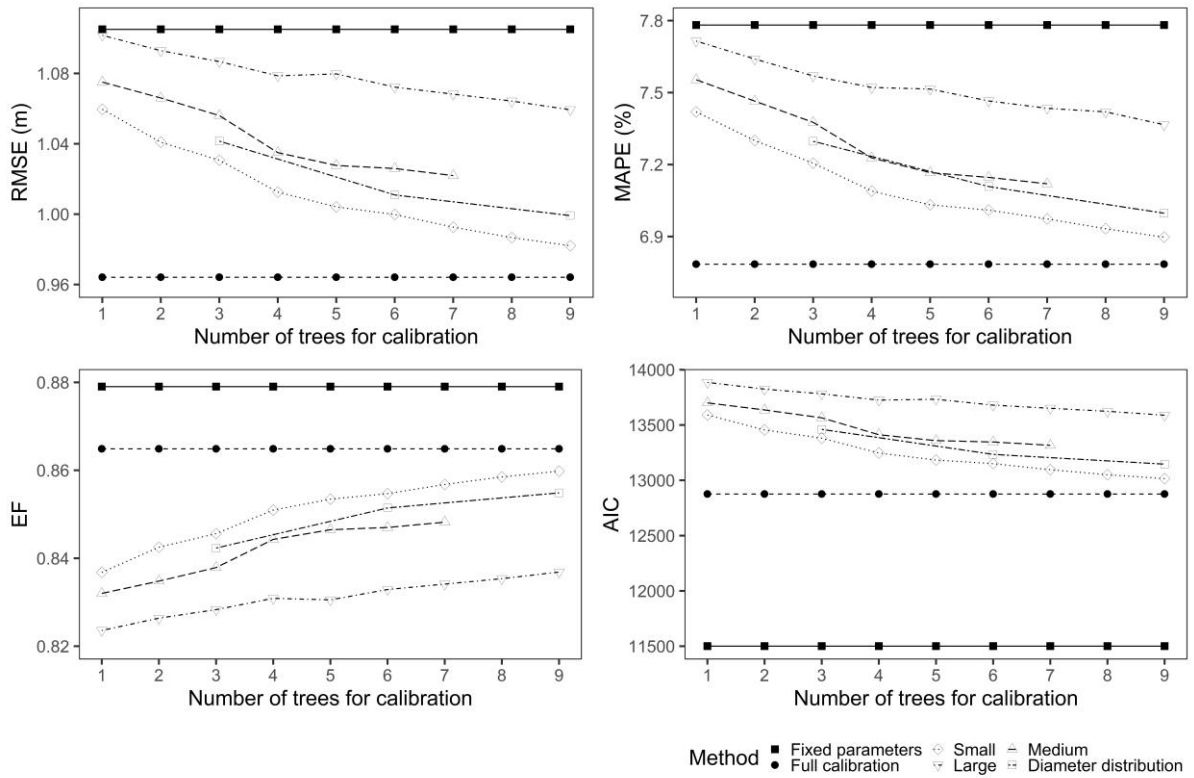
A capacidade de predição dos modelos mistos propostos foi avaliada aplicando-os à base de teste, que não foi utilizada na fase de ajuste dos modelos. No entanto, diferentes alternativas de calibração, em que se utilizou subamostras retiradas da base de teste, foram testadas e o comportamento das métricas estatísticas para estas alternativas são mostrados nas Figuras 5 e 6 para os modelos local e genérico mistos, respectivamente.

Figura 5 – Desempenhos das calibrações do modelo local misto.



Fonte: Do autor (2024).

Figura 6 – Desempenhos das calibrações do modelo genérico misto.



Fonte: Do autor (2024).

Uma redução nos valores de RMSE, MAPE e AIC pode ser notada conforme o tamanho da subamostra utilizada para a calibração aumenta para ambos os modelos. Para o modelo misto básico, os valores de RMSE, MAPE e AIC reduziram 44,21, 54,04 e 20,72%, respectivamente, sendo todas estas reduções máximas encontradas para a calibração que considerou, como subamostra, uma mistura de árvores de classe de tamanho menor, maior e médio, sendo selecionadas três árvores por classe diamétrica de cada UA (nove árvores). Além disso, um aumento de 63% no valor de EF também foi encontrado nesta alternativa de calibração. Isto indica a calibração com nove árvores selecionadas de acordo a distribuição diamétrica da UA como a melhor alternativa de calibração para o modelo misto básico.

Utilizando a calibração completa como referência, é possível perceber que, para todas as métricas avaliadas, os valores resultantes da melhor opção de calibração foram bem próximos aos da calibração completa, com diferenças menores que 7,0 %, indicando que a subamostra foi suficiente para obtenção de predições satisfatórias. Por outro lado, a calibração utilizando as menores árvores por UA, foi a alternativa que resultou no pior desempenho dentre todas as alternativas de calibração avaliadas, sendo superior apenas à ausência de calibração.

Considerando o modelo misto genérico, a calibração considerando as menores árvores por UA resultou em desempenho superior às outras opções de calibração. A calibração com as menores árvores gerou reduções dos valores de RMSE e MAPE na ordem de 11% para ambos. Já para o valor de AIC, houve um aumento de 11501 para 13016. Ainda, uma redução de 0,88 para 0,86 no valor de EF (2.19%) também foi encontrada para esta alternativa de calibração. Dessa forma, a calibração com as menores árvores foi considerada a melhor alternativa.

Diferentemente do modelo misto básico, o modelo misto genérico apresentou uma diferença menor entre o desempenho resultante da melhor opção de calibração e o modelo utilizando apenas os efeitos fixos, o que é esperado, uma vez que a este modelo já estão incorporadas variáveis a nível de UA. Comparando os valores da melhor opção de calibração com a calibração completa, percebe-se que, para todas as métricas, os valores resultantes da melhor opção de calibração foram muito próximos aos da calibração completa, com diferenças inferiores a 2,0 %, indicando que a subamostra selecionada foi suficiente para obtenção de predições satisfatórias. Contudo, a calibração utilizando as maiores árvores por UA, foi a alternativa que resultou em um desempenho muito próximos à ausência de calibração para as métricas RMSE e MAPE e desempenho inferior para as métricas EF e AIC, para subamostras de 1-9 árvores por UA, sendo assim a pior alternativas de calibração para o modelo misto genérico.

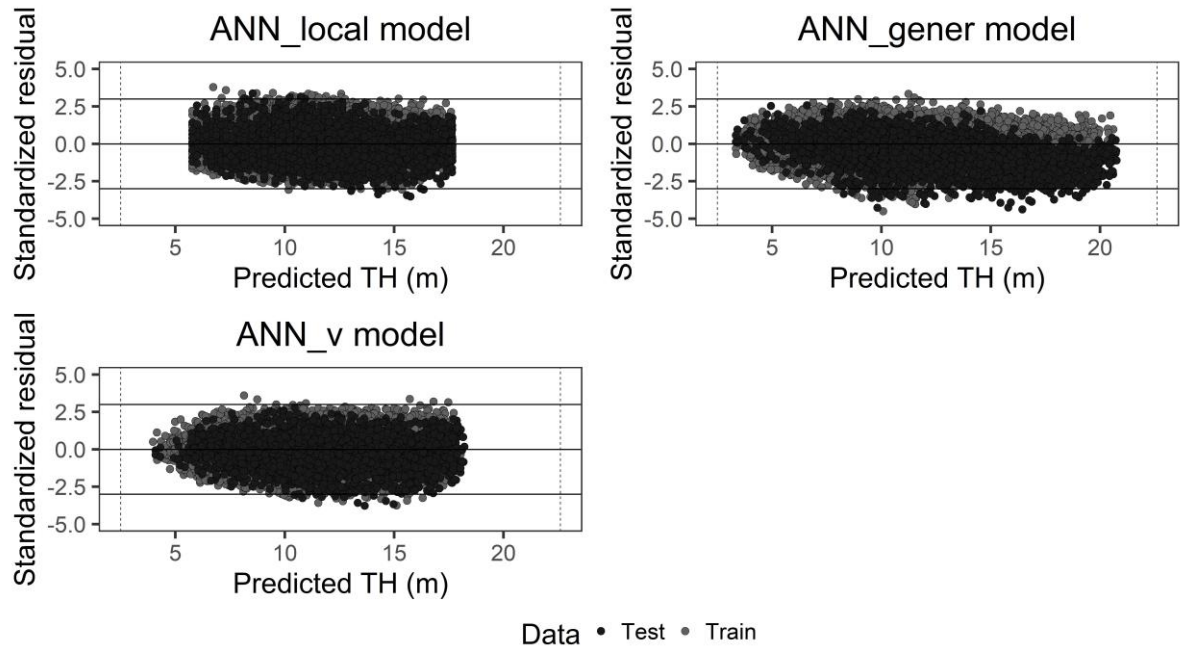
	non linear mixed (Fixed)	MIXED	2,20	15,39	0,5223	0,3137	16698	0,0115	9,00
	gener non linear mixed (Fixed)	MIXED	1,11	7,78	0,8790	0,0518	11501	0,8184	1,00
	ANN_local	ANN	2,01	14,37	0,5991	-0,4380	16041	0,0000	8,19
	ANN_generalized	ANN	1,34	9,23	0,8225	-0,6770	12974	0,0000	4,34
	ANN_v	ANN	1,67	11,74	0,7239	-0,4662	14637	0,0000	5,96
TEST (plot=125; n=3806) after calibration	non linear mixed (Dist. 9)	MIXED	1,23	7,07	0,8513	-0,0166	13239	0,3125	1,71
	gener non linear mixed (Mín.9)	MIXED	0,98	6,90	0,8598	0,0130	13016	0,4140	1,15

A ANN_generalized, que teve como entradas DAP, Hdom e Dq, foi a melhor alternativa dentre as redes e resultou em uma maior diferença, quando comparada à equação genérica ajustada pelo método ONLS, com DAP, Hdom e Ddom como variáveis independentes. A RNA_generalized resultou em uma melhora de 6,58%, 6,51%, 1,92%, 99,11% e 4,35% para as métricas RMSE, MAPE, EF, BIAS e AIC, respectivamente, para a base de treino e piora de 21,92%, 19,34%, 6,60%, 1038,90% e 13,30% para RMSE, MAPE, EF, BIAS e AIC, respectivamente, para a base de teste.

Já a ANN_v, que teve como entradas DAP e DAP_V, foi inferior quando comparada à equação mista local. A RNA_V resultou em uma piora de cerca de 68%, 68%, 19% e 37% para as métricas RMSE, MAPE, EF e AIC, respectivamente, para a base de treino. Comparando com a mesma equação sem calibração, aplicada à base de teste, a RNA atingiu melhora de 24%, 24%, 39% e 12% para RMSE, MAPE, EF e AIC, respectivamente.

Analisando as distribuições de resíduos das RNA (Figura 7), percebe-se que a disposição dos resíduos da ANN_local foi semelhante aos do modelo básico ajustado pelo método OLS, concordando com as métricas estatísticas. A ANN_generalized resultou em uma distribuição uniforme para a base de treino mas, contrariamente, para a base de teste resultou em uma tendência de superestimação que se manifesta de forma mais pronunciada conforme os valores de altura aumentam. Tais resultados corroboram com as métricas estatísticas, em que, de todos os modelos, a RNA_generalized apresentou o pior resultado de BIAS, indicando um viés de superestimação. Já a ANN_V resultou em resíduos simétricos, com disposição mais restrita em relação à amplitude dos valores reais, no entanto não houve tendência.

Figura 7 – Distribuição de resíduos para as RNA ajustadas.



Fonte: Do autor (2024).

Os coeficientes obtidos a partir dos treinamentos dos três tipos de RNA podem ser observados na tabela 12.

Tabela 12 – Coeficientes ajustados para as RNA.

Node	β_0	β_1	β_2	β_3	β_4	β_5
ANN_local						
Neuron 1	1,6423	-0,8850	-	-	-	-
Neuron 2	-0,2093	6,1402	-	-	-	-
Neuron 3	1,4926	-2,9540	-	-	-	-
Output	-0,0335	0,5634	0,1247	-0,3328	-	-
ANN_generalized						
Neuron 1	-3,3771	17,3389	-13,8469	4,0253	-	-
Neuron 2	6,1106	-1,5256	-4,8904	-0,5606	-	-
Neuron 3	-1,1331	-6,0705	-0,2003	3,7169	-	-
Neuron 4	-0,2259	1,4297	2,9516	-0,7158	-	-
Neuron 5	3,7880	-15,3048	-12,7891	-0,7530	-	-
Output	-0,7988	-0,1089	-2,7279	-1,2573	4,2603	-0,8286
ANN_v						
Neuron 1	5,3854	35,9018	-22,1109	-	-	-
Neuron 2	-0,3797	-0,2629	-2,3960	-	-	-
Neuron 3	1,0473	15,9499	-14,1276	-	-	-
Neuron 4	0,1320	-0,4278	0,0380	-	-	-
Output	-0,1174	0,2031	-1,5320	0,3693	0,9556	-

Para melhor compreensão sobre a utilização do sistema de equações extraído da RNA, realizou-se a exemplificação da utilização do sistema (30-37) para obter a predição de altura a partir dos coeficientes obtidos com o treinamento da RNA_local. A arquitetura da RNA_local pode ser observada na figura 8.

$$H = H' * (\max(H) - \min(H)) + \min(H) = H' * (22,6 - 2,5) + 2,5 \quad (30)$$

$$\begin{aligned} H' &= \beta_0 + \beta_1 * z_1 + \beta_2 * z_2 + \beta_3 * z_3 + \epsilon_i \\ &= -0,0335 + 0,5634 * z_1 + 0,1247 * z_2 - 0,3328 * z_3 + \epsilon_i \end{aligned} \quad (31)$$

$$z_1 = \frac{1 - e^{(-2w_1)}}{1 + e^{(-2w_1)}} \quad (32)$$

$$z_2 = \frac{1 - e^{(-2w_2)}}{1 + e^{(-2w_2)}} \quad (33)$$

$$z_3 = \frac{1 - e^{(-2w_3)}}{1 + e^{(-2w_3)}} \quad (34)$$

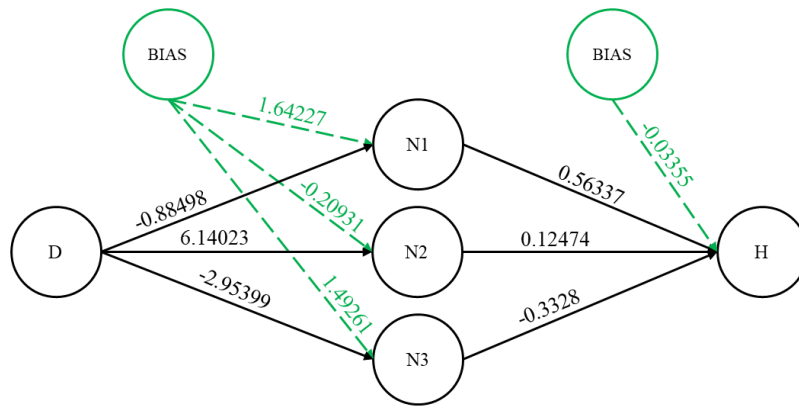
$$w_1 = \beta_{0.1} + \beta_{1.1} * DAP_i' = 1,6423 - 0,8850 * DAP_i' \quad (35)$$

$$w_2 = \beta_{0.2} + \beta_{1.2} * DAP_i' = -0,2093 + 6,1402 * DAP_i' \quad (36)$$

$$w_3 = \beta_{0.3} + \beta_{1.3} * DAP_i' = 1,4926 + -2,9540 * DAP_i' \quad (37)$$

Onde H' é a altura total normalizada, β_0 é o bias, β_{1-3} são os coeficientes associados com os neurônios 1-3, respectivamente, $\beta_{0.1-3}$ são os coeficientes entre a variável de entrada DAP e o neurônio n, z_n é a resposta do n-ésimo neurônio da camada oculta após transformação pela função de ativação, w_{1-3} são as somas dos produtos entre as variáveis de entrada (DAP) e seus respectivos pesos do n-ésimo neurônio.

Figura 8 – Estrutura e parâmetros da RNA_local em forma gráfica.



Onde D é o diâmetro a 1,3 m do solo e H a altura total das árvores.
Fonte: Do autor (2024).

3.6 Avaliação geral

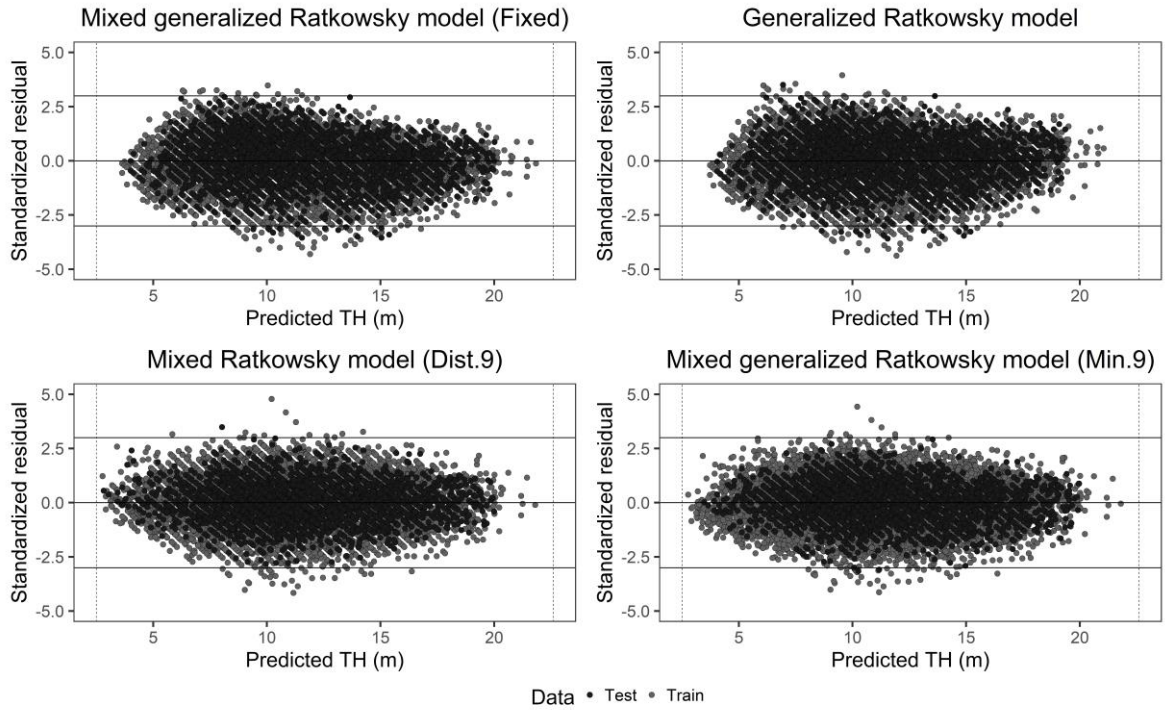
O resumo estatístico de todas as metodologias avaliadas para bases de treino e teste pode ser observado na Tabela 11. É possível notar que, na fase de ajuste, a modelagem mista dos modelos local e genérico foi a que apresentou os melhores resultados juntamente com a ANN_generalized. Já para a base de teste, as melhores abordagens foram o modelo genérico misto sem calibração e os modelos genérico e local, com calibrações incluindo as nove menores árvores e a outra incluindo nove árvores, de acordo a distribuição diamétrica, respectivamente. Estas foram as únicas abordagens que não resultaram em previsões significativamente diferentes dos valores reais, de acordo com o teste t. Embora o modelo genérico de apenas efeitos fixos tenha resultado em um bom valor no *ranking* e uma distribuição semelhante ao modelo genérico de efeito misto, ele apresentou previsões significativamente diferentes dos valores observados.

Resumindo, a inclusão das covariáveis a nível de UA resultaram em uma grande melhora na precisão dos modelos, sendo que os melhores desempenhos foram todos alcançados com a inclusão dessas variáveis e/ou o efeito aleatório a nível de UA, porém, das técnicas utilizadas, apenas os modelos de efeitos mistos foram capazes de resultar em previsões estatisticamente semelhante aos valores reais para as bases de treino e teste.

Mixed Generalized Ratkowsky (Fixed), Mixed Generalized Ratkowsky (Min. 9) e Mixed local Ratkowsky (Dist. 9) foram os que resultaram nos melhores desempenhos, com resultados adequados para a estimação da altura total. Todas estas alternativas apresentaram distribuições de resíduos homogêneas entorno de zero e com distribuições semelhantes à normal (Figuras 8 e 9) tanto para a base de treino quanto de teste. Além disso, os valores de altura total

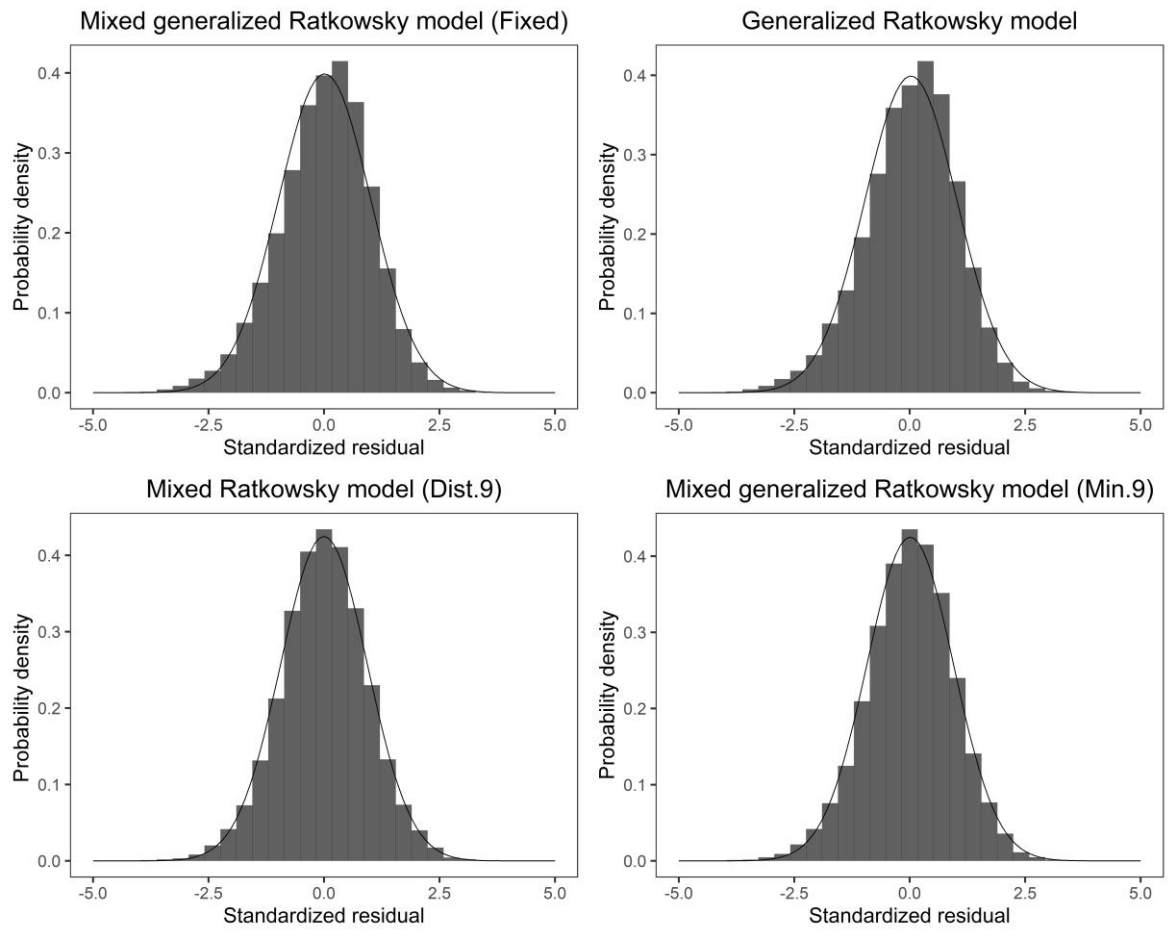
observados foram comparados, graficamente (Figura 10), com os valores preditos por estas alternativas e confirmaram a qualidade dos ajustes.

Figura 8 – Dispersão e distribuição de resíduos das melhores metodologias avaliadas.



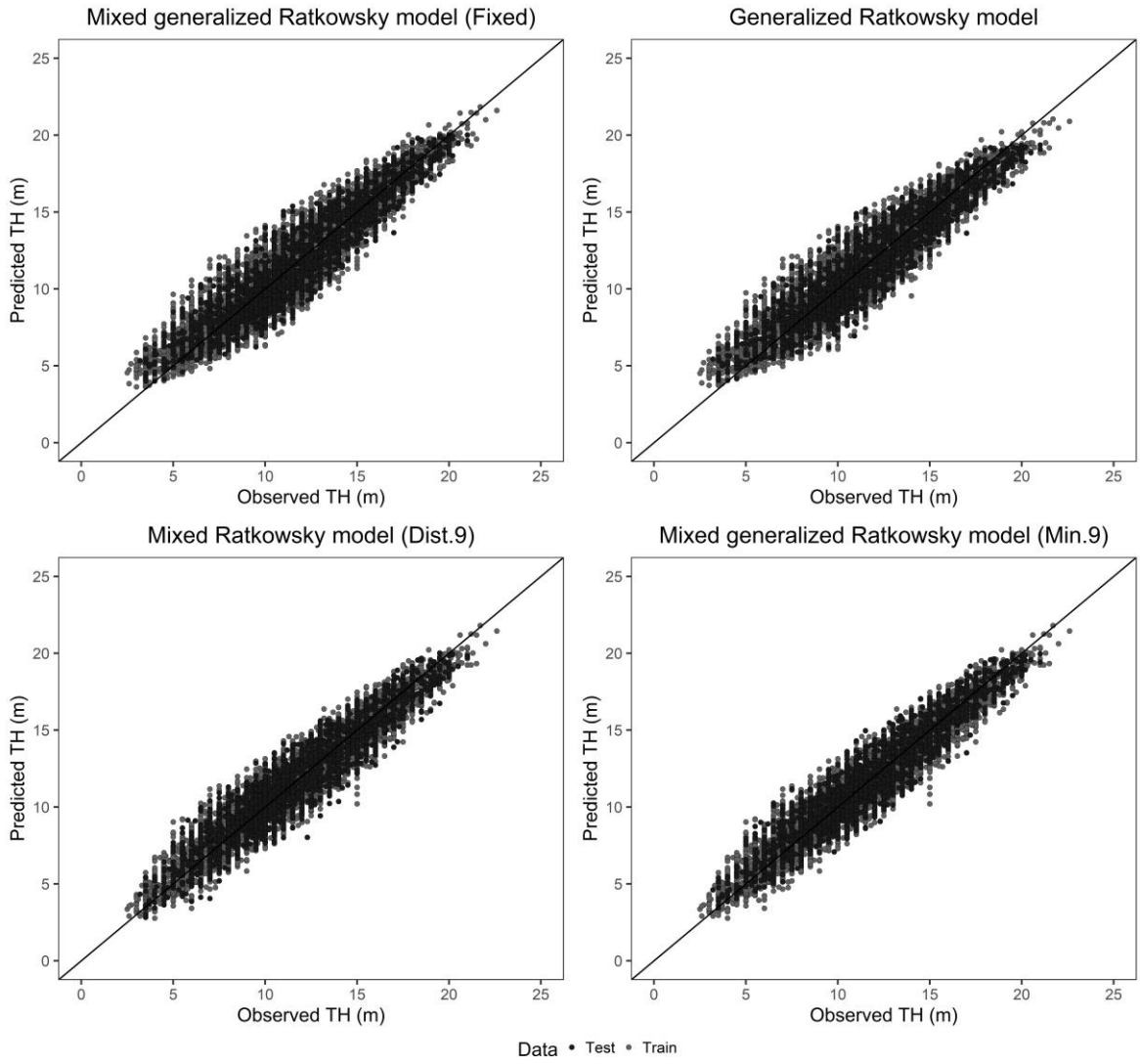
Fonte: Do autor (2024).

Figura 9 – Distribuição de frequência dos resíduos das melhores metodologias avaliadas.



Fonte: Do autor (2024).

Figura 10 – Gráficos de valores observados *versus* preditos das melhores metodologias avaliadas.



Fonte: Do autor (2024).

4 DISCUSSÃO

A partir da estatística paramétrica dos modelos, é possível perceber que todas as equações ajustadas possuem todos os seus parâmetros significativos, indicando que essas equações são capazes de descrever a variabilidade da altura total a partir do DAP (Sanquetta *et al.*, 2014).

Os resultados do presente estudo indicaram que, dentre os modelos não lineares ajustados pelo método ONLS utilizando apenas o DAP como variável preditora, modelo de Ratkowsky (1990) foi o mais adequado para a predição da altura total das árvores de *Pinus caribaea* var. *hondurensis* na faixa de idades avaliada. Quando apenas o DAP é utilizado como variável independente para predição da altura total de árvores, o erro percentual, associado com as predições, normalmente varia entre 10% e 20% (Long *et al.* 2020), como foi observado no presente estudo.

Os resultados aqui encontrados também apontam uma maior eficiência de todas as equações quando comparadas a estudos anteriores para o gênero *Pinus* (Cerqueira *et al.*, 2017; Rodrigues *et al.*, 2017). Esses resultados são consistentes, uma vez que, de modo geral, a relação entre altura e diâmetro tende a uma forma assintótica, com a altura restringida a um limite superior enquanto o diâmetro não (Niklas, 1995) e, contrastando com os trabalhos citados, todas as cinco equações não lineares ajustadas no presente estudo são capazes de descrever este comportamento. Além disso, modelos não lineares frequentemente resultam em melhores ajustes hipsométricos quando comparados aos modelos lineares com o mesmo número de parâmetros (Huang *et al.*, 1992).

A melhora de desempenho a partir da inclusão das variáveis a nível de UA na equação selecionada foi nítida. Esse comportamento é esperado uma vez que a adição dessas variáveis é capaz de contribuir para a explicação da variabilidade dentro da UA (Ogana *et al.*, 2020). Nos estudos de Crecente-Campo *et al.* (2010) e Uzoh (2017), os autores recomendaram a adição de variáveis a nível de UA para aumentar a eficiência da modelagem da relação hipsométrica. Sánchez *et al.* (2003), Canga-Libano *et al.* (2009) e Santiago-García *et al.* (2020), estudando espécies do gênero *Pinus*, recomendaram a inclusão do diâmetro dominante e altura dominante, assim, como no presente estudo. Os autores também ressaltaram que estas são variáveis que não demandam esforços extras para coleta, uma vez que são variáveis já comumente mensuradas nas rotinas dos inventários florestais.

A inclusão dos parâmetros de efeito aleatório resultou em melhoras significativas dos modelos local e genérico em todas as estruturas avaliadas, indicando a importância da

modelagem mista para o aumento da qualidade dos modelos. Esses resultados estão de acordo com a literatura, onde já foi constatado que, normalmente, os modelos de efeito misto resultam em melhor desempenho quando comparados aos ajustes pelo método OLS (Temesgen *et al.*, 2008). De acordo com Uzoh (2017), a combinação de ambos efeitos fixos, que descrevem o comportamento de toda a população, e aleatórios, que fornecem informações a nível de UA, é responsável pelo aumento na qualidade desses modelos. Além disso, a inclusão dos efeitos aleatórios específicos para cada UA permite que as equações superem a falta de independência entre as observações, provenientes da estrutura hierárquica dos dados (árvores nas UA) (Crecente-Campo *et al.*, 2010).

Além da modelagem mista, no presente estudo, foi avaliada a melhor alternativa de calibração, considerando a quantidade e tamanho das árvores presentes na subamostra de calibração, sendo as árvores de tamanho variado e as menores árvores as melhores alternativas, para os modelos local e genérico, respectivamente, sendo exigidas nove árvores por parcela em cada.

Embora não exista um consenso sobre o número de alturas ou a estratégia de calibração a ser utilizada para calibrar as equações hipsométricas, diferentes pesquisas já avaliaram este tema (Calama; Montero, 2004; Temesgen *et al.*, 2008; Ciceu *et al.*, 2020). No presente trabalho, para todas as estratégias testadas, o aumento do número de árvores amostradas resultou em um aumento na precisão das predições, concordando com os resultados encontrados por Ciceu *et al.* (2020).

Além do número de árvores utilizadas para a calibração das equações, quais árvores devem ser selecionadas é um fator que influencia diretamente a qualidade das predições. Dessa forma, no presente estudo, a amostragem aleatória foi desconsiderada, uma vez que, estudos anteriores têm mostrado que o tamanho das árvores amostradas influencia a qualidade das calibrações (Calama; Montero, 2004). Corroborando com esses resultados, as alternativas de calibração testadas neste estudo, que diferiram no tamanho das árvores amostradas, resultaram em precisões diferentes e, para o modelo local misto, as subamostras contendo árvores de mais de uma classe de tamanho resultaram em desempenho superior quando comparadas com as de apenas uma classe. Um comportamento semelhante também foi reportado nos estudos de Castedo-Dorado *et al.* (2006) e Bronisz e Mehtätalo (2020), destacando a importância do tamanho das árvores para o processo de calibração.

Os resultados superiores de EF e AIC resultantes da utilização de apenas os parâmetros fixos do modelo genérico misto pode indicar uma influência do grande número de parâmetros utilizados quando os efeitos aleatórios são considerados, uma vez que essas métricas

penalizaram a quantidade de parâmetros. Desse modo, a grande capacidade de predição utilizando apenas os efeitos fixos da equação genérica mista, representada pelas predições estatisticamente iguais aos valores observados, pode indicar que apenas a inclusão das variáveis a nível de parcela (diâmetro dominante e altura dominante) pode ser suficiente para obtenção de resultados personalizados por UA com alta precisão. A inclusão do diâmetro dominante pode representar o nível de competição na UA (Castedo-Dorado *et al.*, 2006), enquanto que a altura dominante reflete a produtividade e é pouco afetada pela densidade do plantio dentro de um amplo intervalo de densidades (Skovsgaard; Vanclay, 2008). Além disso, para obtenção da altura dominante, poucas medidas adicionais são necessárias.

A RNA que considerou apenas o DAP como entrada foi inferior à equação ajustada pelo método ONLS com apenas o DAP como variável independente em relação às predições para a base de teste. Esses resultados concordam com os de Castaño-Santamaría *et al.* (2013), onde os autores reportaram um desempenho inferior da RNA quando comparada ao método ONLS. Porém, considerando a base de treino, a RNA_local resultou em desempenho levemente superior, indicando que a RNA pode ser comparável com a equação de regressão ONLS, como sugerido por Özçelik *et al.* (2013). Além disso, discordando do presente estudo, o autor reportou que a RNA contendo a medida de variância do DAP foi superior à equação genérica ajustada pelo método ONLS contendo o diâmetro e altura dominantes. Ainda, contrastando com os resultados de Karatepe *et al.* (2022), a RNA contendo DAP, diâmetro e altura dominantes, quando aplicada a base de teste, resultou em predições não estatisticamente iguais aos valores observados e houve a presença de tendência, indicando uma baixa capacidade generalização.

Esperava-se que as RNA genéricas seriam capazes de gerar resultados satisfatórios e pudessem ser facilmente aplicadas a novas parcelas, uma vez que necessitariam apenas das informações de diâmetros e altura dominante. No entanto, os resultados apontaram dificuldades de ajuste e predições contendo viés. Embora uma função não precise ser especificada, a função de ativação, número de camadas ocultas e neurônios por camada devem ser cuidadosamente escolhidos, pois diferentes estruturas podem resultar em grande diferença na qualidade das predições (Aertsen *et al.*, 2010; Leduc *et al.*, 2001). Além disso, mesmo havendo diferentes estratégias para guiar a seleção da melhor estrutura (Silva; Spatti; Flauzino, 2010; Júnior *et al.* 2020; Sahin *et al.*, 2023), a seleção necessita da avaliação de diversas opções por tentativa e erro, levando à necessidade de um grande esforço de modelagem para otimização da acurácia e capacidade de generalização da RNA.

Na literatura, frequentemente as RNA são descritas como “*blackbox technique*”, devido à presença de camadas ocultas e a não linearidade das funções de ativação (Aertsen *et al.*, 2010),

porém a estrutura das RNA pode sim ser interpretada, uma vez que é composta pelas combinações de funções e parâmetros ajustados (pesos) que, resultam em uma equação definida a partir da estrutura da RNA e, se analisada de forma mais detalhada, pode ter seu comportamento interpretado, podendo fornecer informações valiosas, uma vez que as RNA são capazes de extrair características não explícitas a partir dos dados fornecidos para o seu treinamento (Özçelik *et al.*, 2013). A extração dos coeficientes das RNA não é uma prática comum entre os trabalhos de mensuração florestal (Thanh; Tien; Shen, 2019; Socha; Netzel; Cywicka, 2020), limitando a aplicabilidade das equações ajustadas. No entanto, no presente estudo, assim como em trabalhos recentes (Dantas *et al.*, 2020; Costa *et al.*, 2022; Dantas *et al.*, 2024) foi realizada a extração das equações provenientes do treinamento das RNA, onde os seus coeficientes são apresentados.

Predições imprecisas das alturas das árvores pode prejudicar seriamente as estimações do estoque de carbono. Dessa forma, equações hipsométricas acuradas são fundamentais para quantificação de volume e biomassa das florestas (Kearsley *et al.*, 2013; Cui, *et al.*, 2022). Além disso, equações hipsométricas são essenciais para a compreensão da estrutura de várias relações que afetam o desenvolvimento de ecossistemas florestais (Peng *et al.* 2001; Özçelik *et al.* 2013). Dessa forma, os resultados do presente estudo foram capazes de contribuir com o desenvolvimento de equações hipsométricas capazes de fornecer predições confiáveis e personalizadas a nível de parcela para plantios de *Pinus caribaea* var. *hondurensis*.

5 CONCLUSÃO

Dentre as equações não lineares avaliadas nesse estudo, a equação de Ratkowsky (1990) foi a mais adequada e possui comportamento capaz de explicar a interação entre altura e diâmetro das árvores de *Pinus caribaea* var. *hondurensis*.

A inclusão dos atributos altura dominante e diâmetro dominante associados aos parâmetros β_0 e β_1 , respectivamente, resultou em uma melhora nítida das predições do modelo Ratkowsky (1990), sendo recomendada a inclusão dessas variáveis na modelagem da relação hipsométrica de *Pinus caribaea* var. *hondurensis* sob condições semelhantes ao do presente estudo.

A inclusão de efeitos aleatórios, combinada com a calibração utilizando 9 árvores distribuídas em diferentes classes diamétricas por UA, foi capaz de aumentar significativamente o desempenho do modelo Ratkowsky (1990) local, resultando em predições significativamente iguais aos valores observados. A utilização dessa técnica implica em uma redução do número de alturas mensuradas por UA, quando comparada aos inventários tradicionais, onde se costuma medir cerca de 30% das alturas presentes na UA.

A adição dos efeitos aleatórios ao modelo Ratkowsky (1990) genérico foi capaz de gerar predições satisfatórias sem a necessidade de calibração quando aplicada à base de teste, indicando assim a importância da inclusão de variáveis a nível de UA. Além disso, uma vez que essas variáveis já são comumente obtidas nas rotinas de inventário, essa é uma estratégia viável e conveniente para o aumento da precisão das predições de altura total.

Os modelos de redes neurais artificiais não foram capazes de superar os modelos ONLS genéricos nem a modelagem mista e, também, resultaram em uma generalização de predições com viés, não sendo assim recomendadas para a base de dados em estudo. Dessa forma, há a necessidade de estudo mais aprofundados sobre a otimização de estratégias para a escolha da estrutura mais adequada das redes e interpretação de seus parâmetros ajustados.

REFERÊNCIAS

- ADAME, P.; DEL RÍO, M.; CAÑELLAS, I. A mixed nonlinear height–diameter model for pyrenean oak (*Quercus pyrenaica* Willd.). **Forest ecology and management**, v. 256, n. 1-2, p. 88-98, 2008.
- AERTSEN, W. *et al.* Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. **Ecological modelling**, v. 221, n. 8, p. 1119-1130, 2010.
- AKAIKE, H. Maximum likelihood identification of Gaussian autoregressive moving average models. **Biometrika**, v. 60, n. 2, p. 255-265, 1973.
- ASSMANN, E. **The principles of forest yield study: studies in the organic production, structure, increment and yield of forest stands**. New York, NY, USA: Pergamon Press, 1970. 506 p.
- BAILEY, D.; THOMPSON, D. How to develop neural-network applications. **AI expert**, v. 5, n. 6, p. 38-47, 1990.
- BATES, D. M.; WATTS, D. G. Relative curvature measures of nonlinearity. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 42, n. 1, p. 1-16, 1980.
- BISHOP, C. M. *et al.* **Neural networks for pattern recognition**. Oxford university press, 1995.
- BRONISZ, K.; MEHTÄTALO, L. Mixed-effects generalized height–diameter model for young silver birch stands on post-agricultural lands. **Forest Ecology and Management**, v. 460, p. 117901, 2020.
- BRONISZ, K.; MEHTÄTALO, L. Mixed-effects generalized height–diameter model for young silver birch stands on post-agricultural lands. **Forest Ecology and Management**, v. 460, p. 117901, 2020.
- BUONGIORNO, J.; GILLESS, J. K. **Decision Methods for Forest Resource Management**. Elsevier, 2003. 458p.
- BURNHAM, K. P.; ANDERSON, D. R. **Model selection and multimodel inference: a practical information-theoretic approach**. 2. ed. New York, US: Springer, 2002.
- CALAMA, R.; MONTERO, G. Interregional nonlinear height diameter model with random coefficients for stone pine in Spain. **Canadian Journal of Forest Research**, v. 34, n. 1, p. 150-163, 2004.
- CAMACHO, E. A. R. *et al.* Generalized height-diameter models with random effects for natural forests of central Mexico. **Cerne**, v. 28, 2022.
- CAMPOS, J. C. C.; LEITE, H. G. **Mensuração Florestal: perguntas e respostas**. 4. ed. Viçosa, MG: UFV, 2013, 605 p.

- CANGA, L. E. *et al.* Relación altura-diámetro generalizada para *Pinus radiata* D. Don en Asturias (Norte de España). **Sociedad Española de Ciencias Forestales**, v. 23, n. 1, p. 153-158, 2007.
- CASAS, G. G. *et al.* Configuration of the Deep Neural Network Hyperparameters for the Hypsometric Modeling of the *Guazuma crinita* Mart. in the Peruvian Amazon. **Forests**, v. 13, n. 5, p. 697, 2022.
- CASTAÑO-SANTAMARÍA, J. *et al.* Tree height prediction approaches for uneven-aged beech forests in northwestern Spain. **Forest Ecology and Management**, v. 307, p. 63-73, 2013.
- CERQUEIRA, C. L. *et al.* Modelagem Da Altura De *Pinus caribaea* var. *hondurensis* Em Diferentes Espaçamentos. **BIOFIX Scientific Journal**, v. 2, n. 2, p. 93-99, 2017.
- CICEU, A. *et al.* A generalized nonlinear mixed-effects height–diameter model for Norway spruce in mixed-uneven aged stands. **Forest Ecology and Management**, v. 477, p. 118507, 2020.
- CLUTTER, J. L. *et al.* **Timber management: A quantitative approach**. John Wiley & Sons, Inc., 1983.
- COSTA, E. A. *et al.* Enhancing height predictions of Brazilian pine for mixed, uneven-aged forests using artificial neural networks. **Forests**, v. 13, n. 8, p. 1284, 2022.
- CRECENTE-CAMPO, F. *et al.* A generalized nonlinear mixed-effects height–diameter model for *Eucalyptus globulus* L. in northwestern Spain. **Forest Ecology and Management**, v. 259, n. 5, p. 943-952, 2010.
- CRECENTE-CAMPO, F. *et al.* A generalized nonlinear mixed-effects height–diameter model for *Eucalyptus globulus* L. in northwestern Spain. **Forest Ecology and Management**, v. 259, n. 5, p. 943-952, 2010.
- CUI, K. *et al.* Estimating height-diameter relations for structure groups in the natural forests of Northeastern China. **Forest Ecology and Management**, v. 519, p. 120298, 2022.
- CUI, K. *et al.* Estimating height-diameter relations for structure groups in the natural forests of Northeastern China. **Forest Ecology and Management**, v. 519, p. 120298, 2022.
- CURTIS, R. O. Height–diameter and height–diameter–age equations for second growth Douglas-fr. **Forest Science**, v. 13, p. 365–375, 1967.
- DANTAS, D. *et al.* Accuracy of tree height estimation with model extracted from artificial neural network and new linear and nonlinear models. **Acta Scientiarum. Agronomy**, v. 46, n. 1, p. e63286-e63286, 2024.
- DANTAS, D. *et al.* Multilevel nonlinear mixed-effects model and machine learning for predicting the volume of *Eucalyptus* spp. trees. **Cerne**, v. 26, p. 48-57, 2020.

DIAMANTOPOULOU, M. J. Assessing a reliable modeling approach of features of trees through neural network models for sustainable forests. **Sustainable Computing: Informatics and Systems**, v. 2, n. 4, p. 190-197, 2012.

DIAMANTOPOULOU, M. J. *et al.* Estimation of Weibull function parameters for modelling tree diameter distribution using least squares and artificial neural networks methods. **Biosystems Engineering**, v. 133, p. 33-45, 2015.

DIAMANTOPOULOU, M. J. Tree-bole volume estimation on standing pine trees using cascade correlation artificial neural network models. **Agricultural Engineering International: CIGR Journal**, 2006.

DONG, L.; ZHANG, Y.; ZHANG, Z.; XIE, L.; LI, F. Comparison of tree biomass modeling approaches for larch (*Larix olgensis* Henry) trees in Northeast China. **Forests**, v. 11, n. 2, p. 202, 2020.

DORADO, F. C. *et al.* A generalized height–diameter model including random components for radiata pine plantations in northwestern Spain. **Forest Ecology and Management**, v. 229, n. 1-3, p. 202-213, 2006.

FLORIANO, E. P. *et al.* Fitting and selecting traditional models for tree's height time series data. **Ciência Florestal**, v. 16, n. 2, p. 177-199, 2006.

GÓMEZ-GARCÍA, E. *et al.* A comparison of model forms for the development of height–diameter relationships in even-aged stands. **Forest Science**, v. 60, n. 3, p. 560-568, 2014.

HALL, D. B.; BAILEY, R. L. Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models. **Forest science**, v. 47, n. 3, p. 311-321, 2001.

HUANG, S.; TITUS, S. J.; WIENS, D. P. Comparison of nonlinear height–diameter functions for major Alberta tree species. **Canadian Journal of Forest Research**, v. 22, n. 9, p. 1297-1304, 1992.

INDÚSTRIA BRASILEIRA DE ÁRVORES. **Relatório anual IBÁ 2023**. São Paulo, 2023.

JAYALAKSHMI, T.; SANTHAKUMARAN, A. Statistical normalization and back propagation for classification. **International Journal of Computer Theory and Engineering**, v. 3, n. 1, p. 1793-8201, 2011.

JÚNIOR, I. S. T. *et al.* Machine learning and regression models to predict multiple tree stem volumes for teak. **Southern Forests**, v. 83, n. 4, p. 01-09, 2021.

JÚNIOR, I. S. T. *et al.* Machine learning: Modeling increment in diameter of individual trees on Atlantic Forest fragments. **Ecological Indicators**, v. 117, p. 106685, 2020.

KARATEPE, Y. *et al.* Total tree height predictions via parametric and artificial neural network modeling approaches. **iForest-Biogeosciences and Forestry**, v. 15, n. 2, p. 95, 2022.

KEARSLEY, E. *et al.* Conventional tree height–diameter relationships significantly overestimate aboveground carbon stocks in the Central Congo Basin. **Nature communications**, v. 4, n. 1, p. 2269, 2013.

LARSEN, D. R.; HANN, D. W. **Height-diameter equations for seventeen tree species in southwest Oregon**. Forest Research Laboratory, Oregon State University, Corvallis. Res. Pap. 49., 161 p, 1987.

LEDUC, Daniel J. **Predicting diameter distributions of longleaf pine plantations: a comparison between artificial neural networks and other accepted methodologies**. US Department of Agriculture, Forest Service, Southern Research Station, 2001.

LEVIN, I.; NITSAN, J. Use of the Mitscherlich equation in designing factorial fertilizer field experiments to reduce the number of treatments. **Plant and Soil**, v. 21, n. 2, p. 249-252, 1964.

LIMA, R. B. *et al.* Estimating tree volume of dry tropical forest in the Brazilian Semi-Arid Region: a comparison between regression and artificial neural networks. **Journal of Sustainable Forestry**, v. 40, n. 3, p. 281-299, 2021.

LIN, Fu. *et al.* Comparison of Modeling Approaches for the Height–diameter Relationship: An Example with Planted Mongolian Pine (*Pinus sylvestris* var. *mongolica*) Trees in Northeast China. **Forests**, v. 13, n. 8, p. 1168, 2022.

LONG, S. *et al.* Influence of slope, aspect and competition index on the height-diameter relationship of *Cyclobalanopsis glauca* trees for improving prediction of height in mixed forests. **Silva Fennica**, v. 54, n. 1, 2020.

LYNCH, T. B.; MURPHY, P. A. A compatible height prediction and projection system for individual trees in natural, even-aged shortleaf pine stands. **Forest Science**, v. 41, n. 1, p. 194-209, 1995.

MEHTÄTALO, L. A longitudinal height–diameter model for Norway spruce in Finland. **Canadian Journal of Forest Research**, v. 34, n. 1, p. 131-140, 2004.

MENSAH, S. *et al.* Height–Diameter allometry in South Africa’s indigenous high forests: Assessing generic models performance and function forms. **Forest Ecology and Management**, v. 410, p. 1-11, 2018.

NÄSLUND, M. Skogsförsöksanstaltens gallringsförsök i tallskog. **Meddelanden från Statens skogsförsöksanstalt**, 1936.

NICOLETTI, M. F. *et al.* Relação hipsométrica para *Pinus taeda* L. em diferentes fases do ciclo de corte. **Floresta e Ambiente**, v. 23, p. 80-89, 2016.

NIKLAS, K. J. Size-dependent allometry of tree height, diameter and trunk-taper. **Annals of botany**, v. 75, n. 3, p. 217-227, 1995.

OGANA, F. N. A mixed-effects height-diameter model for *Gmelina arborea* Roxb stands in Southwest Nigeria. **Journal of Forest Research**, v. 27, n. 1, p. 1-7, 2021.

OGANA, F. N.; CORRAL-RIVAS, S.; GORGOSO-VARELA, J. Nonlinear mixed-effect height-diameter model for *Pinus pinaster* Ait. and *Pinus radiata* D. Don. **Cerne**, v. 26, p. 150-161, 2020.

OGANA, F. N.; ERCANLI, I. Modelling height-diameter relationships in complex tropical rain forest ecosystems using deep learning algorithm. **Journal of Forestry Research**, v. 33, n. 3, p. 883-898, 2022.

OLIVEIRA, R. R. *et al.* Acurácia de relações hipsométricas para diferentes estratégias de validação em *Eucalyptus urograndis*. **Revista Verde de Agroecologia e Desenvolvimento Sustentável**, v. 11, n. 5, p. 123-127, 2016.

ÖZÇELİK, R. *et al.* Estimating Crimean juniper tree height using nonlinear regression and artificial neural network models. **Forest ecology and management**, v. 306, p. 52-60, 2013.

ÖZÇELİK, R. *et al.* Estimating tree bole volume using artificial neural network models for four species in Turkey. **Journal of environmental management**, v. 91, n. 3, p. 742-753, 2010.

PARRESOL, B. R. Baldcypress height–diameter equations and their prediction confidence intervals. **Canadian journal of forest research**, v. 22, n. 9, p. 1429-1434, 1992.

PENG, C. **Developing ecoregion-based height-diameter models for jack pine and black spruce in Ontario**. Sault Ste. Marie: Ontario Forest Research Institute, 2001.

PENG, C. *et al.* Developing and evaluating tree height-diameter models at three geographic scales for black spruce in Ontario. **Northern Journal of Applied Forestry**, v. 21, n. 2, p. 83-92, 2004.

PINHEIRO, J; BATES, D. **Mixed-effects models in S and S-PLUS**. Springer science & business media, 2006.

RAPTIS, Dimitrios I. *et al.* Mixed-effects height–diameter models for black pine (*Pinus nigra* Arn.) forest management. **Trees**, v. 35, p. 1167-1183, 2021.

RATKOWSKY, D. A. **Handbook of non-linear regression models**. New York, USA: Marcel Dekker, Inc, 1990.

RESENDE, M. D. V.; SILVA, F. F.; AZEVEDO, C. F. **Estatística matemática, biométrica e computacional: Modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão aleatória, seleção genômica, QTL-GWAS, estatística espacial e temporal, competição, sobrevivência**. Viçosa: Suprema gráfica e Editora Ltda, 2014, 882p.

RIEDMILLER, M. Advanced supervised learning in multi-layer perceptrons—from backpropagation to adaptive learning algorithms. **Computer Standards & Interfaces**, v. 16, n. 3, p. 265-278, 1994.

RODRIGUES, C. K. *et al.* Identidade de modelos hipsométricos e volumétricos de povoamentos de *Pinus* com diferentes idades submetidos ao desbaste mecanizado. **BIOFIX Scientific Journal**, v. 2, n. 1, p. 54-59, 2017.

SAHIN, A. *et al.* Estimation of tree height with machine learning techniques in coppice-originated pure sessile oak (*Quercus petraea* (Matt.) Liebl.) stands. **Scandinavian Journal of Forest Research**, v. 38, n. 1-2, p. 87-96, 2023.

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, 2018.

SÁNCHEZ, C. A. L. *et al.* A height-diameter model for *Pinus radiata* D. Don in Galicia (Northwest Spain). **Annals of forest science**, v. 60, n. 3, p. 237-245, 2003.

SANQUETTA, C. R. *et al.* Relação hipsométrica em inventários pré-corte em povoamentos de *Acacia mearnsii* De Wild. **Científica**, v. 42, n. 1, p. 80-90, 2014.

SANTIAGO-GARCÍA, W. *et al.* Generalized height-diameter models for five pine species at Southern Mexico. **Forest Science and Technology**, v. 16, n. 2, p. 49-55, 2020.

SAUD, P. *et al.* Using quadratic mean diameter and relative spacing index to enhance height-diameter and crown ratio models fitted to longitudinal data. **Forestry**, v. 89, n. 2, p. 215-229, 2016.

Schumacher, F. X. A new growth curve and its application to timber yield studies. **Journal of Forestry**, v. 37, p. 819-20, 1939.

SHARMA, M.; PARTON, J. Height-diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach. **Forest Ecology and Management**, n. 249, p. 187-198, 2007.

SHEN, J. *et al.* Modeling height-diameter relationship for poplar plantations using combined-optimization multiple hidden layer back propagation neural network. **Forests**, v. 11, n. 4, p. 442, 2020.

SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. **Redes neurais artificiais para engenharia e ciências aplicadas**. São Paulo: Artliber Editora, 2010.

SILVA, J. P. M. *et al.* Prognosis of forest production using machine learning techniques. **Information Processing in Agriculture**, 2021.

SKOVSGAARD, J. P.; VANCLAY, J. K. Forest site productivity: a review of the evolution of dendrometric concepts for even-aged stands. **Forestry**, v. 81, n. 1, p. 13-31, 2008.

SOARES, F. A. A. M. N. *et al.* Recursive diameter prediction and volume calculation of eucalyptus trees using Multilayer Perceptron Networks. **Computers and electronics in agriculture**, v. 78, n. 1, p. 19-27, 2011.

SOARES, G. M. *et al.* Artificial Neural Networks (ANN) For Height Estimation in A Mixed-Species Plantation of *Eucalyptus globulus* Labill and *Acacia mearnsii* De Wild. **Revista Árvore**, v. 45, 2021.

SOARES, P.; TOMÉ, M. Height–diameter equation for first rotation eucalypt plantations in Portugal. **Forest Ecology and Management**, v. 166, n. 1-3, p. 99-109, 2002.

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, 2020.

STAGE, A. R. **Prediction of height increment for models of forest growth**. Intermountain Forest and Range Experiment Station, Forest Service, US Department of Agriculture, 1975.

TEMESGEN, H. *et al.* Generalized height–diameter models—an application for major tree species in complex stands of interior British Columbia. **European Journal of Forest Research**, v. 123, n. 1, p. 45-51, 2004.

TEMESGEN, H.; MONLEON, V. J.; HANN, D. W. Analysis and comparison of nonlinear tree height prediction strategies for Douglas-fir forests. **Canadian Journal of Forest Research**, v. 38, n. 3, p. 553-565, 2008.

THANH, T. N.; TIEN, T. D.; SHEN, H. L. Height-diameter relationship for *Pinus koraiensis* in Mengjiagang Forest Farm of Northeast China using nonlinear regressions and artificial neural network models. **Journal of Forest Science**, v. 65, n. 4, p. 134-143, 2019.

TIMILSINA, N.; STAUDHAMMER, C. L. Individual tree-based diameter growth model of slash pine in Florida using nonlinear mixed modeling. **Forest Science**, v. 59, n. 1, p. 27-37, 2013.

TRIM, K. R. *et al.* A new site index model for intensively managed loblolly pine (*Pinus taeda*) plantations in the West Gulf Coastal Plain. **Forest Science**, v. 66, n. 1, p. 2-13, 2020.

TRINCADO, G.; VANDERSCHAAF, C. L.; BURKHART, H. E. Regional mixed-effects height–diameter models for loblolly pine (*Pinus taeda* L.) plantations. **European Journal of Forest Research**, v. 126, n. 2, p. 253-262, 2007.

UZOH, F. Height-diameter model for managed even-aged stands of Ponderosa pine for the Western United States using hierarchical nonlinear mixed-effects model. **Australian Journal of Basic and Applied Science**, v. 11, n. 14, p. 69-87, 2017.

VONESH, E.; CHINCHILLI, V. M. **Linear and nonlinear models for the analysis of repeated measurements**. CRC press, 1996.

WANG, C. H.; HANN, D. W. **Height-diameter equations for sixteen tree species in the central western Willamette valley of Oregon**. Forest Research Laboratory Oregon State University, Corvallis, 1988.

WANG, C. H.; HANN, D. W. **Height-diameter equations for sixteen tree species in the central western Willamette Valley of Oregon**. 1988.

WINSOR, C. P. The Gompertz curve as a growth curve. **Proceedings of the national academy of sciences**, v. 18, n. 1, p. 1-8, 1932.

YANG, S. I.; BURKHART, H. E. Evaluation of total tree height subsampling strategies for estimating volume in loblolly pine plantations. **Forest ecology and management**, v. 461, p. 117878, 2020.

ZHANG, X.; LEI, Y.; LIU, X. Modeling stand mortality using Poisson mixture models with mixed-effects. *iForest-Biogeosciences and Forestry*, v. 8, n. 3, p. 333, 2015.

ARTIGO 2 – Modelagem do afileamento do fuste de espécies de *Pinus* com modelos de efeito misto e redes neurais artificiais

ARTIGO FORMATADO DE ACORDO COM AS NORMAS DA NBR 6022 (ABNT, 2018)

RESUMO

O presente estudo desenvolveu e analisou funções de afilamento/taper baseadas em técnicas de modelagem mista e redes neurais artificiais para povoamentos de *Pinus caribaea* var. *caribaea*, *Pinus caribaea* var. *hondurensis* e *Pinus oocarpa*. Os dados utilizados no presente estudo foram coletados em plantios de situados no município de Nova Ponte, Minas Gerais. Foram utilizados dados provenientes da cubagem rigorosa de 485 árvores de pinus. A base de dados foi separada em treino (80%) e teste (20%). Para descrever o perfil individual das árvores, por meio dos modelos mistos, foi incorporada, ao modelo selecionado por espécie, a variabilidade entre árvore, onde foram consideradas todas as combinações possíveis de um e dois parâmetros aleatórios. Além disso, a estrutura de erro autoregressiva contínua de primeira ordem CAR(1) foi adicionada para obtenção de estimativas sem viés dos parâmetros e as predições calibradas foram obtidas e avaliadas para diferentes alternativas de amostragem e diferentes tamanhos das subamostras dentro de cada árvore da base de teste. Modelos de redes neurais artificiais também foram desenvolvidos para estimar os diâmetros a qualquer altura dos fustes neste estudo, onde avaliou-se diferentes configurações de parâmetros. O desempenho dos diferentes tipos de modelos testados foi avaliado com base na raiz do erro quadrático médio (RMSE), erro médio absoluto percentual (MAPE), desvio médio (BIAS), índice de eficiência do modelo (EF) e Critério de informação de Akaike (AIC). Para comparação dos modelos mistos entre si, foi utilizado o critério de Informação de Akaike (AIC), o critério Bayesiano de Schwarz (BIC), o logaritmo da verossimilhança maximizada (LogLik) e o teste da razão de verossimilhança (MLRT). Também, gráficos de distribuição dos resíduos foram construídos para auxiliar na avaliação dos modelos avaliados. Por fim, testes de médias foram aplicados, comparando os valores observados com os preditos. Além disso, os valores preditos de volume total também foram obtidos e avaliados com os mesmos critérios utilizados para as predições de diâmetro. A equação de forma variável de Kozak (2004) foi a alternativa mais adequada dentre as técnicas de regressão avaliadas e, com a adição de parâmetros de efeito aleatório e estrutura de erro autoregressiva contínua de primeira ordem houve melhoras significativas, tornando os modelos mais coerentes com a natureza dos dados. Após a calibração das equações, os modelos resultaram em predições de diâmetro ao longo do fuste e volume total estatisticamente iguais aos valores observados para as espécies *Pinus caribaea* var. *caribaea* e *Pinus oocarpa*. As redes neurais artificiais promoveram resultados satisfatórios de diâmetro ao longo do fuste e volume total para a espécie *Pinus caribaea* var. *hondurensis*, com valores preditos não diferindo estatisticamente dos observados. O comportamento dos resíduos ao longo do fuste e a avaliação da propagação do erro para as predições volumétricas devem ser considerados na escolha do modelo de afilamento para uma seleção mais adequada. Além disso, recomenda-se a adição de outros atributos a nível de parcela, a fim de melhorar a representação do afilamento dos fustes das árvores, desde que a coleta destas não comprometam a viabilidade das operações de inventário florestal.

Palavras-chave: Efeito aleatório; Aprendizagem de máquina; Kozak; Equação taper.

ABSTRACT

The present study developed and analyzed functions based on mixed modeling techniques and artificial neural networks for *Pinus caribaea* var. *caribaea*, *Pinus caribaea* var. *hondurensis* and *Pinus oocarpa*. The data used in the present study were collected in plantations located in the municipality of Nova Ponte, Minas Gerais. Data from the rigorous cubing of 485 pine trees were used. The database was separated into training (80%) and testing (20%). To describe the individual profile of the trees, through mixed models, the variability between trees was incorporated into the model selected by species, where all possible combinations of one and two random parameters were considered. Furthermore, the first-order continuous autoregressive error structure CAR(1) was added to obtain unbiased parameter estimates and calibrated predictions were obtained and evaluated for different sampling alternatives and different subsample sizes within each tree in the testing database. Artificial neural network models were also developed to estimate the diameters at any height of the stem in this study, where different parameter configurations were evaluated. The performance of the different types of models tested was evaluated based on root mean square error (RMSE), mean absolute percentage error (MAPE), mean deviation (BIAS), model efficiency index (EF), and Akaike Information Criterion (AIC). To compare the mixed models with each other, the Akaike Information Criterion (AIC), the Schwarz Bayesian Criterion (BIC), the maximized logarithm of likelihood (LogLik), and the likelihood ratio test (MLRT) were used. Also, residual distribution graphs were constructed to assist in the evaluation of the evaluated models. Finally, mean tests were applied, comparing the observed values with the predicted ones. Furthermore, the predicted total volume values were also obtained and evaluated with the same criteria used for diameter predictions. The variable form equation of Kozak (2004) was the most appropriate alternative among the regression techniques evaluated and, with the addition of random effect parameters and continuous first-order autoregressive error structure, there were significant improvements, making the models more coherent with the nature of the data. After calibrating the equations, the models resulted in predictions of diameter along the stem and total volume statistically equal to the values observed for the species *Pinus caribaea* var. *caribaea* and *Pinus oocarpa*. Artificial neural networks provided satisfactory results for diameter along the stem and total volume for the species *Pinus caribaea* var. *hondurensis*, with predicted values not statistically different from those observed. The behavior of residues along the stem and the evaluation of error propagation for volumetric predictions must be considered when choosing the tapering model for a more appropriate selection. Furthermore, it is recommended to add other attributes at the plot level, to improve the representation of the tapering of tree boles, if their collection does not compromise the viability of forest inventory operations.

Keywords: Random effect; Machine learning; Kozak; Taper equation.

1 INTRODUÇÃO

Medidas de volume de árvores acuradas é essencial para a descrição de plantios comerciais. Para a determinação do volume é necessária a disponibilidade de informações provenientes de métodos de amostragem a nível de árvore. No entanto, esse processo possui alto custo, devido à demanda de tempo e investimento de mão de obra (Schikowski *et al.*, 2018). Uma alternativa de uso muito frequente na área florestal para estimação deste atributo está relacionada à utilização de métodos indiretos como equações volumétricas e de afilamento. As equações de afilamento desempenham um papel muito importante em inventários florestais e projeções de crescimento, assim como no próprio planejamento florestal. A partir dessas equações é possível obter informações relevantes para nortear a tomada de decisões nos níveis individual de árvore e da floresta (Trincado; Burkhart, 2006; Salekin *et al.*, 2021).

As funções de afilamento permitem modelar a redução do diâmetro do fuste em função do aumento da altura ao longo do tronco da árvore. Isso possibilita a predição do diâmetro a qualquer altura ao longo do fuste, permitindo assim a obtenção de volumes comerciais para qualquer tamanho de seção e diâmetro mínimo desejado. Dessa forma, é possível realizar o sortimento dos multiprodutos da floresta e otimizar a destinação destes produtos, maximizando assim a receita econômica dos empreendimentos florestais. Além disso, essas equações também possuem aplicabilidade em estudos da qualidade do fuste, modelagem da alocação de carbono em diferentes porções do tronco das árvores e, também, na avaliação de impactos de tratamentos silviculturais na forma do fuste (Fonweban *et al.* 2011).

Ao longo das últimas décadas, equações de afilamento têm sido intensivamente estudadas ao redor do mundo. Conseqüentemente, diversas funções têm sido desenvolvidas e aperfeiçoadas, permitindo o ganho de flexibilidade e aplicabilidade para diversas espécies comerciais (*Pinus*, *Eucalyptus*, *Tectona*) (Li *et al.*, 2012; Stefanello *et al.*, 2021; Hirigoyen *et al.*, 2021; Rocha *et al.*, 2022; Santos *et al.*, 2023).

A modelagem por meio da análise de regressão é o método de estimação de parâmetros mais utilizado no desenvolvimento de equações de afilamento (Kozak, 2004; Amarioarei *et al.*, 2020; Liu *et al.*, 2020). No entanto, as inferências estatísticas a partir de equações de regressão não linear deve obedecer pressuposições como homocedasticidade da variância e ausência de autocorrelação dos resíduos e, nos estudos florestais, essas pressuposições são raramente atendidas nas aplicações práticas (Sandoval; Acuña, 2022).

Os dados utilizados para a construção das equações de afilamento do fuste possuem medidas repetidas de diâmetro na mesma árvore em diferentes alturas, conferindo a esses dados

uma estrutura hierárquica onde espera-se a presença de correlação. Dessa forma, a pressuposição de independência do erro exigidas pelos métodos da regressão clássica são violadas, levando à erros na estimação dos parâmetros (Yang *et al.*, 2009). A técnica de modelos de efeito misto têm sido testada para modelagem do afilamento do fuste. Esses modelos são uma alternativa eficiente de lidar com estruturas hierárquicas e medidas repetidas, permitindo assim uma estimativa mais confiável dos parâmetros (Zhang *et al.*, 2021).

Além da modelagem mista, técnicas de aprendizagem de máquinas, como as redes neurais artificiais (RNA) têm sido cada vez mais estudadas na área florestal, uma vez que elas podem resultar em boas previsões sem a necessidade de atender pressupostos estatísticos. As RNA também se destacam por serem comparáveis às técnicas tradicionais de modelagem estatística e por possuírem grande capacidade de lidar com relações não lineares complexas entre variáveis, o que não pode ser obtido com modelos de regressão, onde uma função não linear apropriada deve ser encontrada previamente (Lecun *et al.*, 2015; Özçelik *et al.*, 2010).

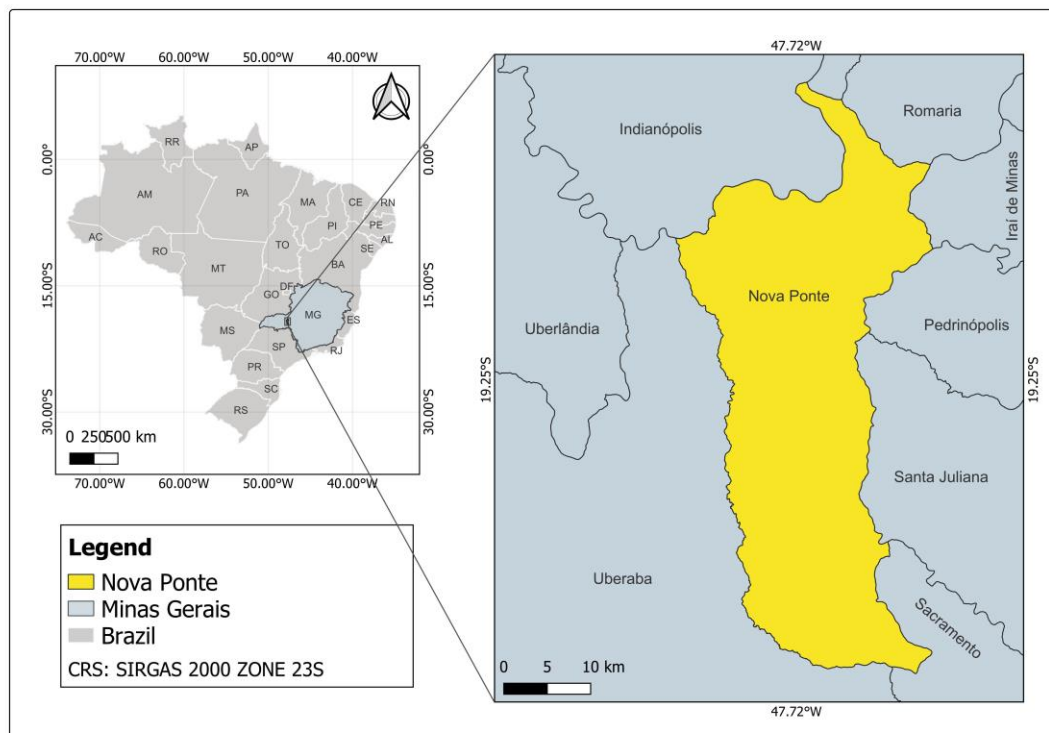
Em um levantamento realizado por Terra *et al.* (2020), dos trabalhos realizados em plantios brasileiros de *Pinus*, os modelos predominantemente utilizados são os modelos de Hradetzky (1976) e Schöpfer (1966), evidenciando a necessidade de estudos considerando outros tipos de modelos e técnicas. Desse modo, a fim contribuir com informações sobre técnicas mais eficientes para a modelagem do perfil do fuste de árvores comerciais no Brasil, o presente estudo desenvolveu e analisou funções baseadas em técnicas de modelagem mista e redes neurais artificiais para povoamentos de *Pinus caribaea* var. *caribaea*, *Pinus caribaea* var. *hondurensis* e *Pinus oocarpa* localizados na região sudeste do país.

2 MATERIAL E MÉTODOS

2.1 Caracterização da área de estudo

Os dados utilizados neste estudo foram obtidos a partir de plantios de pinus pertencentes à empresa Caxuana, localizada no município de Nova Ponte, Minas Gerais (Figura 1). O local possui uma precipitação média anual de 1.520, temperatura média anual de 22,2 °C, altitude média de 980 m e umidade relativa média igual a 51%. Além disso, o relevo é, predominantemente, plano a suave ondulado e os solos são dos tipos LATOSSOLO Vermelho escuro e LATOSSOLO Vermelho-amarelo (Horle *et al.*, 2010).

Figura 1 – Mapa de localização do município de Nova Ponte, MG.



Fonte: Do autor (2024).

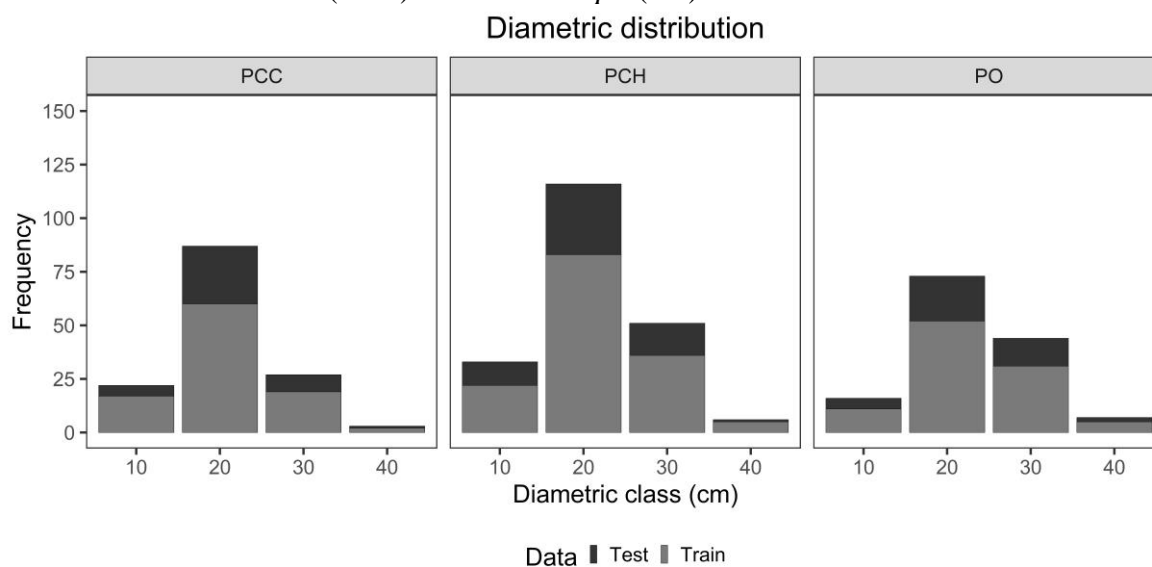
2.2 Base de dados

Foram utilizados dados provenientes da cubagem rigorosa de 139, 206 e 140 árvores de *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO), respectivamente, com idade entre 8 e 30 anos e com espaçamento inicial de 3 x 2 metros. Na cubagem foram medidos, com o auxílio de suta, diâmetros com casca nas alturas relativas de 0,5%; 1%; 5%; 10%; 15%; 25%; 35%; 45%; 55%; 65%; 75%; 85% e 95% da altura

total da árvore, diâmetro a 1,30 m de altura do solo (DAP) e a altura total (HT) de todas as árvores cubadas.

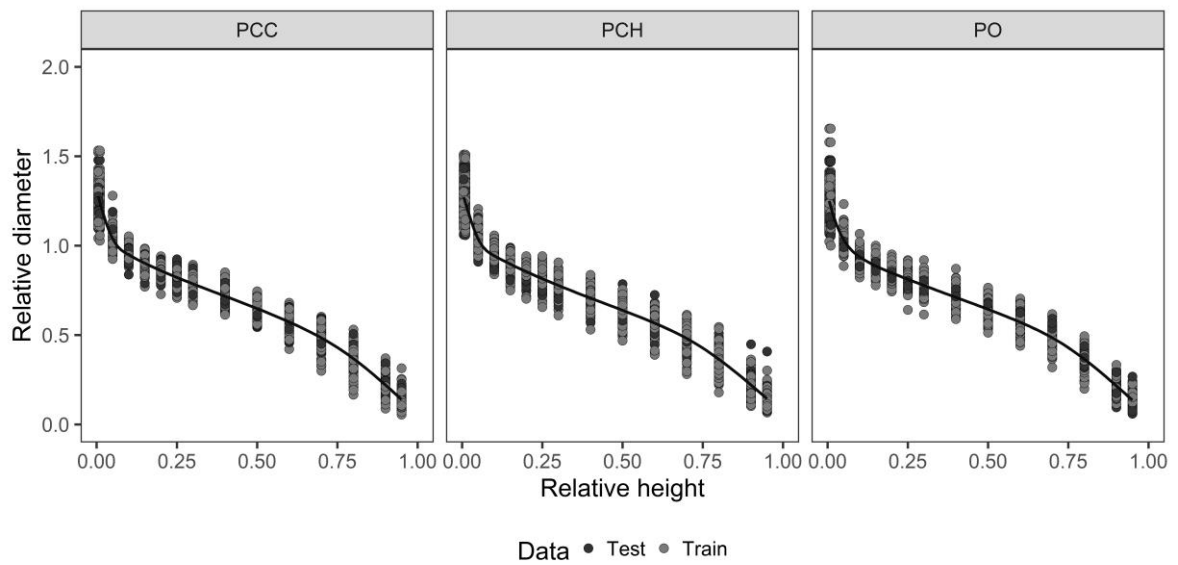
De posse dos dados, a presença de medidas anormais foi avaliada visualmente por meio de gráficos de altura *versus* diâmetro ao longo do fuste para todas as árvores. As anomalias foram corrigidas, se fossem consideradas como erros de medição ou de anotações de campo, ou removidas, se fossem consideradas como deformações causadas pela presença de nós ou outros tipos de defeitos severos no fuste. Também, foi realizada a distribuição de frequência de indivíduos por classe diamétrica por espécie, onde foram definidas quatro classes diamétricas (Figura 2), sendo os valores das classes diamétricas representados pelo limite superior de cada classe. Além disso, o volume total de cada indivíduo foi calculado a partir da soma dos volumes de todas as seções, realizado pelo método de Smalian. Os valores de diâmetro relativo (diâmetro com casca / DAP) e altura relativa (altura da seção / HT) foram calculados (Figura 3). Os gráficos de dispersão dos dados de volume total em função do DAP e HT podem ser observados na Figura 4.

Figura 2 – Distribuições diamétricas das espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).



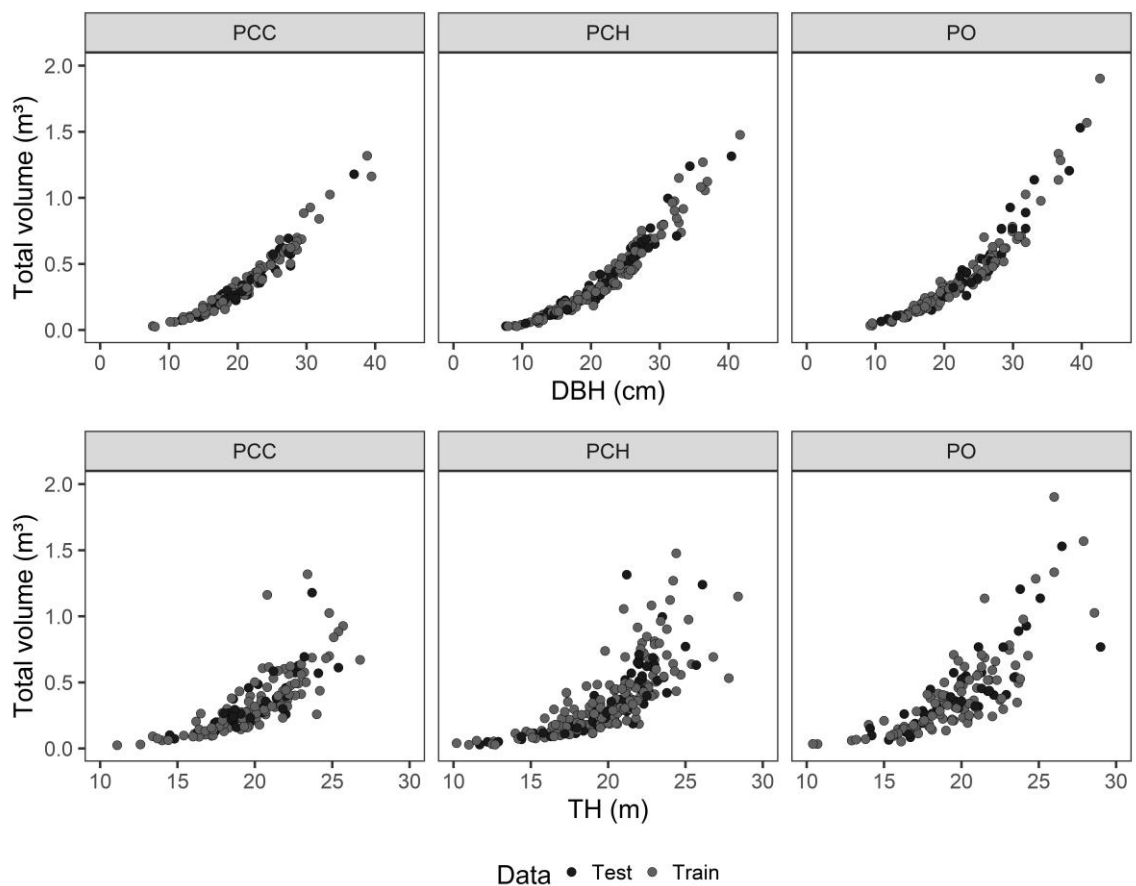
Fonte: Do autor (2024).

Figura 3 – Descrição dos fustes das espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).



Fonte: Do autor (2024).

Figura 4 – Gráfico de distribuição do volume total em função do DAP e altura total para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).



Fonte: Do autor (2024).

A base de dados foi separada em 80% para a base de treino/ajuste e 20% para a base de teste, de modo que a seleção das árvores ocorreu de forma proporcional por classe diamétrica para as três espécies e, quando uma árvore foi selecionada, todas as suas medidas ao longo do fuste foram incluídas também, de modo a preservar a estrutura de correlação das medidas repetidas. A estatística descritiva dos dados pode ser observada na Tabela 1.

Tabela 1 – Medidas de dispersão e variabilidade das principais variáveis dendrométricas utilizadas no presente estudo.

		TRAIN (trees =342, n = 5145)						TEST (trees =143, n = 2130)					
Species	Variable	n	Min.	Max.	Mean	Std. Dev.	CV (%)	n	Min.	Max.	Mean	Std. Dev.	CV (%)
PCC	di (cm)	1470	1,27	44,88	15,11	8,19	54,23	615	1,59	46,15	15,09	7,76	51,45
	DBH (cm)		7,64	39,47	20,59	6,04	29,32		11,78	36,92	20,58	4,89	23,77
	TH (m)		11,10	26,80	19,95	3,01	15,09		14,50	25,40	19,74	2,38	12,07
	Volume (m ³)		0,0242	1,3183	0,3543	0,2467	69,64		0,0734	1,1789	0,3315	0,2087	62,97
PCH	di (cm)	2190	0,95	47,75	15,61	8,60	55,11	900	1,27	49,34	15,61	8,50	54,43
	DBH (cm)		7,96	41,70	21,45	6,39	29,79		7,64	40,43	21,24	6,31	29,72
	TH (m)		10,20	28,40	19,97	3,16	15,84		11,70	26,10	20,22	3,10	15,33
	Volume (m ³)		0,0275	1,4763	0,3795	0,2746	72,37		0,0286	1,3147	0,3869	0,2729	70,54
PO	di (cm)	1485	1,27	48,38	16,14	8,85	54,82	615	1,27	48,38	17,09	9,06	53,00
	DBH (cm)		9,23	42,65	22,27	6,77	30,38		10,82	39,79	23,49	6,32	26,93
	TH (m)		10,40	28,60	19,58	3,26	16,62		14,10	29,00	20,07	3,21	16,01
	Volume (m ³)		0,0328	1,9027	0,4069	0,3253	79,96		0,0640	1,5295	0,4581	0,3225	70,39

2.3 Funções de afilamento

Dentre as diversas técnicas estatísticas de modelagem aplicadas à descrição do perfil do fuste, pode-se citar (i) Modelos Polinomiais (Schöepfer 1966; Bruce *et al.*, 1968; Hradetzky, 1976), (ii) Modelos Segmentados (Max; Burkhart, 1976; Parresol *et al.*, 1987; Clark *et al.*, 1991; Fang *et al.*, 2000), (iii) Modelos Sigmoidais (Ormerod, 1973; Guimarães; Garcia, 1992), (iv) Modelos Compatíveis (Demaerschalk, 1972), (v) Modelos de Forma Variável Altamente Flexível (Bi, 2000; Kozak, 2004) e (vi) Modelos de razão entre volumes (Amateis *et al.*, 1987). Neste estudo, oito equações de afilamento foram avaliadas, para cada espécie, por meio de regressão linear e não linear, contemplando as técnicas i, ii, iii, iv e v (Tabela 2).

Tabela 2 – Oito funções de afilamento do fuste e suas respectivas expressões matemáticas.

Model	Expression
Shöepfer (1996)	$\frac{d_{ij}}{D_i} = \beta_0 + \beta_1(Z) + \beta_2(Z)^2 + \beta_3(Z)^3 + \beta_4(Z)^4 + \beta_5(Z)^5 + \varepsilon_{ij}$
Hradetzky (1976)	$\frac{d_{ij}}{D_i} = \beta_0 + \beta_1(Z)^{p_1} + \beta_2(Z)^{p_2} + \dots + \beta_n(Z)^{p_n} + \varepsilon_{ij}$
Max and Burkhart (1976)	$\left(\frac{d_{ij}}{D_i}\right)^2 = \beta_1(Z - 1) + \beta_2((Z)^2 - 1) + \beta_3(a_1 - Z)^2 I_1 + \beta_4(a_2 - Z)^2 I_2 + \varepsilon_{ij}$ $I_1 = \begin{cases} 1 & \text{se } Z \leq a_1 \\ 0 & \text{se } Z > a_1 \end{cases} \quad I_2 = \begin{cases} 1 & \text{se } Z \leq a_2 \\ 0 & \text{se } Z > a_2 \end{cases}$
Parresol (1987)	$\left(\frac{d_{ij}}{D_i}\right)^2 = (1 - Z)^2(\beta_1 + \beta_2(1 - Z))$ $+ ((1 - Z) - a_0)^2\{\beta_3 + \beta_4((1 - Z) + 2a_0)\}I + \varepsilon_{ij}$ $I = \begin{cases} 1 & \text{se } 1 - Z \geq a_0 \\ 0 & \text{se } 1 - Z < a_0 \end{cases}$
Ormerod (1973)	$\frac{d_{ij}}{D_i} = \left[\frac{H_i - h_{ij}}{H_i - 1,3}\right]^{\beta_1} + \varepsilon_{ij}$
Modified Ormerod (1992)	$\frac{d_{ij}}{D_i} = \left\{e^{\left[1 - e^{(1,3 - h_{ij})}\right]}\right\}^{\beta_1} \left[\frac{H_i - h_{ij}}{H_i - 1,3}\right]^{\beta_2} + \varepsilon_{ij}$
Demaerschalk (1972)	$\frac{d_{ij}}{D_i} = \beta_1 D_i^{\beta_2} (H_i - h_{ij})^{\beta_3} H_i^{\beta_4} + \varepsilon_{ij}$

Kozak (2004)

$$\frac{d_{ij}}{D_i} = \alpha_0 D_i^{\alpha_1} H_i^{\alpha_2} X^{[\beta_1 Z^4 + \beta_2 \left(\frac{1}{\exp\left(\frac{D_i}{H_i}\right)}\right) + \beta_3 X^{0,1} + \beta_4 \left(\frac{1}{D_i}\right) + \beta_5 H_i^Q + \beta_6 X]} + \varepsilon_{ij}$$

Onde $p_1 - p_n$ = expoente variando de 0,004 a 25. $X = \frac{\left(1 - \left(\frac{h_{ij}}{H_i}\right)^{\frac{1}{3}}\right)}{1 - q^{\frac{1}{3}}}$, $Q = \left(1 - Z^{\frac{1}{3}}\right)$, $Z = \frac{h_{ij}}{H_i}$, $q = \frac{1,3}{H_i}$; $\alpha_0 - \alpha_2$ e $\beta_0 - \beta_6$ são os parâmetros a serem estimados; q é a altura relativa onde as árvores mudam de neilóide para parabolóide, ou ponto de inflexão.

Os expoentes testados para o modelo de Hradetzky (1976), após uma análise prévia, foram 0,004; 0,005; 0,006; 0,007; 0,008; 0,009; 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1; 2; 3; 4; 5; 6; 7; 8; 9; 2; 21; 22; 23; 24 e 25. O método “Stepwise” foi aplicado para a seleção das variáveis preditoras do modelo.

As equações de estrutura linear foram ajustadas com o auxílio da função `lm()`, enquanto as de estrutura não linear foram ajustadas com auxílio da função `nls()` e o método “Stepwise” por meio da função `step()`, todas do pacote `stats`. Para os modelos não lineares, foi utilizado o algoritmo de Gauss-Newton.

O modelo de melhor desempenho por espécie foi selecionado e a este foi aplicada a modelagem mista, considerando o nível individual de árvore como fonte de variabilidade.

2.4 Modelos de efeitos mistos

Para descrever o perfil individual das árvores, por meio dos modelos mistos, foi incorporada, ao modelo selecionado anteriormente, a variabilidade entre árvore. O diâmetro da j -ésima seção na i -ésima árvore será modelado como:

$$d_{ij} = f(\boldsymbol{\beta}_{ij}, h_{ij}, D_i, H_i) + \varepsilon_{ij}, \quad (1)$$

$$i = 1, \dots, M, j = 1, \dots, n_i, \quad (2)$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \quad (3)$$

Onde $\boldsymbol{\beta}_{ij}$ é um vetor de p árvores e possíveis parâmetros específicos por seção, f é uma função (função do modelo selecionado) que relaciona o diâmetro de uma seção com sua respectiva altura, diâmetro a 1,3m do solo e altura total da árvore e ε_{ij} é o erro entre árvores que será considerado independente e com distribuição normal.

O k -ésimo elemento do vetor de parâmetros $\boldsymbol{\beta}_{ij}$ foi então modelado como uma função linear de efeitos fixo e aleatório:

$$\beta_{ij} = \mathbf{x}'_{ijk} \boldsymbol{\beta}_k + \mathbf{z}'_{ijk} \mathbf{b}_{ik}, \quad (4)$$

$$k = 0, \dots, p - 1, \quad (5)$$

$$\mathbf{b}_{ik} \sim N(\mathbf{0}, \sigma_k^2) \quad (6)$$

Onde $\boldsymbol{\beta}_k$ é um vetor de efeitos fixos e \mathbf{b}_{ik} é um vetor de efeitos aleatórios associados com a i -ésima árvore. Assume-se que \mathbf{b}_{ik} tem distribuição normal, com média 0 e matriz de covariância σ_k^2 e é independente de ε_{ij} . \mathbf{x}'_{ijk} e \mathbf{z}'_{ijk} possivelmente com a j -ésima seção.

Na forma matricial, temos:

$$\mathbf{d}_i = \mathbf{f}(\boldsymbol{\beta}_i, \mathbf{h}_i, \mathbf{D}_i, \mathbf{H}_i) + \boldsymbol{\varepsilon}_i \quad (7)$$

$$\boldsymbol{\beta}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \quad (8)$$

$$\mathbf{b}_i \sim N(\mathbf{0}, \sigma_k^2) \quad (9)$$

Combinando essas duas equações, temos:

$$\mathbf{H}_i = \mathbf{f}(\boldsymbol{\beta}, \mathbf{b}_i, \mathbf{X}_i, \mathbf{Z}_i) + \boldsymbol{\varepsilon}_i \quad (10)$$

Onde \mathbf{d}_i é o vetor de diâmetros das seções das árvores i , $\boldsymbol{\beta}_i$ é o vetor de parâmetros a nível de árvore, \mathbf{h}_i é o vetor de alturas das seções das árvores i , \mathbf{D}_i é o vetor de diâmetros a 1,3m do solo e \mathbf{H}_i é o vetor de alturas totais das árvores i , $\boldsymbol{\varepsilon}_i$ é o vetor de erros entre árvores, $\boldsymbol{\beta}$ é o vetor de parâmetros fixos que não dependem das árvores, \mathbf{b}_i é o vetor de efeitos aleatórios a nível de árvore e \mathbf{X}_i e \mathbf{Z}_i são as matrizes das variáveis independentes.

Uma etapa extremamente importante no ajuste de modelos de efeitos mistos é a escolha de qual ou quais parâmetros devem ser considerados como fixos e aleatórios (Pinheiro; Bates, 2000). Existem três métodos para realizar esta seleção: i) ajustar, individualmente, um modelo de efeitos fixos para cada árvore e selecionar os parâmetros que possuem maior variabilidade para serem expandidos com efeitos aleatórios (Fang; Bailey, 2001), ii) avaliar como o afilamento do fuste varia com variações individuais nos valores de vários parâmetros da função de afilamento (De-Miguel *et al.*, 2012) e iii) avaliar diferentes combinações de parâmetros a serem expandidos com efeitos aleatórios.

Como também observado por Arias-Rodil *et al.* (2015), a segunda alternativa apresentou duas limitações: 1) o ajuste individual não foi possível devido o baixo número de observações por árvore (15 observações) que, conseqüentemente, impossibilitou a obtenção de parâmetros significativos e 2) a alta variabilidade de parâmetros entre indivíduos pode não estar relacionada a uma alta variabilidade na forma do fuste, uma vez que os parâmetros podem se comportar de forma não linear nas equações. Dessa forma, no presente trabalho, foi utilizada a terceira alternativa, onde foram consideradas todas as combinações de expansão possíveis de um e dois parâmetros aleatórios.

O ajuste dos modelos mistos e as suas análises estatísticas foram conduzidas com auxílio do pacote nlme da linguagem de programação R (Pinheiro; Bates, 2000).

2.5 Autocorrelação

Em análises de regressão, assume-se que os resíduos são independentes, possuem distribuição normal e que são aleatórios, porém, para o desenvolvimento de equações de afilamento, utiliza-se o conjunto de múltiplas observações de cada árvore (ou seja, dados longitudinais). Dessa forma, espera-se que as observações dentro de cada árvore sejam espacialmente relacionadas, violando assim a pressuposição de independência dos resíduos.

Dois métodos podem ser citados para lidar com dados longitudinais contínuos, desbalanceados e multiníveis. A primeira alternativa é a adição de efeitos aleatório (Gregoire et al. 1995) e a segunda é a modelagem direta da estrutura de correlação. No entanto, a adição de efeitos aleatórios pode não ser suficiente para modelar a autocorrelação presente nos dados (Arias-Rodil *et al.*, 2015), dessa forma, avaliou-se a necessidade de adição de uma estrutura de correlação. A estrutura de correlação intra-individual mais comumente utilizada quando se trata de dados florestais é a estrutura de erro autorregressiva de primeira ordem (AR1) (Li; Weiskittel, 2010), porém, como os dados de afilamento ao longo do fuste foram coletados com distâncias distintas ao longo do fuste, optou-se a utilização da estrutura de erro autoregressiva contínua de primeira ordem CAR(1) para minimizar a autocorrelação dos dados utilizados, por ser uma versão contínua da AR1. A CAR(1) é dada por:

$$\text{Corr}(e_t e_s) = \varphi^{|t-s|} \quad (11)$$

Onde φ representa a correlação entre duas observações com distância de uma unidade, t e s representam duas observações de altura obtidas do mesmo indivíduo e $|t - s|$ é a distância entre elas, ou seja, o valor de correlação entre duas observações da mesma árvore depende da distância absoluta entre elas.

A utilização de CAR(1) assume que a autocorrelação diminui conforme a distância entre medidas sucessivas aumenta. Dessa forma, sua utilização objetivou a obtenção de estimativas sem viés dos parâmetros (Huang 1997; Parresol; Vissage 1998) e, para avaliar a presença de autocorrelação e o efeito da estrutura de autocorrelação dos resíduos utilizada, foram utilizados gráficos dos resíduos vs resíduos defasados de observações anteriores por árvore, além do teste da razão de verossimilhança entre o modelo sem e com a CAR(1).

A estrutura de correlação foi adicionada por meio da função corCAR1() repassada ao argumento *correlation* na função nlme().

2.6 Calibração dos modelos de efeito misto

Uma vantagem dos modelos de efeitos mistos é que, havendo a disponibilidade de, no mínimo, uma observação dos dados, estes podem ser utilizados para a predição do vetor de efeitos aleatórios (Özçelik *et al.*, 2013; Vonesh; Chinchilli, 1997). Este método pode ser chamado de calibração ou localização e, na área florestal, esta técnica pode ser empregada para a estimação dos parâmetros aleatórios de uma nova árvore que não estava presente anteriormente na base de ajuste do modelo (Hall; Bailey 2001). No presente estudo, a calibração foi realizada com a utilização de, no mínimo, os valores dos atributos de uma seção presente na nova árvore, juntamente de seu DAP e altura total, para estimação dos novos parâmetros aleatórios. Os valores da seção podem ser coincidentes aos da altura do DAP ou de quaisquer outras alturas desejadas.

Foi utilizado o algoritmo de Lindstrom-Bates por meio do método FOCE para o processo de calibração, onde os valores do vetor de parâmetros aleatório \mathbf{b}_i , associado à árvore, pode ser estimado a partir da equação abaixo (Camacho *et al.*, 2022; Vonesh; Chinchilli, 1997).

$$\mathbf{b}_i = \widehat{\mathbf{D}}\widehat{\mathbf{Z}}_i^T(\widehat{\mathbf{R}}_i + \widehat{\mathbf{Z}}_i\widehat{\mathbf{D}}\widehat{\mathbf{Z}}_i^T)^{-1}\widehat{\mathbf{e}}_i \quad (12)$$

Onde $\widehat{\mathbf{D}}$ é a matriz de variância e covariância $q \times q$ associada com os parâmetros aleatórios (q é o número de parâmetros aleatórios no modelo) de todas as árvores, $\widehat{\mathbf{R}}_i$ é a matriz de variância e covariância $m \times m$ para a variabilidade dentro das árvores, $\widehat{\mathbf{e}}_i$ é o vetor de resíduos $m \times 1$ (onde é calculado pela diferença entre os diâmetros observados de cada seção dentro da subamostra e os valores preditos a partir do modelo contendo apenas os efeitos fixos), $\widehat{\mathbf{Z}}_i$ é a matriz $m \times q$ das derivativas parciais dos parâmetros aleatórios avaliados em $\widehat{\mathbf{b}}_i$. Uma vez que $\widehat{\mathbf{b}}_i$ é predito, os valores dos diâmetros podem ser calculados a partir do modelo calibrado.

Para os modelos de efeitos mistos selecionados, as predições calibradas foram obtidas e avaliadas para diferentes alternativas de amostragem e diferentes tamanhos das subamostras dentro de cada árvore da base de teste. As alternativas de calibração foram: 1) Diâmetro a 1,3 m do solo por árvore. 2) Diâmetro a 5,2 m do solo por árvore. 3) Diâmetros a 1,3 e 5,2 m do solo por árvore e 4) Sem calibração, utilizando apenas os parâmetros de efeito fixo para a predição. Preferiu-se a utilização de diâmetros a alturas absolutas ao invés de alturas relativas, pois suas obtenções são mais práticas nas operações florestais.

No presente estudo, o diâmetro 5,2 m do solo é uma variável necessária em duas das três alternativas de calibração dos modelos de efeito misto. No entanto, essa variável é de difícil

acesso e requer uma maior quantidade de tempo para ser obtida diretamente no campo. Dessa forma, assim como Li e Weiskittel (2010), o diâmetro a 5,2 m do solo foi inicialmente obtido por meio de uma interpolação linear e, posteriormente, uma equação de predição foi utilizada para a obtenção dessa variável de forma mais prática e econômica, uma vez que esta variável não é frequentemente coletada durante as rotinas de inventário florestal convencionais (Jiang *et al.*, 2005). A interpolação linear foi realizado com auxílio da função `approxfun()` do pacote `stats` da linguagem R (Becker *et al.*, 1988).

O modelo desenvolvido por Clark *et al.* (1991) (13) para a predição do diâmetro a 5,2 m a partir da medida do DAP e altura total foi utilizado neste estudo.

$$D_{5,2m_i} = D_i \left(\beta_1 + \beta_2 \left(\frac{5,2}{H_i} \right)^2 \right) \quad (13)$$

2.7 Redes Neurais Artificiais

Modelos de redes neurais artificiais foram desenvolvidos para estimar os diâmetros a qualquer altura dos fustes neste estudo. Existem diversos critérios como o número de camadas, algoritmo de aprendizagem, funções de ativação, número de neurônios nas camadas ocultas e o tamanho das bases de dados utilizadas para treinamento, validação e teste para definir qual a melhor arquitetura da rede neural (Sakici; Ozdemir, 2018).

O treinamento das redes neurais funcionam de forma mais eficiente quando os dados utilizados estão normalizados em um pequeno intervalo próximo de zero (Jayalakshmi; Santhakumaran, 2011). A normalização dos dados irá acelerar o processo de treinamento e minimizar vieses. Existem diferentes tipos de normalização de dados, sendo uma delas a normalização min-max, que transforma os dados para uma escala de 0 a 1 ou -1 a 1, considerando os valores mínimos e máximos de cada variável, de modo a preservar a distribuição original dos dados (Lantz, 2019; Valença, 2007). Nesse estudo, a normalização min-max foi aplicada para todos as variáveis de entrada e saída, por meio da equação 14 (Soares *et al.*, 2011). Dessa forma, todas as variáveis passaram para a escala de 0 a 1. Como os dados normalizados foram utilizados no treinamento das redes, as saídas das redes precisam ser desnormalizadas para obtenção dos valores reais de saída. A desnormalização dos valores de saída foi obtida por meio da equação 15.

$$x' = \frac{(x - x_{min}) * (b - a)}{(x_{max} - x_{min})} + a \quad (14)$$

$$x = \frac{(x' - a)(x_{max} - x_{min})}{(b - a)} + x_{min} \quad (15)$$

Em que: x' = valor normalizado; x = valor original; x_{min} = valor mínimo da variável; x_{max} = valor máximo da variável; a = limite inferior do intervalo de normalização, igual a 0; b = limite superior do intervalo de normalização, igual a 1.

O treinamento das redes neurais foi realizado por meio do pacote *neuralnet* (Günther; Fritsch, 2010), onde as RNA consideradas foram do tipo *Multilayer Perceptron* (MLP) (Bishop, 1995) e com uma camada de entrada, uma camada intermediária (camada oculta) e uma camada de saída, sendo o valor predito. O número de neurônios na camada intermediária foi definido a partir da equação (16) (Júnior *et al.* 2021; Sahin *et al.*, 2023).

$$2n^{0,5} + n_2 \leq n_1 \leq 2n + 1 \quad (16)$$

Em que n é o número de variáveis de entrada na RNA, n_1 é o número de neurônios na camada oculta e n_2 é o número de neurônios na camada de saída.

A partir do número da definição da quantidade mínima e máxima do número de neurônios na camada oculta, foram testadas todas as configurações para, no máximo, duas camadas ocultas. Também foram testadas as funções de ativação logística (17) e tangente hiperbólica (18), tanto nas camadas ocultas, quanto na camada de saída. Para a camada de saída, também foi testada a alternativa sem transformação, ou seja, a função linear (19). As funções de ativação avaliadas são demonstradas a seguir:

$$f(s) = \frac{1 - e^{(-2s)}}{1 + e^{(-2s)}} \quad (17)$$

$$f(s) = \frac{1}{1 + e^{(-s)}} \quad (18)$$

$$f(s) = s \quad (19)$$

As variáveis de entrada foram altura do respectivo diâmetro (h_{ij}), diâmetro à altura de 1,3m do solo (D_i), altura total da árvore (H_i), $\frac{H_i - h_{ij}}{H_i - 1,3}$ e variáveis dummy referentes às três espécies (PCC, PCH, PO), resultando em sete variáveis de entrada. Como exemplo, para uma árvore da espécie PCH, com $D_i=15$ e $H_i=25$, onde se deseja estimar o diâmetro a 5m de altura do fuste o vetor de entrada na rede deve ser como a seguir:

$$\begin{bmatrix} h_{ij} \\ D_i \\ H_i \\ \frac{H_i - h_{ij}}{H_i - 1,3} \\ PCC \\ PCH \\ PO \end{bmatrix} = \begin{bmatrix} 5 \\ 15 \\ 25 \\ \frac{25 - 5}{25 - 1,3} \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad (20)$$

O algoritmo utilizado no ajuste dos parâmetros da RNA foi o *Resilient Backpropagation with weigh backtracking*, que é baseado no algoritmo *Backpropagation*, ajustando os pesos da rede neural com o objetivo de minimizar o erro, onde diferentes taxas de aprendizado são utilizadas para cada peso, como é abordado em mais detalhes em Riedmiller (1994).

O critério de parada das redes neurais, durante o treinamento, foi atingido pelo número máximo de 100.000 iterações ou quando todas as derivativas parciais absolutas da função de erro quadrático médio (E) em relação aos pesos (w) ($\partial E/\partial w$) fossem menores que 1%. Uma vez que, uma dessas condições foi atendida, a etapa de treinamento foi concluída e a melhor configuração de parâmetros foi selecionada, baseando-se no menor erro quadrático médio. Além disso, como os pesos iniciais de cada treinamento são valores retirados aleatoriamente de uma distribuição normal padrão, cada estrutura de rede neural foi treinada três vezes e apenas a de menor erro quadrático médio foi considerada como resultado final para posterior comparação com as demais estruturas.

Após o ajuste de todas as possíveis estruturas das redes neurais, todas foram comparadas entre si, onde foi avaliado o desempenho tanto para base de treinamento, quanto para base de teste.

Tabela 3 – Configurações das RNA avaliadas.

Model	Neurons	Layers	Inputs	Output	Activation function		Total combinations
					Hidden layer	Output layer	
ANN	7 to 15	1 to 2	DAP, HT, hi, H_hi_H_1.3, spp_PCH, spp_PCC, spp_PO	di	Logistic, Tanh	Logistic, Tanh, Linear	540

2.8 Avaliação dos modelos

O desempenho dos diferentes tipos de modelos testados (equações de afilamento de efeitos fixos, alternativas de calibração dos modelos mistos e RNA) foi avaliado com base na

raiz do erro quadrático médio (RMSE – Root Mean Squared Error) , erro médio absoluto percentual (MAPE – Mean Absolute Percentage Error) , desvio médio (BIAS) e índice de eficiência do modelo (EF - model efficiency) (21-26), conforme estudado por Raptis *et al.* (2021), Camacho *et al.* (2022) e Ogana e Ercanli (2022).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (d_{obs} - d_{est})^2}{n - p}} \quad (21)$$

$$RMSE_ANN = \sqrt{\frac{\sum_{i=1}^n (d_{obs} - d_{est})^2}{n - k}} \quad (22)$$

$$MAPE (\%) = \frac{\sum_{i=1}^n \frac{|d_{obs} - d_{est}|}{d_{obs}}}{\bar{d}_{obs}} \times 100 \quad (23)$$

$$BIAS = \frac{\sum_{i=1}^n (d_{obs} - d_{est})}{n} \quad (24)$$

$$EF = 1 - \frac{(n - 1) \sum_{i=1}^n (d_{obs} - d_{est})^2}{(n - p) \sum_{i=1}^n (d_{obs} - \bar{d}_{obs})^2} \quad (25)$$

$$EF_ANN = 1 - \frac{(n - 1) \sum_{i=1}^n (d_{obs} - d_{est})^2}{(n - k) \sum_{i=1}^n (d_{obs} - \bar{d}_{obs})^2} \quad (26)$$

Em que: d_{obs} = diâmetro observado; d_{est} = diâmetro estimado; \bar{d}_{obs} = diâmetro médio observado; n = número de observações; p = número de parâmetros.; k = número de pesos.

Para comparação dos modelos em relação à qualidade das previsões, tanto para a base de treino quanto para a de teste, o método de ranking relativo (Poudel; Cao, 2013; Sahin *et al.*, 2023) (27) foi utilizado para a avaliação simultânea das cinco métricas utilizadas.

$$R_i = 1 + \frac{(m - 1) - (S_i - S_{min})}{S_{max} - S_{min}} \quad (27)$$

Onde R_i = valor do *ranking* relativo do i -ésimo modelo avaliado ($i = 1, 2, 3 \dots m$).

S_i valor da métrica estatística do modelo i . S_{min} e S_{max} = valores mínimo e máximo de S_i , respectivamente. m = quantidade de modelos sendo comparados.

Este método permite que a melhor e pior técnicas avaliadas possuam um ranking relativo de 1 e m , respectivamente. O *rank* das demais técnicas avaliadas foi expressado por um valor real entre 1 e m e, uma vez que, além da posição no *ranking*, o tamanho de S_i também é considerado, este método fornece mais informações que o método tradicional de *ranking* que utiliza apenas valores inteiros referentes às posições das métricas. Este método considera como

melhor, a métrica de menor valor, dessa forma, para as métricas em que o melhor resultado é o maior valor, como o EF, os valores foram multiplicados por -1 e, para o BIAS, em que o melhor valor é o mais próximo de zero, foi utilizado o módulo desses valores. Após o ajuste das equações de efeito fixo, a equação com melhor desempenho foi selecionada para ser utilizada como base nas etapas seguintes.

Para comparação dos modelos mistos entre si, foi utilizado o critério de Informação de Akaike (AIC) (28), Critério Bayesiano de Schwarz (BIC) (29), o logaritmo da verossimilhança maximizada (LogLik) e o teste da razão de verossimilhança (MLRT). O AIC é obtido a partir do valor da verossimilhança e na quantidade de parâmetros, além de derivar-se da informação ou distância de Kullback-Leibler, que verifica a diferença entre dois modelos (Akaike, 1973; Floriano *et al.*, 2006), com base nesta métrica, o melhor modelo apresentará o menor valor de AIC. O Critério Bayesiano de Schwarz (BIC) é uma métrica que auxilia na seleção de modelos e, geralmente, é empregada juntamente com o AIC e descreve a associação entre a variável dependente e as variáveis independentes, aumentando as chances de se escolher o modelo mais consistente dentre os avaliados. Avaliando este critério, deve-se optar pelo modelo que apresentar o menor valor de *BIC*.

$$AIC = -2l(\hat{\theta}) + 2p \quad (28)$$

$$BIC = -2l(\hat{\theta}) + 2p \ln(n) \quad (29)$$

Em que: $l(\hat{\theta})$ = logaritmo da verossimilhança maximizada; \ln = logaritmo neperiano; n = número de observações.

O MLRT teve como objetivo avaliar o resultado da inclusão do efeito aleatório nos modelos (Resende *et al.*, 2014) e foi feito verificando a significância da diferença entre os valores de *deviance* ($-2l(\hat{\theta})$) dos modelos com e sem a incorporação da variável aleatória, comparando o valor calculado com o valor tabelado, a partir do teste qui-quadrado (χ^2), a 5% de significância.

Além das medidas globais, foram calculados os valores de BIAS para as porções de 0 a 100% da altura do fuste, variando de 10 em 10%. Essas medidas tiveram como objetivo auxiliar na verificação da estabilidade de cada função ao longo do fuste.

Por fim, para validar a eficácia dos métodos selecionados, testes pareados de média foram aplicados, comparando os valores observados com os valores preditos pelas metodologias propostas, utilizando a base de teste. Uma técnica comum de verificar a significância de resíduos é considerar a hipótese nula de que “não há diferença significativa”

entre os valores preditos e observados contra a hipótese alternativa de que há “diferença significativa” entre esses valores (Sahin *et al.*, 2023).

Assumindo-se que os resíduos possuem uma distribuição normal e ausência de *outliers*, o teste t pareado foi utilizado para aceitar ou rejeitar a hipótese nula. No entanto, o teste t é sensível a *outliers* (Schabenberger; Pierce, 2002), não recomendado para dados com distribuição não normal, portanto, a presença de erros elevados foi verificada considerando como erros aceitáveis aqueles que estavam entre os limites de -3 e 3 resíduos padronizados e a normalidade dos dados foi verificada por meio do teste de Shapiro-Wilk ($p > 0,05$).

Desse modo, quando uma das pressuposições anteriores foi violada, o teste de Wilcoxon *signed rank* pareado foi aplicado, uma vez que é recomendado quando há a existência de *outliers* e a falta de normalidade (Hollander; Wolfe, 1973). Ou seja, aceitando a hipótese nula ($p > 0,05$), seja pelo teste t ou de Wilcoxon *signed rank*, significa que as previsões resultantes dos modelos podem ser consideradas como satisfatórias a um nível de 5% de significância.

2.9 Avaliação da predição do volume total

No presente estudo, os diâmetros foram preditos por meio de duas técnicas (modelos de efeito misto e RNA) enquanto que o volume foi obtido da mesma forma para ambas essas técnicas. Comumente o volume é obtido integrando a área seccional obtida a partir da função de afilamento da seguinte forma: $V_i = k \int_{h_1}^{h_2} d_i^2 \partial h$, onde $k = \pi/40000$ resulta em uma estimativa contínua infinitésima entre duas seções do fuste (entre h_1 e h_2). Porém, neste estudo, os volumes seccionais foram obtidos utilizando os diâmetros preditos pelos modelos propostos, nas mesmas seções onde os dados da cubagem rigorosa foram obtidos e aplicando-se o método de Smalian. Posteriormente, os valores preditos de volume total foram avaliados com os mesmos critérios utilizados para as previsões de diâmetro.

2.10 Avaliação de predição dos métodos

Para visualizar a capacidade preditiva do modelo misto e da RNA selecionados, três árvores de tamanhos pequeno, médio e grande, por espécie, foram selecionadas da base de teste e os modelos foram aplicados a essas árvores para verificar o quão eficiente foram as previsões. Desse modo, foram gerados gráficos de dispersão das medidas reais de diâmetros das seções e suas respectivas alturas e as curvas correspondentes às previsões dos modelos foi simulada de centímetro a centímetro para verificar o comportamento das curvas em relação aos valores reais.

3 RESULTADOS

3.1 Equações de afilamento de efeitos fixos

As oito equações de afilamento do fuste foram ajustadas para cada espécie pelos métodos OLS e ONLS a partir da base de treino, enquanto a capacidade preditiva dos modelos foi verificada a partir da base de teste. Inicialmente, todos os modelos avaliados foram ajustados assumindo-se variância homogênea dos erros e resíduos sem correlação. As estatísticas são apresentadas na Tabela 4, onde o *ranking* relativo permitiu avaliar o desempenho geral das equações ajustadas.

Tabela 4 – Desempenho estatístico das equações de afilamento por espécie.

Species	Model	TRAIN				TEST				Summed ranks	Overall rank
		RMSE	MAPE	EF	BIAS	RMSE	MAPE	EF	BIAS		
PCC	Kozak	1,06	5,11	0,9832	0,0172	1,08	5,1	0,9808	0,0104	8,00	1,00
	Shoepfer	1,16	5,58	0,9799	-0,0421	1,15	5,54	0,9779	-0,0595	13,05	1,66
	Modified Ormerod	1,16	5,6	0,9799	0,1099	1,13	5,4	0,9788	0,0743	13,71	1,75
	Hradetzy	1,1	5,17	0,982	0,2117	1,11	5,14	0,9794	0,1958	14,65	1,87
	Max and Burkhart	1,17	5,62	0,9795	-0,0889	1,16	5,59	0,9778	-0,1085	14,78	1,89
	Parresol	1,36	6,75	0,9723	0,3318	1,37	6,81	0,969	0,3142	30,89	4,00
	Demaerschalk	1,78	9,06	0,9529	0,0552	1,88	9,38	0,9413	0,0429	46,08	6,00
	Ormerod	1,92	8,29	0,9451	0,5168	2,02	8,59	0,9323	0,4881	61,33	8,00
PCH	Kozak	1,18	5,61	0,9812	0,0166	1,15	5,48	0,9817	0,1817	9,39	1,00
	Hradetzy	1,25	5,84	0,9787	-0,2097	1,21	5,63	0,9797	-0,0541	13,13	1,50
	Shoepfer	1,28	6,11	0,9779	-0,0218	1,27	6,1	0,9778	0,1338	13,89	1,60
	Modified Ormerod	1,24	5,79	0,9792	0,0788	1,26	5,88	0,9782	0,253	14,34	1,66
	Max and Burkhart	1,29	6,19	0,9776	-0,0786	1,27	6,1	0,9778	0,0841	14,36	1,66
	Parresol	1,54	7,26	0,9681	0,3775	1,59	7,58	0,965	0,5291	36,25	4,58
	Demaerschalk	1,96	9,49	0,948	0,0611	1,93	9,36	0,9483	0,2244	48,30	6,19
	Ormerod	2,07	8,82	0,9421	0,5329	2,06	8,85	0,9411	0,6958	61,88	8,00
PO	Kozak	1,17	5,22	0,9826	0,0234	1,19	4,83	0,9826	0,0908	9,33	1,00
	Hradetzy	1,26	5,5	0,9796	-0,0231	1,25	5,24	0,9808	0,0428	12,17	1,37
	Shoepfer	1,31	5,85	0,9782	-0,0231	1,31	5,54	0,9792	0,0428	14,71	1,71
	Max and Burkhart	1,32	5,92	0,9779	-0,0754	1,31	5,63	0,9791	-0,0072	15,58	1,82
	Modified Ormerod	1,3	5,77	0,9785	0,0063	1,35	5,72	0,9779	0,0998	15,73	1,84
	Parresol	1,6	7,47	0,9674	0,3836	1,64	6,93	0,9671	0,4716	40,77	5,14
	Demaerschalk	1,93	8,84	0,9522	0,0446	2,04	8,8	0,9494	0,1269	48,96	6,21
	Ormerod	2,01	8,4	0,9483	0,4482	2,14	8,45	0,944	0,5572	62,53	8,00

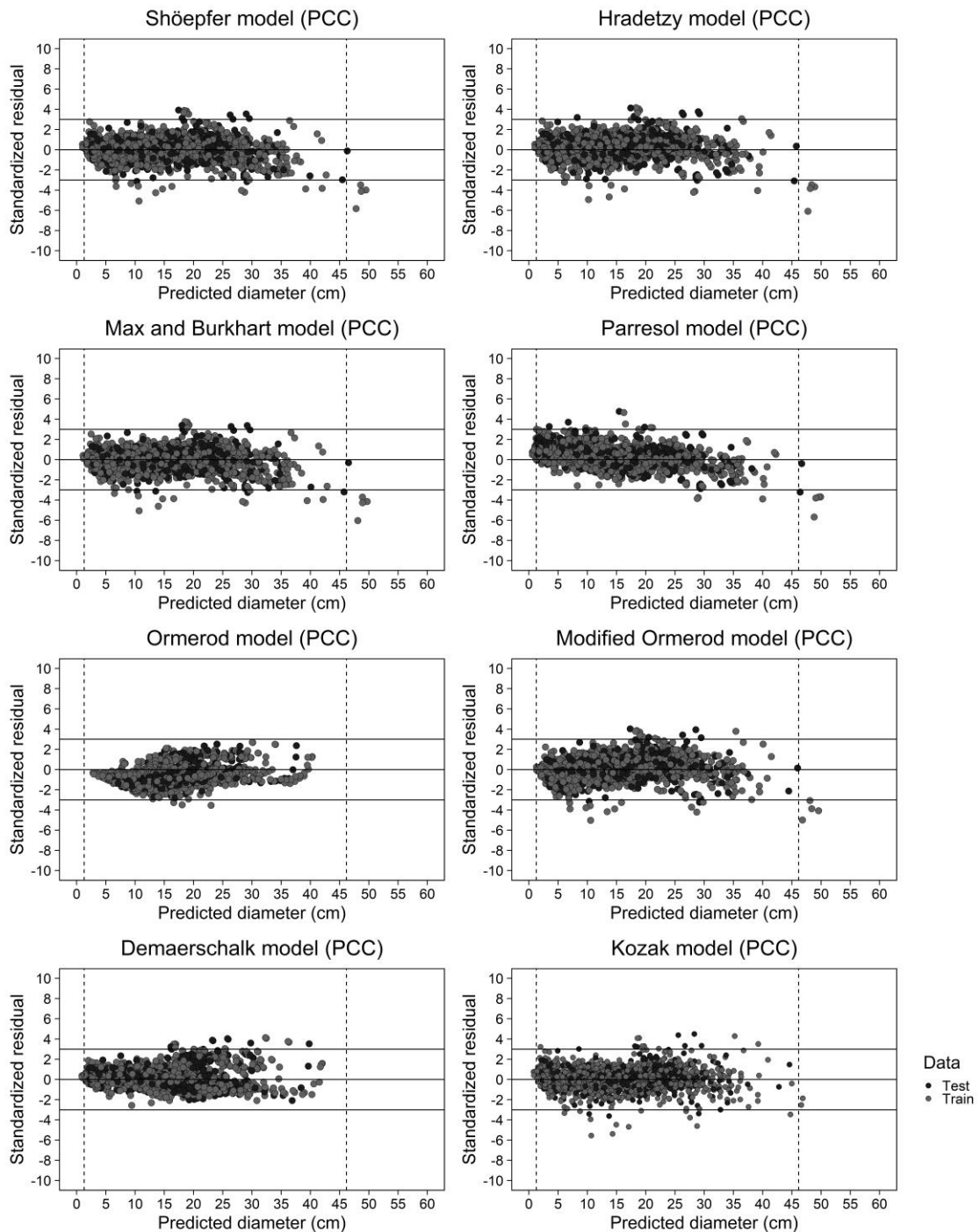
De modo geral, as equações ajustadas no presente estudo foram capazes de explicar a variação dos diâmetros ao longo do fuste com uma eficiência média de cerca de 97% para a base de treino e teste para todas as espécies. Além disso, os modelos na estrutura fixa, quando avaliados para a base de treino, resultaram em erros médios absolutos com valores variando de 5,11 a 9,06%, 5,61 a 9,49% e 5,22 a 8,84% para as espécies PCC, PCH e PO, respectivamente. Semelhantemente, para a base de teste, os valores variaram de 5,10 a 9,38%, 5,48 a 9,36% e 4,83 a 8,80% para as espécies PCC, PCH e PO, respectivamente. Considerando os valores de RMSE para a base de treino, houve uma variação de 1,06 a 1,92 cm para PCC, 1,18 a 2,07 cm para PCH e 1,17 a 2,01 cm para PO, enquanto que para a base de teste a variação foi de 1,08 a 2,02 cm, 1,15 a 2,06 cm e 1,19 a 2,14 cm. Ainda, os valores absolutos de BIAS se distanciaram na ordem 0,02 a 0,52 cm para PCC, 0,02 a 0,53 cm e 0,01 a 0,45 cm para PO para a base de treino, enquanto que, para a base de teste os valores variaram de 0,01 a 0,49 cm, 0,05 a 0,70 e 0,01 a 0,56 cm.

Dentre os oito modelos avaliados para cada espécie, o modelo de Kozak (2004) foi o que resultou no melhor desempenho para a base de treino, com os menores valores de RMSE (PCC = 1,06 cm, PCH = 1,18 cm e PO = 1,17 cm) e MAPE (PCC = 5,11 %, PCH = 5,61 % e PO = 5,22 %) para todas as espécies, enquanto que, para o BIAS, os melhores resultados foram alcançados para as espécies PCC e PCH (0,0172 cm e 0,0166 cm, respectivamente) e para PO o modelo Modified Ormerod (1992) resultou no melhor resultado (0,0063 cm). Ainda considerando a base de treino, o modelo Kozak (2004) resultou nos maiores valores de EF para todas as espécies com uma eficiência superior a 98 %. Considerando a base de teste, o modelo Kozak (2004) resultou em desempenho superior para RMSE (PCC = 1,08 cm, PCH = 1,15 cm e PO = 1,19 cm), MAPE (PCC = 5,10 %, PCH = 5,48 % e PO = 4,83 %) e EF (PCC = 0,9808, PCH = 0,9817 cm e PO = 0,9826) para todas as espécies e para o BIAS o desempenho foi superior apenas para a espécie PCH (0,0104 cm), enquanto que os modelos Hradetzky (1976) e Max e Burkhart (1976) resultaram nos melhores resultados para PCC (-0,0541) e PO (-0,0072), respectivamente.

De acordo com os valores do *ranking* geral, o modelo Ormerod (1973) foi o menos adequado para todas as espécies, enquanto que, por outro lado, o modelo Kozak (2004) foi o que demonstrou melhor desempenho para todas as espécies. Estes resultados podem ser confirmados, quando se analisa a distribuição dos resíduos padronizados das equações ajustadas, em que, o modelo Ormerod (1973) é o que possui as distribuições menos simétricas e as predições mais distantes das amplitudes observadas para todas as espécies (Figuras 4-6), concordando com seus maiores valores de BIAS, indicando uma tendência à subestimação das

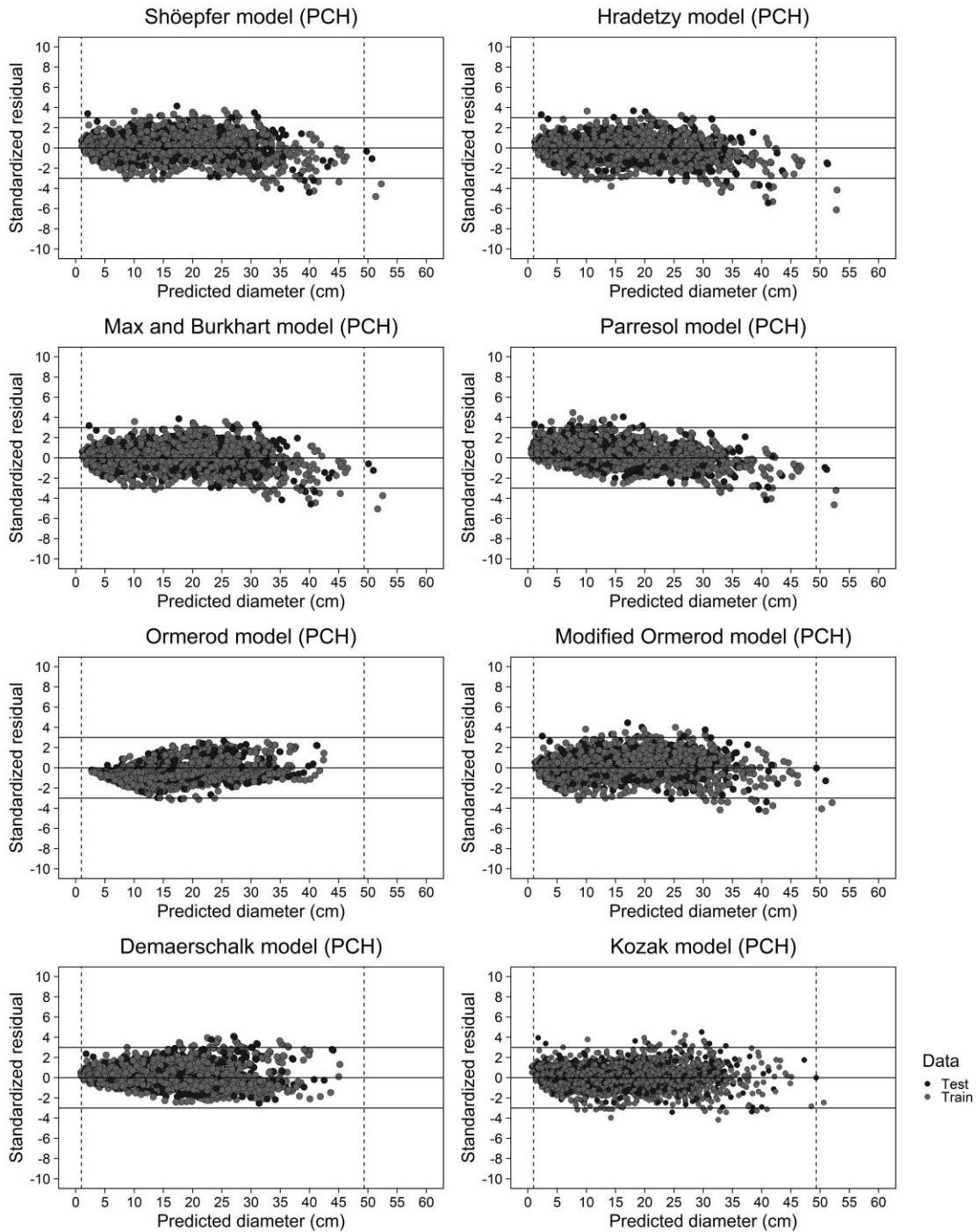
predições para ambas as bases de treino e teste. Dessa forma, por possuir os melhores resultados de *rank* e as distribuições de resíduos mais simétrica que os demais modelos, o modelo de forma variável de Kozak (2004) foi selecionado como o modelo base para a modelagem mista das três espécies de *Pinus* avaliadas.

Figura 4 – Distribuição residual para as predições diâmetro pelas oito equações de afilamento ajustadas para a espécie *Pinus caribaea* var. *caribaea* (PCC).



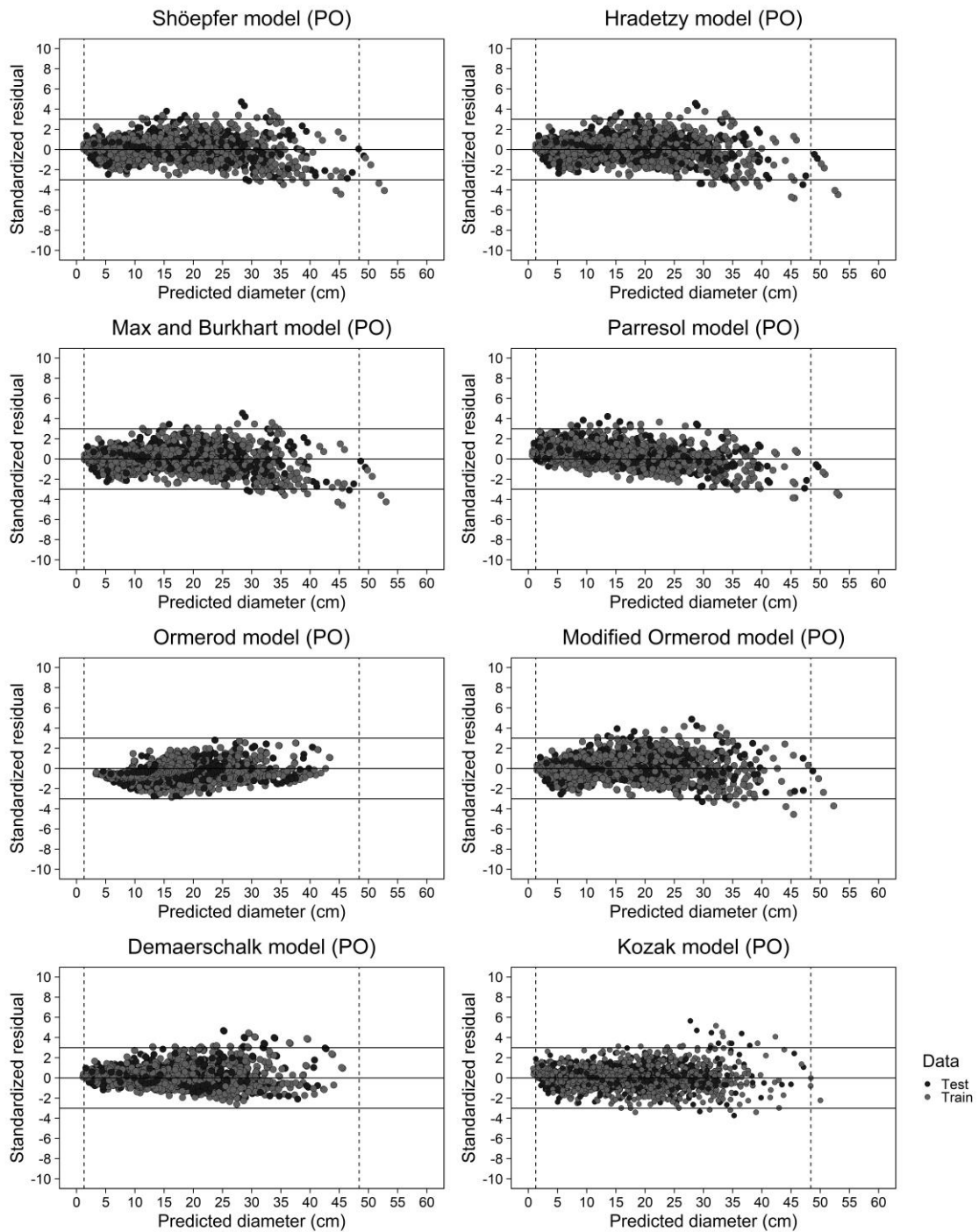
Fonte: Do autor (2024).

Figura 5 – Distribuição residual para as previsões diâmetro pelas oito equações de afilamento ajustadas para a espécie *Pinus caribaea* var. *hondurensis* (PCH).



Fonte: Do autor (2024).

Figura 6 – Distribuição residual para as previsões diâmetro pelas oito equações de afilamento ajustadas para a espécie *Pinus oocarpa* (PO).



Fonte: Do autor (2024).

Após a seleção do modelo Kozak (2004), seus parâmetros foram avaliados de acordo sua significância a um nível de confiança de 95%, como pode ser observado na Tabela 4. Para todas as espécies, dois parâmetros não foram significativos, sendo β_4 e β_5 para PCC e PO e β_2 e β_5 para PCH, de modo que esses parâmetros foram removidos dos modelos um a um para verificar a influência da remoção. Após a verificação, percebeu-se que a remoção dos parâmetros não significativos não resultou em impacto relevante na qualidade das predições

dos modelos nem no comportamento do BIAS ao longo das alturas relativas dos fustes. Dessa forma, novos modelos Kozak (2004) modificados foram ajustados sem a presença dos parâmetros não significativos (Tabela 5).

Tabela 5 – Estatísticas paramétricas para o modelo Kozak (2004) e suas modificações com a remoção dos parâmetros não significativos.

Species	Parameter estimates	Kozak (2004)				Modified Kozak (2004)			
		Values	Standard error	t value	p value	Values	Standard error	t value	p value
PCC	α_0	0,7252	0,0336	21,5813	0,0000	0,7067	0,0282	25,0533	0,0000
	α_1	0,9214	0,0095	97,0969	0,0000	0,9212	0,0094	98,1980	0,0000
	α_2	0,1924	0,0206	9,3511	0,0000	0,2018	0,0185	10,9301	0,0000
	β_1	0,4060	0,0191	21,2491	0,0000	0,4216	0,0147	28,7760	0,0000
	β_2	0,1522	0,0748	2,0331	0,0422	0,1940	0,0381	5,0960	0,0000
	β_3	0,2692	0,0180	14,9959	0,0000	0,2650	0,0173	15,3443	0,0000
	β_4	0,3353	0,4381	0,7654	0,4442	-	-	-	-
	β_5	0,0069	0,0055	1,2555	0,2095	-	-	-	-
	β_6	0,1665	0,0466	3,5712	0,0004	0,2229	0,0138	16,1603	0,0000
PCH	α_0	0,8508	0,0328	25,9136	0,0000	0,8375	0,0280	29,8999	0,0000
	α_1	0,9693	0,0081	119,1984	0,0000	0,9737	0,0077	126,1300	0,0000
	α_2	0,0891	0,0171	5,2000	0,0000	0,0907	0,0150	6,0598	0,0000
	β_1	0,3705	0,0163	22,7679	0,0000	0,3909	0,0126	31,1459	0,0000
	β_2	0,0581	0,0648	0,8968	0,3699	-	-	-	-
	β_3	0,2673	0,0149	17,9920	0,0000	0,2781	0,0121	23,0198	0,0000
	β_4	1,4458	0,4012	3,6034	0,0003	1,5495	0,1894	8,1818	0,0000
	β_5	0,0088	0,0047	1,8682	0,0619	-	-	-	-
	β_6	0,1555	0,0401	3,8798	0,0001	0,2233	0,0120	18,5776	0,0000
PO	α_0	0,6872	0,0280	24,5293	0,0000	0,6707	0,0232	28,8629	0,0000
	α_1	0,9129	0,0100	90,9309	0,0000	0,9093	0,0100	91,3529	0,0000
	α_2	0,2174	0,0194	11,1878	0,0000	0,2288	0,0175	13,0752	0,0000
	β_1	0,4089	0,0186	21,9700	0,0000	0,4048	0,0146	27,8179	0,0000
	β_2	0,2411	0,0793	3,0420	0,0024	0,3865	0,0404	9,5752	0,0000
	β_3	0,2025	0,0169	11,9482	0,0000	0,1956	0,0166	11,7872	0,0000
	β_4	0,9295	0,4797	1,9376	0,0529	-	-	-	-
	β_5	-0,0028	0,0053	-0,5222	0,6016	-	-	-	-
	β_6	0,2503	0,0447	5,5986	0,0000	0,2337	0,0139	16,8333	0,0000

3.2 Modelos de efeito misto

A partir dos modelos Kozak (2004) modificados, que passaram a ser compostos por sete parâmetros de efeito fixo, foram adicionados os parâmetros de efeito aleatório a nível de árvore. No total 28 combinações foram possível a partir de diferentes combinações de um e dois parâmetros de efeito aleatório. Das 28 combinações, 21 convergiram para a espécie PCC, 16 para PCH e e 23 para PO (Tabela 6). De posse dessas combinações, por meio do teste MLRT, foi verificado se a adição dos parâmetros de efeito aleatório resultou em uma melhora significativa dos modelos e então foram selecionadas as melhores estruturas mistas por espécies.

Tabela 6 – Resultados do teste MLRT para o modelo Kozak (2004) modificado para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).

Model	Random Parameters	LogLik	AIC	Test	MLRT	p-value
PCC						
1	a1	-2062,58	4143,16	1 vs 23	215,93	0,0000
2	a2	-2061,70	4141,40	2 vs 23	217,69	0,0000
3	b1	-2054,11	4126,22	3 vs 23	232,87	0,0000
4	b2	-1978,97	3975,94	4 vs 23	383,14	0,0000
5	b3	-1979,66	3977,32	5 vs 23	381,77	0,0000
6	b6	-2043,50	4105,00	6 vs 23	254,08	0,0000
7	a0, a2	-2061,70	4145,40	7 vs 23	217,69	0,0000
8	a1, a2	-2060,42	4142,84	8 vs 23	220,25	0,0000
9	a1, b1	-1908,08	3838,17	9 vs 23	524,92	0,0000
10	a1, b2	-1807,65	3637,29	10 vs 23	725,80	0,0000
11	a1, b3	-1794,80	3611,61	11 vs 23	751,48	0,0000
12	a1, b6	-1845,12	3712,24	12 vs 23	650,85	0,0000
13	a2, b1	-1906,88	3835,76	13 vs 23	527,32	0,0000
14	a2, b2	-1806,56	3635,12	14 vs 23	727,96	0,0000
15	a2, b3*	-1793,69	3609,38	15 vs 23	753,71	0,0000
16	a2, b6	-1844,09	3710,18	16 vs 23	652,91	0,0000
17	b1, b2	-1911,79	3845,59	17 vs 23	517,50	0,0000
18	b1, b3	-1911,57	3845,14	18 vs 23	517,95	0,0000
19	b1, b6	-1920,79	3863,57	19 vs 23	499,51	0,0000
20	b2, b3	-1906,74	3835,47	20 vs 23	527,61	0,0000
21	b2, b6	-1810,96	3643,91	21 vs 23	719,17	0,0000
22	b3, b6	-1807,68	3637,36	22 vs 23	725,73	0,0000
23	None	-2170,54	4357,09	-	-	-
PCH						
1	a1	-3283,73	6585,45	1 vs 18	372,17	0,0000
2	a2	-3282,71	6583,43	2 vs 18	374,20	0,0000
3	b1	-3383,60	6785,19	3 vs 18	172,43	0,0000

4	b3	-3181,87	6381,75	4 vs 18	575,88	0,0000
5	b6	-3230,80	6479,59	5 vs 18	478,03	0,0000
6	a0, b1	-3383,60	6789,19	6 vs 18	172,43	0,0000
7	a0, b3	-3181,87	6385,75	7 vs 18	575,88	0,0000
8	a0, b6	-3230,80	6483,59	8 vs 18	478,03	0,0000
9	a1, b1	-3163,96	6349,91	9 vs 18	611,71	0,0000
10	a1, b3	-2846,44	5714,88	10 vs 18	1246,74	0,0000
11	a1, b6	-2877,52	5777,04	11 vs 18	1184,58	0,0000
12	a2, b1	-3162,45	6346,89	12 vs 18	614,73	0,0000
13	a2, b3*	-2844,61	5711,23	13 vs 18	1250,40	0,0000
14	a2, b6	-2876,12	5774,25	14 vs 18	1187,38	0,0000
15	b1, b3	-3110,74	6243,49	15 vs 18	718,14	0,0000
16	b1, b6	-3105,61	6233,22	16 vs 18	728,40	0,0000
17	b3, b6	-2921,84	5865,68	17 vs 18	1095,94	0,0000
18	None	-3469,81	6955,63	-	-	-
PO						
1	a1	-2197,60	4413,19	1 vs 23	274,90	0,0000
2	a2	-2195,88	4409,76	2 vs 23	278,33	0,0000
3	b1	-2307,83	4633,67	3 vs 23	54,42	0,0000
4	b2	-2206,05	4430,10	4 vs 23	257,99	0,0000
5	b3	-2193,12	4404,25	5 vs 23	283,84	0,0000
6	b6	-2137,79	4293,58	6 vs 23	394,52	0,0000
7	a0+b1	-2307,84	4637,67	7 vs 23	54,42	0,0000
8	a1+a2	-2191,97	4405,94	8 vs 23	286,16	0,0000
9	a1+b1	-2131,51	4285,02	9 vs 23	407,08	0,0000
10	a1+b2	-1934,18	3890,35	10 vs 23	801,74	0,0000
11	a1+b3	-1901,40	3824,79	11 vs 23	867,30	0,0000
12	a1+b6	-1843,01	3708,02	12 vs 23	984,07	0,0000
13	a2+b1	-2129,55	4281,11	13 vs 23	410,98	0,0000
14	a2+b2	-1932,07	3886,13	14 vs 23	805,96	0,0000
15	a2+b3	-1899,41	3820,83	15 vs 23	871,26	0,0000
16	a2+b6*	-1841,14	3704,29	16 vs 23	987,80	0,0000
17	b1+b2	-2145,94	4313,87	17 vs 23	378,22	0,0000
18	b1+b3	-2139,82	4301,63	18 vs 23	390,46	0,0000
19	b1+b6	-2106,62	4235,23	19 vs 23	456,86	0,0000
20	b2+b3	-2160,54	4343,08	20 vs 23	349,01	0,0000
21	b2+b6	-2047,13	4116,25	21 vs 23	575,84	0,0000
22	b3+b6	-2041,49	4104,98	22 vs 23	587,11	0,0000
23	None	-2335,05	4686,09	-	-	-

Onde os prâmetros seguidos por * indicam as estruturas de melhor desempenho de acordo o Critério de Informação de Akaike (AIC).

A partir da tabela 6 também é possível observar que para todas as espécies, a expansão de dois parâmetros obteve uma performance superior à expansão de apenas um parâmetro com efeito aleatório. Para as espécies PCC e PCH, as melhores estruturas mistas foram obtidas a partir das expansões dos parâmetros α_2 e β_3 e, para PO, dos parâmetros α_2 e β_6 , com valores de

MLRT de 753,71, 1250,40 e 987,80, respectivamente, quando comparados com os modelos apenas na forma fixa. Os novos parâmetros estimados para os modelos na forma mista podem ser verificados na tabela 7.

Tabela 7 – Estatísticas paramétricas para o modelo Kozak (2004) modificado na forma mista sem e com a modelagem da estrutura de autocorrelação (CAR(1)) para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).

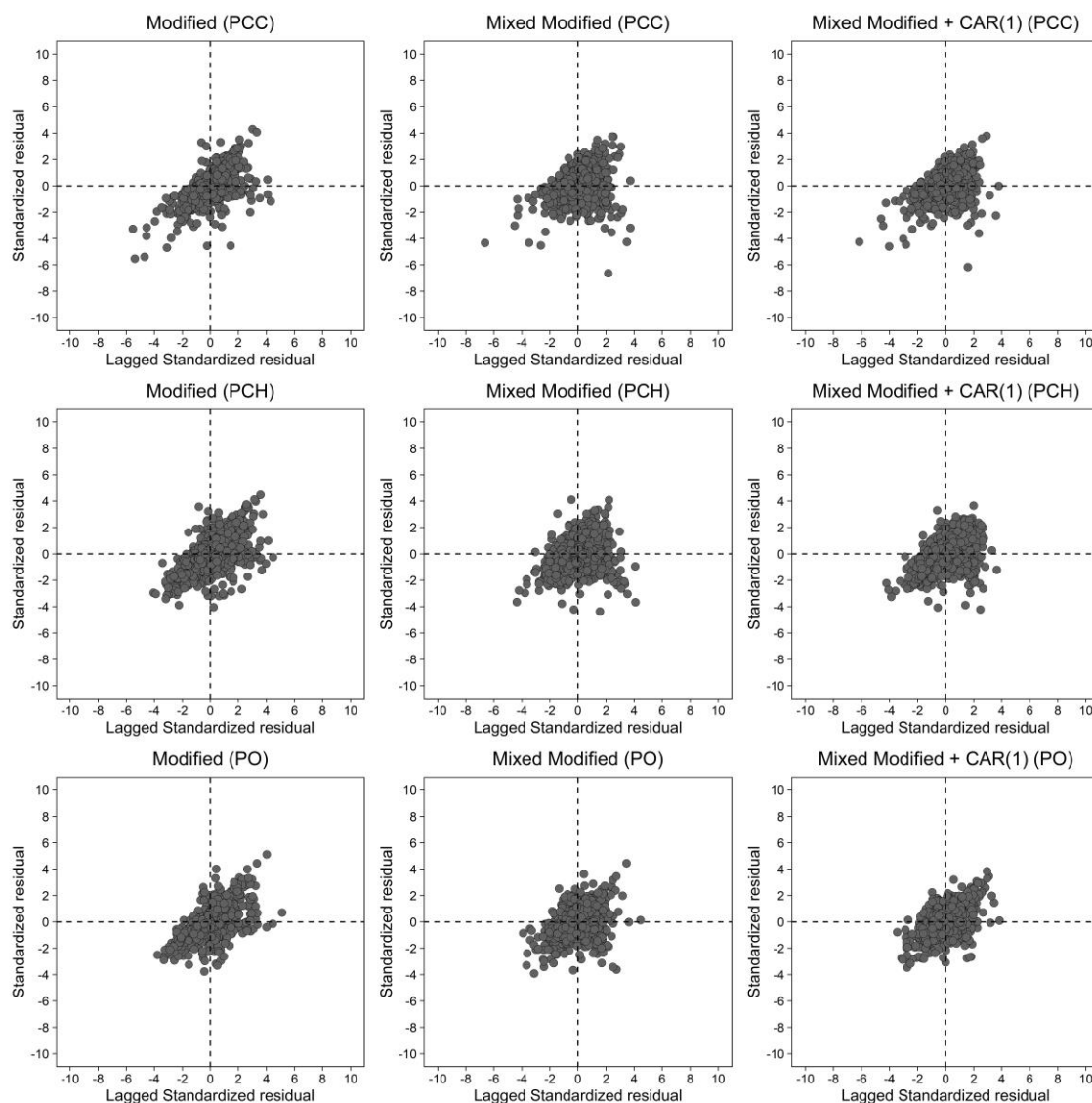
Species	Parameter estimates	Mixed Modified Kozak (2004)				Mixed Modified Kozak (2004) + CAR(1)			
		Values	Standard error	t value	p value	Values	Standard error	t value	p value
PCC	α_0	0,6866	0,0514	13,3635	0,0000	0,6656	0,0585	11,3759	0,0000
	α_1	0,9400	0,0209	45,0492	0,0000	0,9197	0,0231	39,8453	0,0000
	α_2	0,1927	0,0393	4,9044	0,0000	0,2220	0,0435	5,0988	0,0000
	β_1	0,4655	0,0121	38,3513	0,0000	0,4403	0,0157	28,0301	0,0000
	β_2	-0,4013	0,1025	-3,9137	0,0001	0,0615	0,1008	0,6107	0,5415
	β_3	0,5191	0,0443	11,7098	0,0000	0,3099	0,0427	7,2585	0,0000
	β_6	0,1918	0,0103	18,6640	0,0000	0,2503	0,0139	18,0658	0,0000
	σ_{2a2}	0,0001	-	-	-	0,0001	-	-	-
	σ_{2b3}	0,0070	-	-	-	0,0040	-	-	-
	σ_{a2b3}^2	0,2250	-	-	-	0,5310	-	-	-
Correlation Structure (φ)		-	-	-	-	0,6472	-	-	-
PCH	α_1	0,7732	0,0466	16,5784	0,0000	0,9710	0,0190	51,2023	0,0000
	α_2	0,9697	0,0176	55,1940	0,0000	0,1231	0,0346	3,5612	0,0004
	β_1	0,1216	0,0308	3,9412	0,0001	0,4112	0,0121	34,0251	0,0000
	β_3	0,4099	0,0085	47,9487	0,0000	0,3120	0,0264	11,8337	0,0000
	β_4	0,3571	0,0218	16,3465	0,0000	0,7855	0,4413	1,7801	0,0752
	β_6	0,1784	0,3512	0,5079	0,6116	0,2480	0,0112	22,0567	0,0000
	σ_{2a2}	0,2164	0,0080	27,0414	0,0000	0,0001	-	-	-
	σ_{2b3}	0,0002	-	-	-	0,0058	-	-	-
	σ_{a2b3}^2	0,0058	-	-	-	0,5910	-	-	-
	Correlation Structure (φ)		-	-	-	-	0,6567	-	-
PO	α_0	0,6512	0,0451	14,4486	0,0000	0,6528	0,0495	13,1971	0,0000
	α_1	0,9402	0,0210	44,8219	0,0000	0,9492	0,0218	43,5335	0,0000
	α_2	0,2060	0,0366	5,6233	0,0000	0,1948	0,0382	5,0995	0,0000
	β_1	0,4119	0,0089	46,0498	0,0000	0,4131	0,0128	32,1647	0,0000
	β_2	0,1066	0,0362	2,9410	0,0033	0,1550	0,0673	2,3037	0,0214
	β_3	0,2981	0,0137	21,7000	0,0000	0,2749	0,0245	11,2238	0,0000
	β_6	0,2338	0,0151	15,4962	0,0000	0,2503	0,0168	14,8663	0,0000
	σ_{2a2}	0,0002	-	-	-	0,0001	-	-	-
	σ_{2b6}	0,0154	-	-	-	0,0142	-	-	-
	σ_{a2b6}^2	0,2550	-	-	-	0,5210	-	-	-
Correlation Structure (φ)		-	-	-	-	0,6490	-	-	-

Após a expansão dos parâmetros dos modelos modificados para a forma mista, foi adicionada a estrutura de autocorrelação CAR(1) e seus parâmetros também podem ser observados na tabela 7. A inclusão da estrutura de autocorrelação aos modelos mistos se mostrou significativa para todas as espécies de acordo com o teste MLRT (PCC = 381,17, PCH = 615,91 e PO = 457,45) (tabela 8), indicando que, mesmo após a inclusão dos efeitos aleatórios, um certo grau de autocorrelação dos resíduos permaneceu presente. Tal fato também pode ser verificado na Figura 7, por meio da análise gráfica da autocorrelação dos modelos modificados na forma fixa, mista e mista com adição da estrutura de autocorrelação, onde há uma redução da autocorrelação dos resíduos, evidenciado pela centralização da dispersão dos resíduos defasados no centro do gráfico após a adição dos efeitos aleatórios e da estrutura de autocorrelação.

Tabela 8 – Resultados do teste MLRT para o modelo Kozak (2004) modificado na forma mista sem e com a modelagem da estrutura de autocorrelação (CAR(1)) para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).

Species	Model	logLik	BIC	AIC	Test	MLRT	p value
PCC	Mixed Modified Kozak	-1793,69	3667,60	3609,38	-	-	-
	Mixed Modified Kozak + CAR(1)	-1603,10	3293,73	3230,21	1 vs 2	381,17	0,0000
PCH	Mixed Modified Kozak	-2844,61	5773,84	5711,23	-	-	-
	Mixed Modified Kozak + CAR(1)	-2536,66	5165,62	5097,32	1 vs 2	615,91	0,0000
PO	Mixed Modified Kozak	-1841,14	3762,62	3704,29	-	-	-
	Mixed Modified Kozak + CAR(1)	-1612,42	3312,48	3248,84	1 vs 2	457,45	0,0000

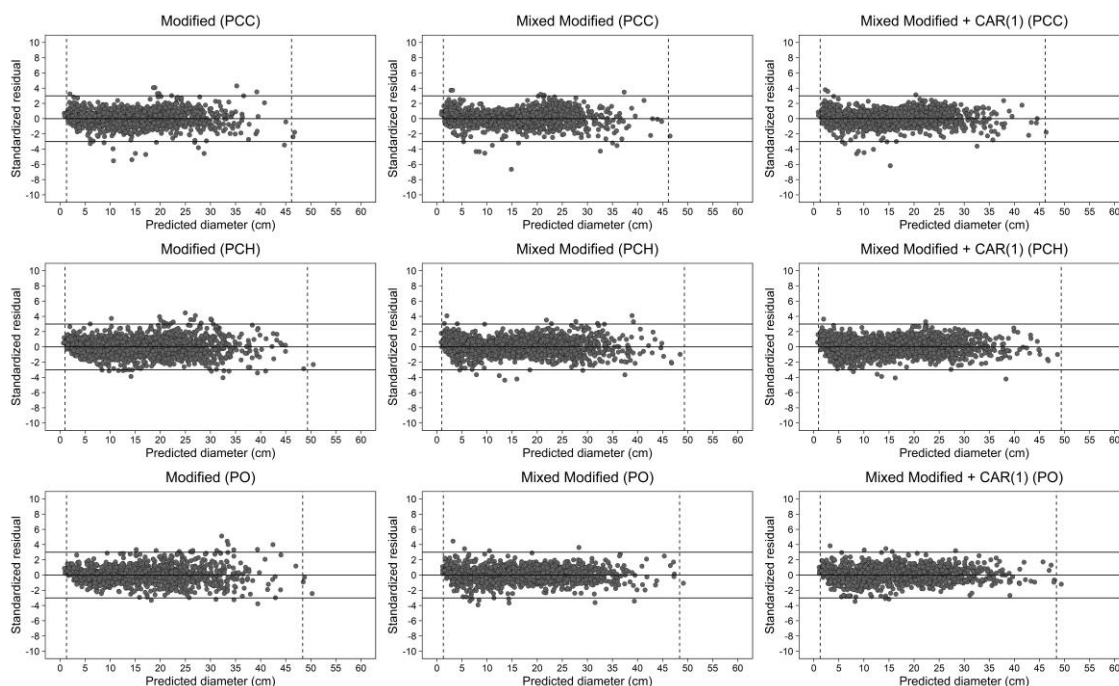
Figura 7 – Distribuição dos resíduos padronizados vs resíduos defasados para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).



Fonte: Do autor (2024).

Analisando a dispersão dos resíduos padronizados em função dos diâmetros preditos (Figura 8), também é possível observar que houve uma maior centralização dos resíduos em torno de zero, conforme os efeitos aleatórios e a estrutura de autocorrelação foram adicionados para todas as espécies.

Figura 8 – Distribuição dos resíduos padronizados em função dos diâmetros preditos para a base de treino para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).



Fonte: Do autor (2024).

Após a modelagem dos efeitos aleatórios e da estrutura de autocorrelação, as distribuições dos resíduos para as três espécies demonstram um comportamento desejável, uma vez que não há evidências de heterocedasticidades dos resíduos, todos estão simetricamente entorno de zero ao longo dos valores preditos, apenas poucos valores ultrapassaram os valores de -3 e 3 resíduos padronizados e os valores preditos são compatíveis com a amplitude dos valores observados, indicando a ausência de alguma tendência ao longo da amplitude dos dados utilizados no treinamento.

Posteriormente a obtenção dos modelos mistos com estrutura de autocorrelação, estes foram calibrados para posterior avaliação da capacidade preditiva. No entanto, dentre as três alternativas de calibração sugeridas neste estudo, duas utilizaram o diâmetro a 5,2 m de altura do solo ($D_{5,2}$) como variável. Desse modo, após a obtenção de $D_{5,2}$ por meio da interpolação linear, ajustou-se equações não lineares para descrever a sua relação com o DAP e altura total (Clark *et al.*, 1991). Os parâmetros estimados e as métricas estatísticas das equações de predição de $D_{5,2}$ podem ser verificados na tabela 9.

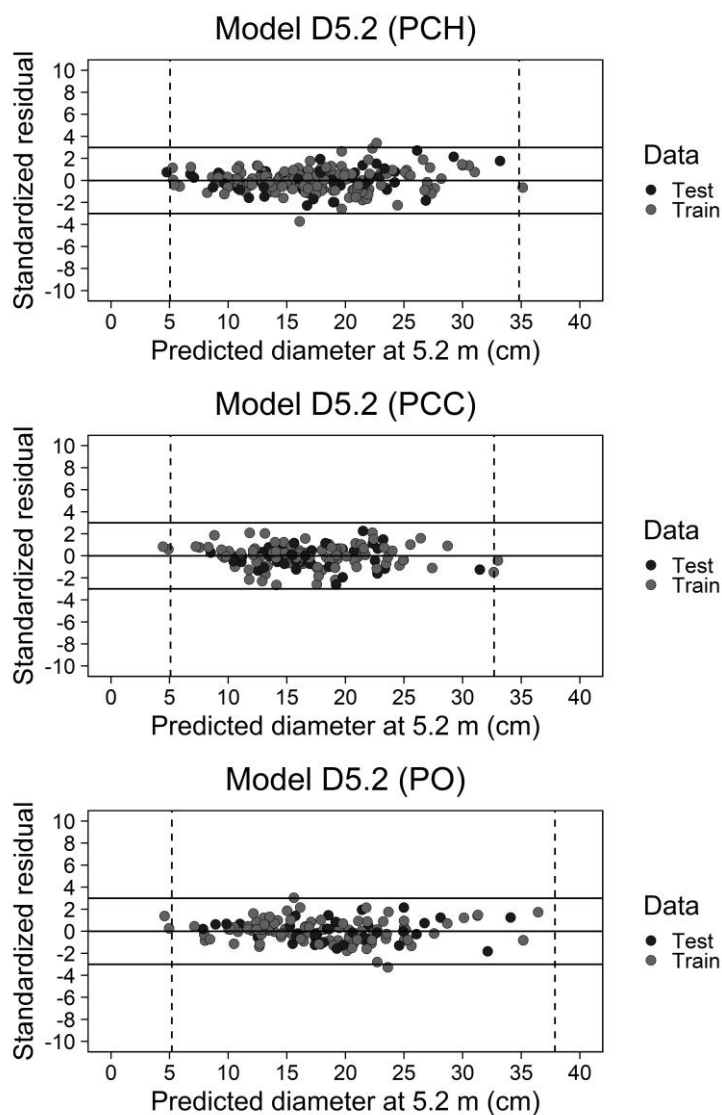
Tabela 9 – Resumo estatístico do desempenho dos ajustes dos modelos de predição de $D_{5,2}$ para as bases de treino e teste para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).

Parameter estimates	Values	Standard error	t value	p value	TRAIN				TEST				
					RMSE	MAPE	EF	BIAS	RMSE	MAPE	EF	BIAS	
PCC													
β_1	0,9349	0,0033	283,7977	0,0000	0,7745	3,57	0,9804	0,0177	0,7599	3,60	0,9717	-0,1193	
β_2	-1,7212	0,0498	-34,5783	0,0000									
PCH													
β_1	0,9079	0,0028	327,7430	0,0000	0,9176	4,02	0,9755	-0,0213	0,9185	4,07	0,9762	0,0887	
β_2	-1,4462	0,0414	-34,9030	0,0000									
PO													
β_1	0,9225	0,0029	315,7103	0,0000	0,8551	3,67	0,9810	0,0136	0,7877	3,40	0,9817	0,0347	
β_2	-1,7058	0,0422	-40,4694	0,0000									

Onde foram considerados significantes os parâmetros com valor $p \leq 0,05$.

Todos os parâmetros das equações de predição de $D_{5,2}$ foram significativos a 95% de nível de significância. Além disso, as equações resultaram em valores médios de 0,8490 e 0,8221 cm para RMSE, 3,76 e 3,69% para MAPE e cerca de 98% de eficiência para base de treino e teste, respectivamente, para as três espécies analisadas. Essas informações apontam uma grande capacidade das variáveis DAP e altura total em explicar $D_{5,2}$, juntamente com os gráficos de resíduos padronizados, em que é possível observar ausência de viés ao longo dos valores preditos, bem como apontam a presença de variâncias homogêneas para as bases de treino e teste. Dessa forma, essas equações foram utilizadas para obtenção dos valores de $D_{5,2}$ para a base de teste para, posteriormente, serem utilizados na calibração dos modelos mistos selecionados.

Figura 9 – Distribuição dos resíduos padronizados em função dos diâmetros a 5,2 m preditos para a base de treino e teste para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).



Fonte: Do autor (2024).

As métricas de desempenho das predições de diâmetro ao longo do fuste e volume total para as diferentes alternativas de diâmetros testados no processo de calibração são mostradas na tabela 10. Inicialmente é possível perceber que, independentemente da espécie ou opção de calibração, as predições de diâmetro possuem maior precisão em relação às de volume total. De acordo o *rank* geral, para a espécie PCC, a ausência de calibração resultou no melhor desempenho geral com valores superiores de RMSE (1,07 cm), MAPE (5,10 %) e EF (0,9810) para as predições de diâmetro, enquanto que o melhor valor de BIAS (-0,0214) foi alcançado por meio da calibração utilizando as medidas de diâmetro a 1,3 e 5,2 m de altura. Já para as predições de volume total, a opção sem calibração gerou resultados superiores apenas para RMSE (0,0272 m³) e EF (0,9826), enquanto que a calibração utilizando o diâmetro a 5,2 m de altura resultou nos melhores valores de MAPE (5,41 %) e BIAS (-0,0029 cm). Os gráficos de

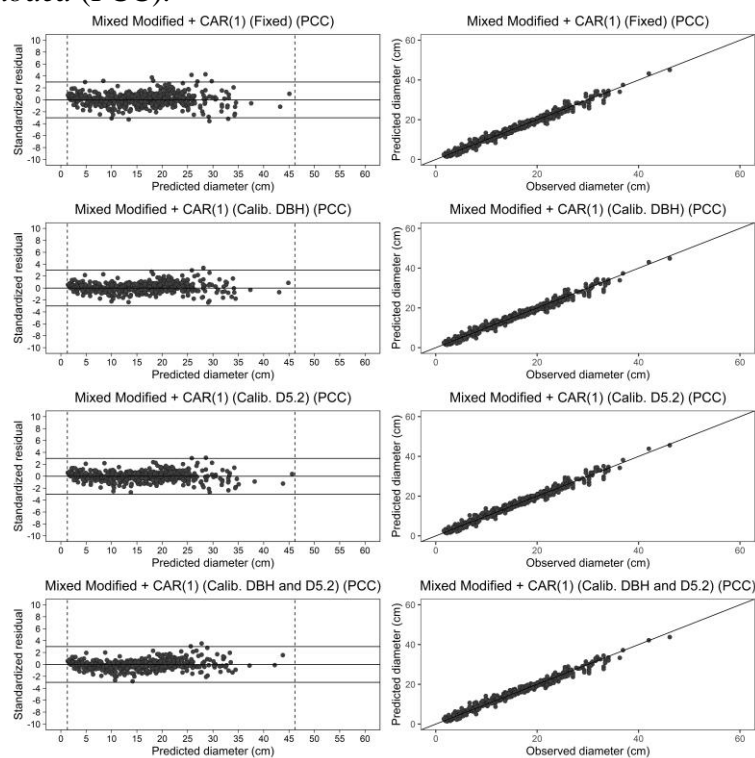
distribuição de resíduos padronizados evidenciam um bom comportamento para as predições de diâmetro com e sem calibração, enquanto que para as predições de volumes seccionais e total há um aumento dos resíduos conforme os valores preditos aumentam, mas mesmo assim os resíduos para o volume total permaneceram entre o intervalo de -3 e 3 para todas as alternativas de calibração (Figuras 10-12).

Tabela 10 – Resumo estatístico do desempenho das alternativas de calibração do modelo Kozak (2004) modificado na forma mista com a modelagem da estrutura de autocorrelação (CAR(1)) para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).

PCC												
Calibration method	Diameter					Total volume					Summed ranks	Overall rank
	RMSE	MAPE	EF	BIAS	p value	RMSE	MAPE	EF	BIAS	p value		
Calib. (DBH)	1,48	5,17	0,9638	0,1301	0,0063	0,0389	6,2357	0,9645	0,0058	0,1965	29,71	4,00
Calib. (D5.2)	1,46	5,11	0,9648	-0,0281	0,3283	0,0342	5,4180	0,9725	-0,0029	0,4599	16,87	1,69
Calib. (DBH and D5.2)	1,52	5,25	0,9617	-0,0214	0,2091	0,0348	5,4189	0,9716	-0,0053	0,1884	23,24	2,84
Fixed effects*	1,07	5,10	0,9810	0,0838	0,0682	0,0272	6,0717	0,9826	0,0038	0,3772	13,01	1,00
PCH												
Calib. (DBH)	1,61	5,63	0,9640	0,3067	0,0000	0,0723	8,8559	0,9287	0,0207	0,0013	31,83	3,63
Calib. (D5.2)	1,57	5,53	0,9660	0,1329	0,0004	0,0618	7,6427	0,9479	0,0110	0,0449	19,29	1,63
Calib. (DBH and D5.2)*	1,55	5,38	0,9668	0,1414	0,0005	0,0588	7,2805	0,9529	0,0081	0,1399	15,34	1,00
Fixed effects	1,18	5,65	0,9807	0,2498	0,0000	0,0505	8,4490	0,9652	0,0181	0,0030	17,64	1,37
PO												
Calib. (DBH)	1,66	5,02	0,9662	0,1574	0,0041	0,0685	6,6701	0,9538	0,0179	0,0214	31,90	3,48
Calib. (D5.2)*	1,61	4,84	0,9685	0,0618	0,3905	0,0589	5,7466	0,9659	0,0088	0,1797	16,26	1,00
Calib. (DBH and D5.2)	1,67	4,91	0,9659	0,1117	0,2102	0,0608	5,9255	0,9636	0,0106	0,1480	21,58	1,85
Fixed effects	1,20	4,97	0,9824	0,1424	0,0032	0,0472	6,4422	0,9781	0,0153	0,0256	17,13	1,14

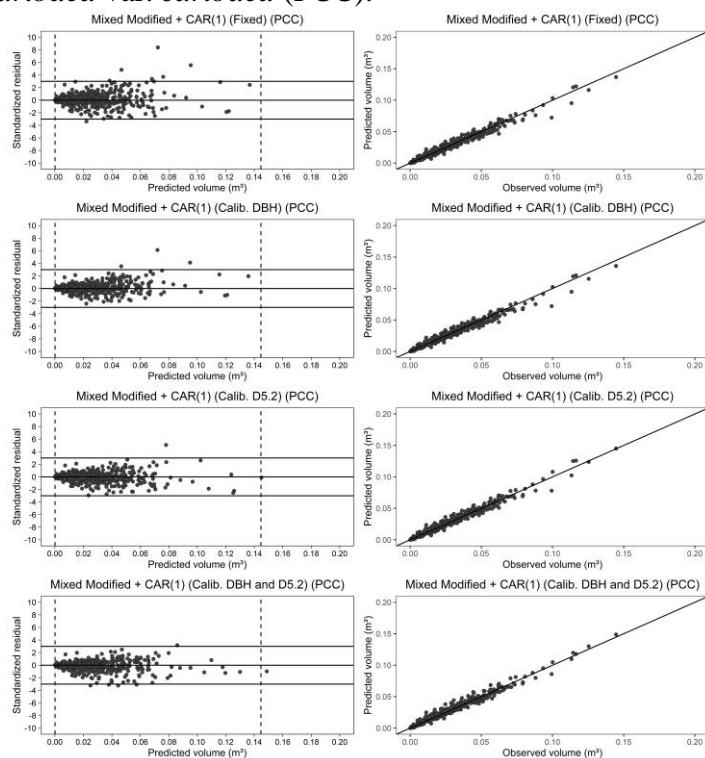
Onde os métodos seguidos por * indicam as alternativas de calibração de melhor desempenho de acordo o rank geral.

Figura 10 – Distribuição dos resíduos padronizados em função dos diâmetros preditos e valores de diâmetro preditos comparados com os valores observados para a base de treino para *Pinus caribaea* var. *caribaea* (PCC).



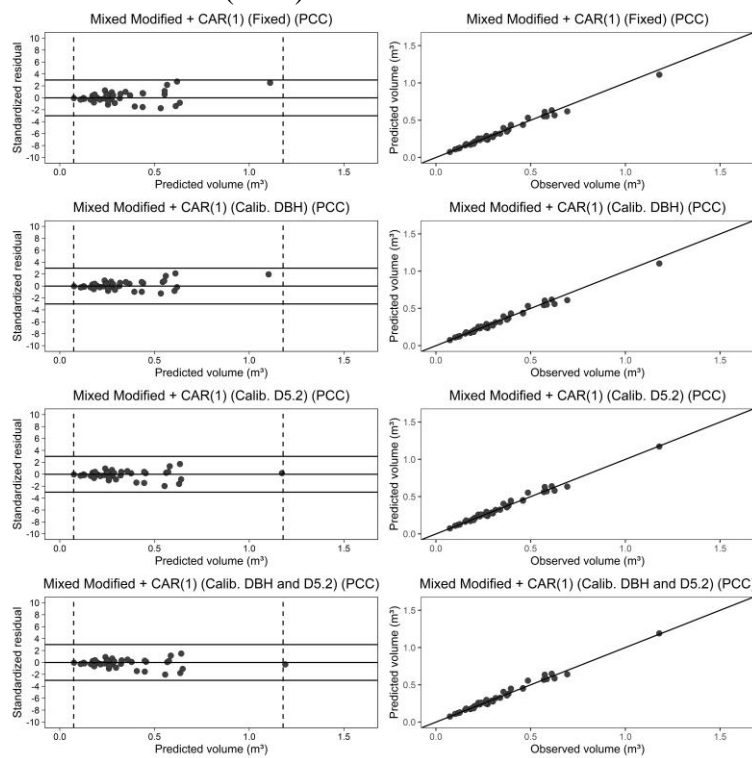
Fonte: Do autor (2024).

Figura 11 – Distribuição dos resíduos padronizados em função dos volumes seccionais preditos e valores de volumes seccionais preditos comparados com os valores observados para a base de treino para *Pinus caribaea* var. *caribaea* (PCC).



Fonte: Do autor (2024).

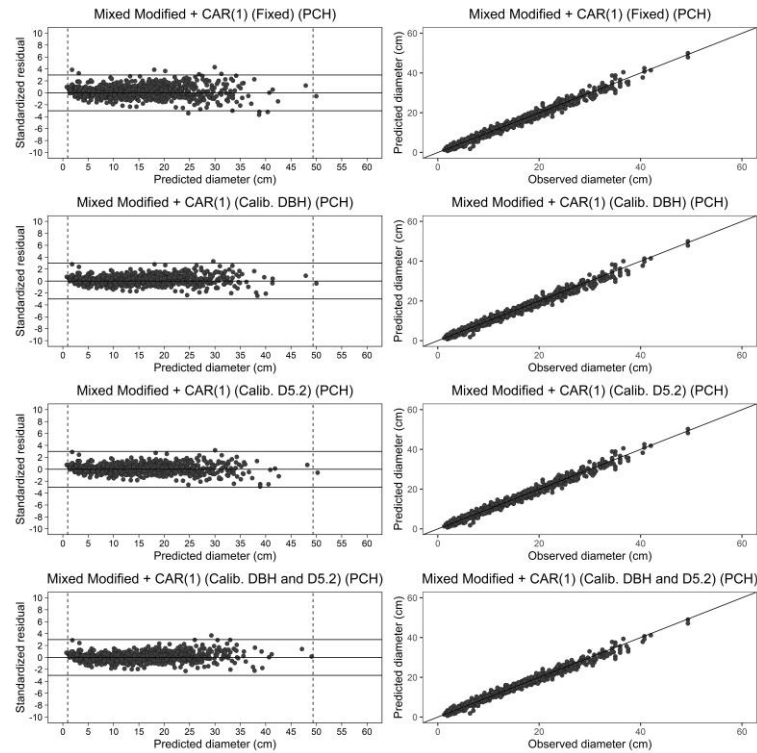
Figura 12 – Distribuição dos resíduos padronizados em função dos volumes totais preditos e valores de volumes totais preditos comparados com os valores observados para a base de treino para *Pinus caribaea* var. *caribaea* (PCC).



Fonte: Do autor (2024).

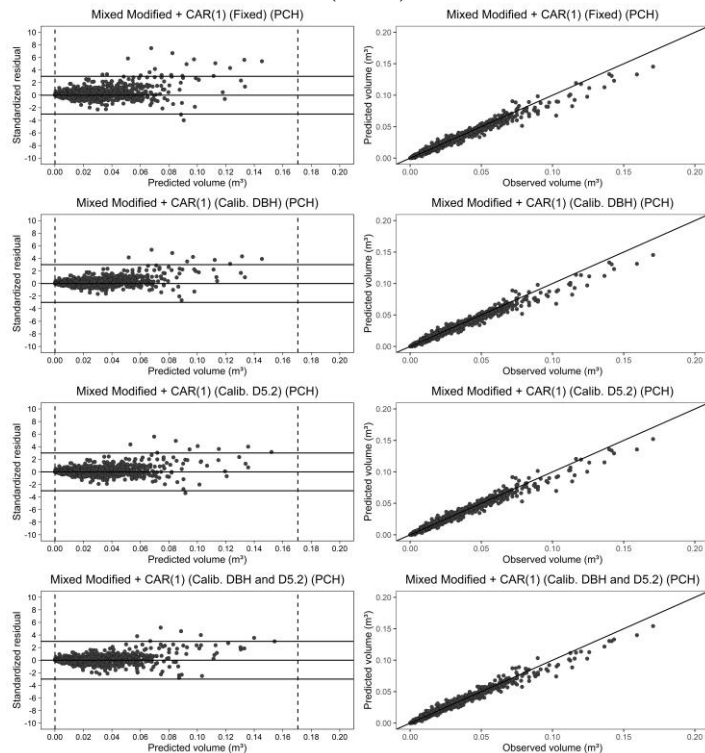
Considerando a espécie PCH, para as predições de diâmetro, a ausência de calibração resultou nos melhores valores de RMSE (1,18 cm) e EF (0,9897), enquanto que as calibrações com os diâmetros a 1,3 e 5,2 m e apenas o diâmetro a 5,2 m resultaram nos melhores valores de MAPE (5,38 %) e BIAS (0,1329 cm), respectivamente. Já para as predições de volume, a ausência de calibração resultou nos melhores valores de RMSE (0,0505 m³) e EF (0,9652), enquanto que desempenhos superiores de MAPE (7,28 %) e BIAS (0,0081 m³) foram observados para a calibração com os diâmetros a 1,3 e 5,2 m que, de acordo com o rank geral, resultou na melhor alternativa de calibração. A distribuição de resíduos padronizados resultante da calibração com diâmetros a 1,3 e 5,2 m mostram uma melhora no comportamento dos resíduos em relação à ausência de calibração para as predições de diâmetros, enquanto que para as predições de volumes seccionais e total, a melhora com a utilização da calibração reduz a tendência de aumento dos resíduos conforme os valores preditos aumentam e mantem os valores dos resíduos do volume total entre -3 e 3 (Figuras 13-15).

Figura 13 – Distribuição dos resíduos padronizados em função dos diâmetros preditos e valores de diâmetro preditos comparados com os valores observados para a base de treino para *Pinus caribaea* var. *hondurensis* (PCH).



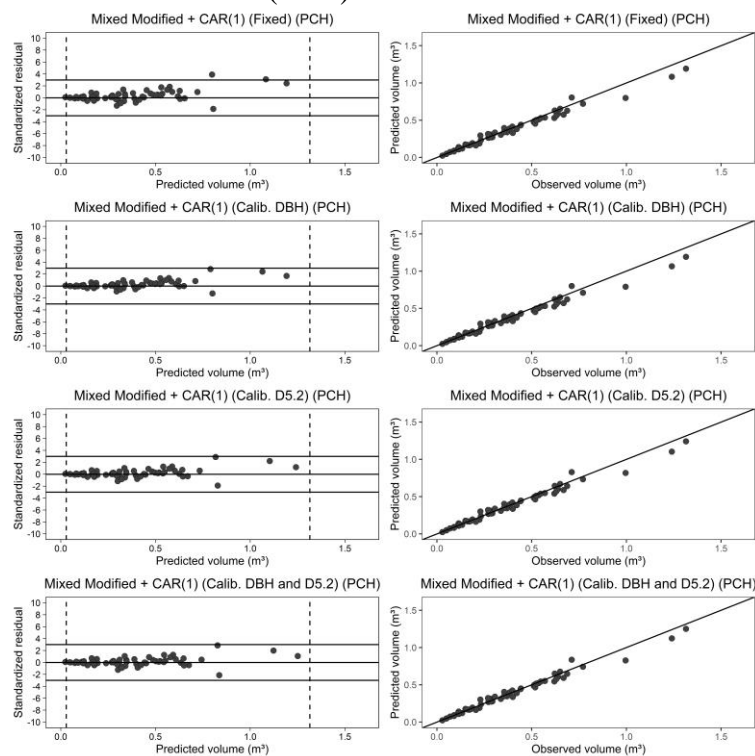
Fonte: Do autor (2024).

Figura 14 – Distribuição dos resíduos padronizados em função dos volumes seccionais preditos e valores de volumes seccionais preditos comparados com os valores observados para a base de treino para *Pinus caribaea* var. *hondurensis* (PCH).



Fonte: Do autor (2024).

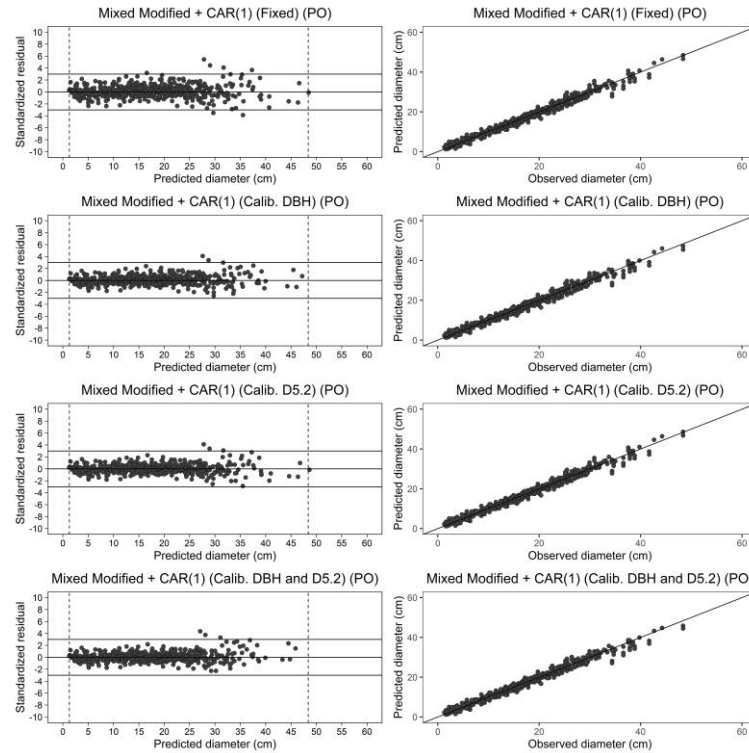
Figura 15 – Distribuição dos resíduos padronizados em função dos volumes totais preditos e valores de volumes totais preditos comparados com os valores observados para a base de treino para *Pinus caribaea* var. *hondurensis* (PCH).



Fonte: Do autor (2024).

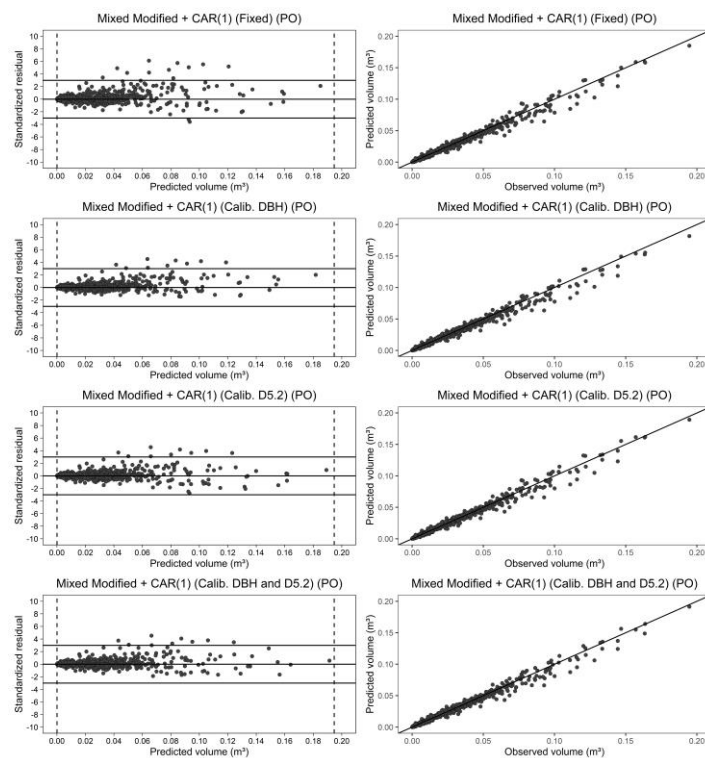
As previsões de diâmetro, para a espécie PO, com ausência de calibração geraram valores de RMSE (1,20 cm e 0,0472 m³) e EF (0,9824 e 0,9781) superiores para as previsões de diâmetro e volume, enquanto que a calibração utilizando os diâmetros a 5,2 m resultaram melhores valores de MAPE (4,84 e 5,75) e BIAS (0,0618 cm e 0,0088 m³). Observando os valores do rank geral, a calibração com os diâmetros a 5,2 m de altura foi a de melhor desempenho. A distribuição de resíduos padronizados resultante da calibração com diâmetros a 5,2 m possui resíduos mais concentrados em volta de zero em relação à ausência de calibração para as previsões de diâmetros, enquanto que para as previsões de volumes seccionais e total, o mesmo padrão é observado com a utilização da calibração, além de uma redução da tendência de aumento dos resíduos conforme os valores preditos aumentam, mantendo os resíduos entre -3 e 3 (Figuras 16-18).

Figura 16 – Distribuição dos resíduos padronizados em função dos diâmetros preditos e valores de diâmetro preditos comparados com os valores observados para a base de treino para as espécies *Pinus oocarpa* (PO).



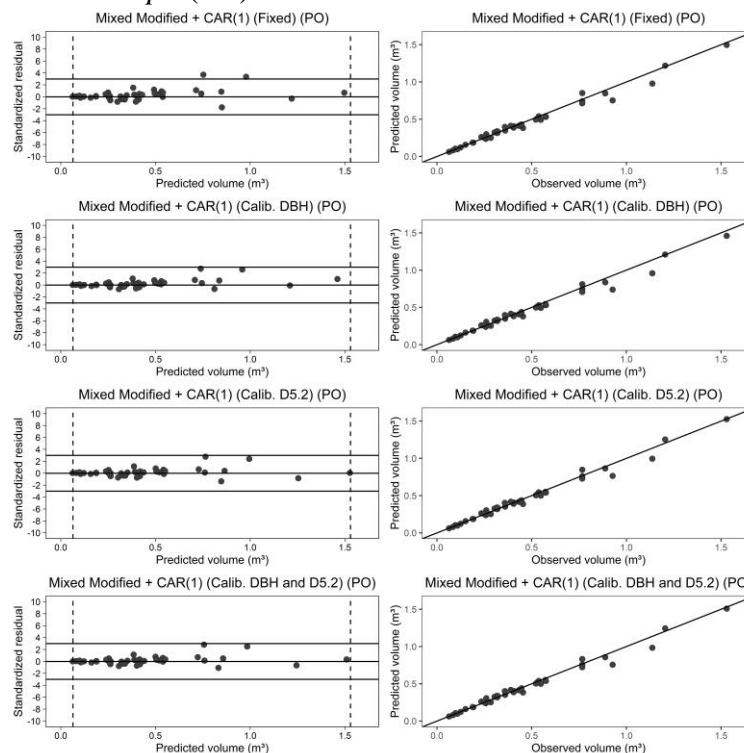
Fonte: Do autor (2024).

Figura 17 – Distribuição dos resíduos padronizados em função dos volumes seccionais preditos e valores de volumes seccionais preditos comparados com os valores observados para a base de treino para as espécies *Pinus oocarpa* (PO).



Fonte: Do autor (2024).

Figura 18 – Distribuição dos resíduos padronizados em função dos volumes totais preditos e valores de volumes totais preditos comparados com os valores observados para a base de treino para as espécies *Pinus oocarpa* (PO).



Fonte: Do autor (2024).

Dentre as melhores opções de calibração para cada espécie, apenas as predições de diâmetro para a espécie PCH foram diferentes significativamente, de acordo o teste de médias pareadas (Figura 19). As demais alternativas selecionadas resultaram em predições estatisticamente semelhantes aos valores observados, além de preservar as características da distribuição dos dados.

Além das métricas estatísticas médias para as predições de diâmetro e volume avaliadas para as bases de dados, foram extraídos os valores de BIAS para as porções relativas dos fustes, variando 10 a 100 %, com intervalos de 10%.

Os valores de BIAS para cada alternativa de calibração, para cada espécie ao longo do fuste podem ser observados na tabela 11 e, para uma análise mais eficiente, esses valores foram representados de forma gráfica (Figura 20). É possível observar que a variação dos valores de BIAS resultantes das predições de diâmetro, para a espécie PCC, confirmam os resultados anteriores, pois os valores de BIAS para a alternativa sem calibração resultou em uma curva mais estável, com valores mais próximos de zero, principalmente nas porções de 50 a 70%. Já para os valores de BIAS resultantes das predições dos volumes seccionais (Figura 21), a

ausência de calibração, além de apresentar valores semelhantes à calibração com o diâmetro a 5,2 m, resultou em valores com menos viés nas porções de 10% e 40 a 60%. Todos esses resultados corroboram com as métricas médias extraídas anteriormente, onde a opção sem calibração foi apontada como a mais eficiente.

Figura 19 – Resultados dos testes de médias pareadas comparando as previsões de diâmetro e volume total, realizadas pelas alternativas de calibração selecionadas, com os dados observados para a base de teste para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).

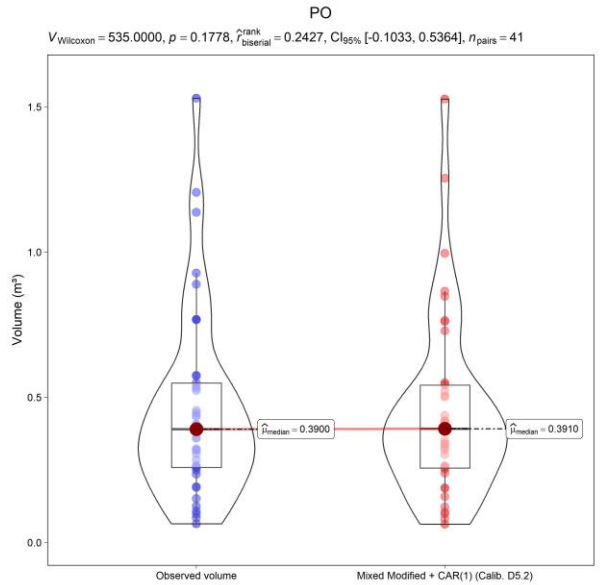
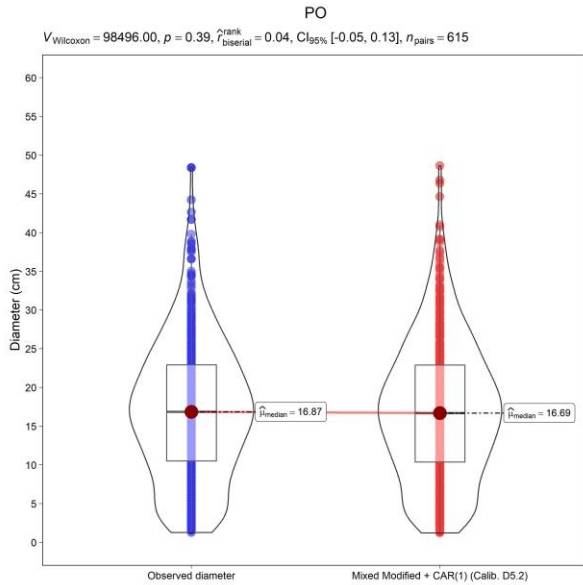
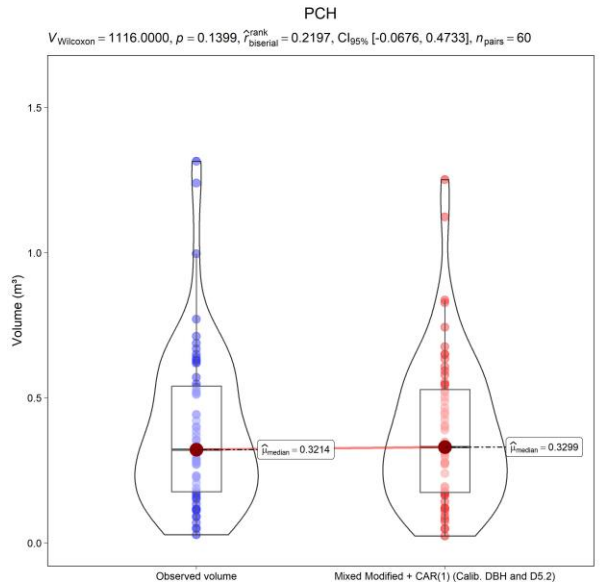
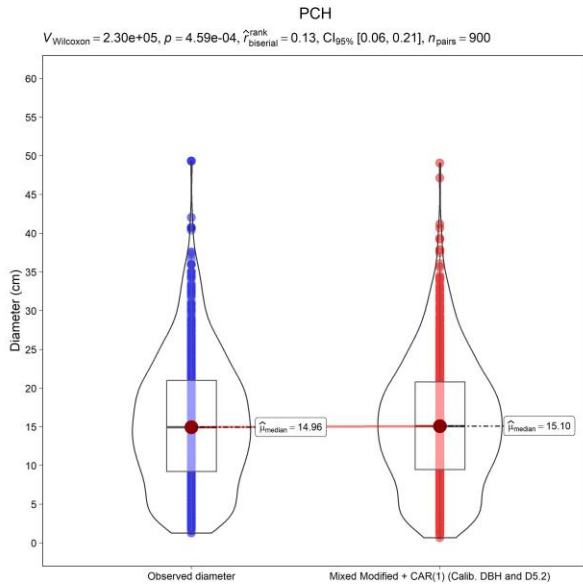
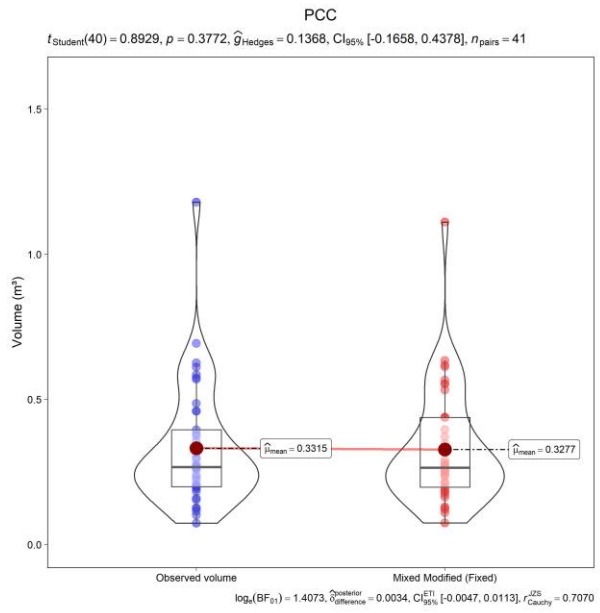
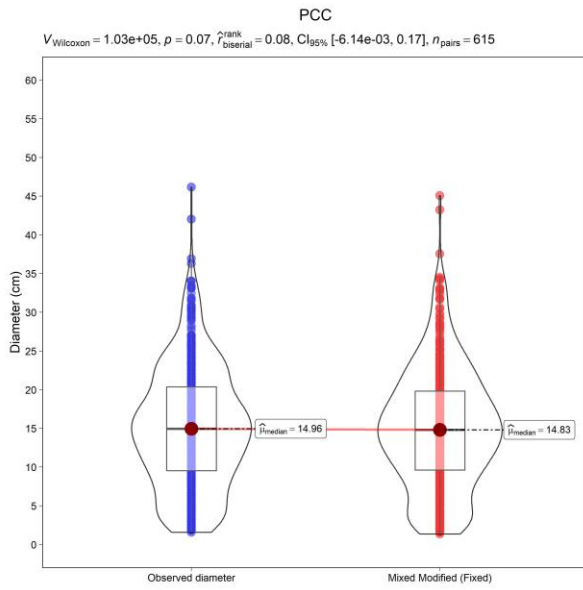
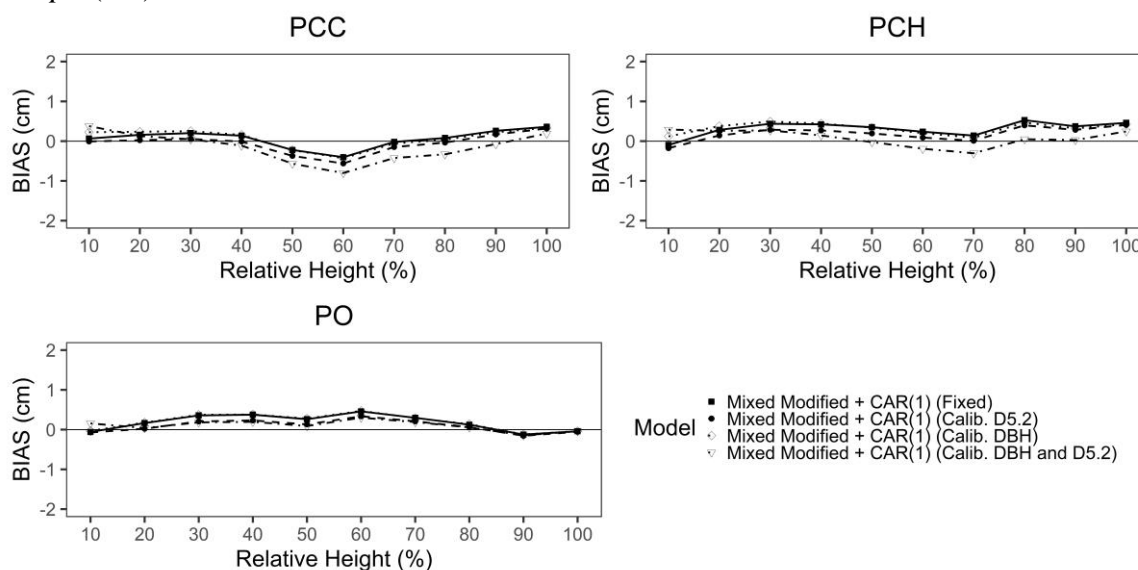


Tabela 11 – Valores de BIAS em porções relativas do fuste resultantes das predições de diâmetros e volumes seccionais realizadas pelas alternativas de calibração avaliadas para a base de treino para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).

		BIAS							
RH	n	Diameter (cm)				Volume (m ³)			
		Calibrated (DBH)	Calibrated (D5.2)	Calibrated (DBH and D5.2)	Fixed effects	Calibrated (DBH)	Calibrated (D5.2)	Calibrated (DBH and D5.2)	Fixed effects
PCC									
10	154	0,2253	-0,0074	0,3794	0,0644	0,0003	-0,0003	0,0005	0,0000
20	79	0,2370	0,0235	0,1285	0,1583	0,0009	0,0002	0,0004	0,0007
30	80	0,2507	0,0543	0,0461	0,2020	0,0010	0,0002	0,0000	0,0008
40	43	0,1568	-0,0098	-0,1115	0,1376	0,0005	-0,0005	-0,0013	0,0004
50	43	-0,2253	-0,3708	-0,5652	-0,2224	-0,0008	-0,0016	-0,0028	-0,0008
60	38	-0,4288	-0,5614	-0,8018	-0,4048	-0,0007	-0,0014	-0,0026	-0,0006
70	39	-0,0568	-0,1472	-0,4243	-0,0193	0,0010	0,0006	-0,0006	0,0011
80	46	0,0363	-0,0362	-0,3308	0,0804	0,0006	0,0003	-0,0005	0,0007
90	45	0,2250	0,1687	-0,0736	0,2603	0,0002	0,0001	-0,0002	0,0002
100	48	0,3369	0,3167	0,1915	0,3611	0,0001	0,0001	0,0000	0,0001
PCH									
10	224	0,1179	-0,1772	0,2907	-0,0969	0,0005	-0,0004	0,0007	-0,0001
20	112	0,3796	0,1446	0,2584	0,2836	0,0018	0,0009	0,0012	0,0015
30	118	0,4933	0,2863	0,2717	0,4367	0,0024	0,0016	0,0013	0,0022
40	63	0,4427	0,2702	0,1432	0,4225	0,0031	0,0021	0,0010	0,0030
50	62	0,3397	0,1931	-0,0246	0,3508	0,0025	0,0017	0,0003	0,0026
60	57	0,2024	0,0893	-0,1918	0,2371	0,0019	0,0014	-0,0002	0,0021
70	59	0,0888	0,0104	-0,3055	0,1404	0,0010	0,0007	-0,0007	0,0012
80	67	0,4659	0,3973	0,0495	0,5280	0,0015	0,0013	0,0003	0,0017
90	67	0,3235	0,2844	0,0284	0,3734	0,0003	0,0002	-0,0001	0,0003
100	71	0,4268	0,4073	0,2447	0,4606	0,0001	0,0001	0,0000	0,0001
PO									
10	158	-0,0287	-0,0701	0,1566	-0,0586	0,0003	-0,0003	0,0004	0,0000
20	75	0,1781	0,0221	0,0400	0,1656	0,0010	0,0003	0,0003	0,0009
30	80	0,3743	0,2069	0,1849	0,3576	0,0020	0,0011	0,0009	0,0018
40	43	0,3812	0,2319	0,1966	0,3771	0,0028	0,0016	0,0014	0,0026

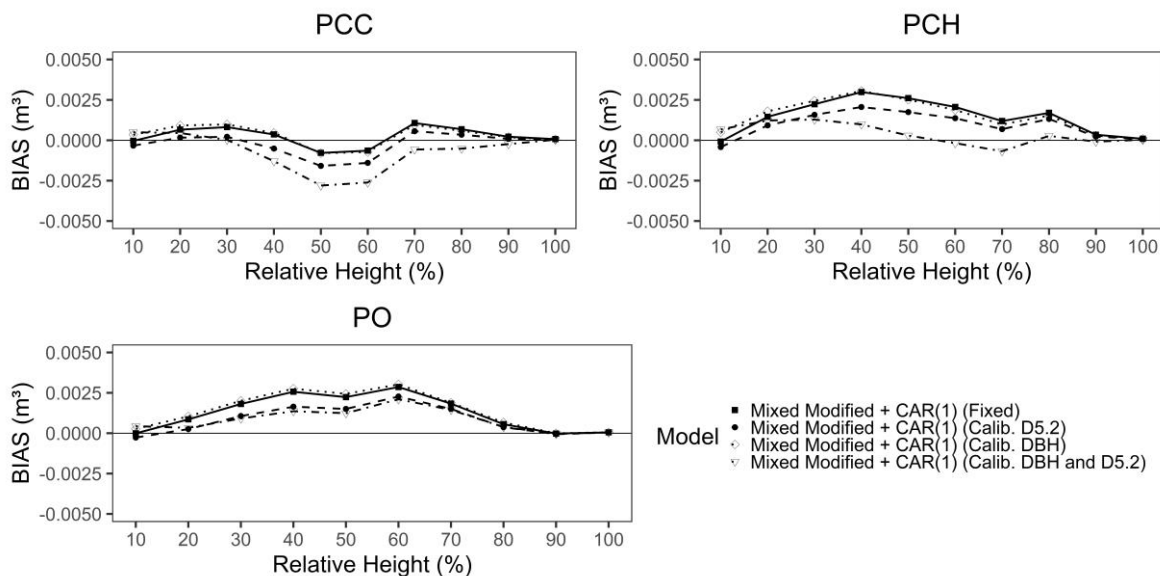
50	45	0,2788	0,1331	0,0902	0,2636	0,0024	0,0015	0,0012	0,0022
60	40	0,4708	0,3378	0,3014	0,4584	0,0030	0,0023	0,0021	0,0029
70	34	0,2939	0,2089	0,1908	0,2972	0,0019	0,0015	0,0014	0,0018
80	52	0,1368	0,0628	0,0565	0,1234	0,0007	0,0004	0,0004	0,0006
90	42	-0,1197	-0,1543	-0,1433	-0,1240	0,0000	-0,0001	0,0000	0,0000
100	46	-0,0393	-0,0538	-0,0388	-0,0409	0,0001	0,0000	0,0001	0,0001

Figura 20 – Comportamento dos valores de BIAS em porções relativas resultantes das predições de diâmetro realizadas pelas alternativas de calibração avaliadas para a base de teste para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).



Fonte: Do autor (2024).

Figura 21 – Comportamento dos valores de BIAS em porções relativas resultantes das predições de volumes seccionais realizadas pelas alternativas de calibração avaliadas para a base de teste para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).



Fonte: Do autor (2024).

Para a espécie PCH, também corroborando com os resultados anteriores, os valores de BIAS resultantes das predições de diâmetro, para a alternativa de calibração utilizando os diâmetros a 1,3 e 5,2 m resultaram em valores mais próximos de zero, principalmente nas seções

de 30 a 50% e 80 a 100%, onde as outras alternativas demonstraram maiores desvios. Já as predições de volume resultaram em BIAS, mais próximos de zero em quase todas as porções do fuste (30 a 100%), enquanto que as outras alternativas resultaram em vieses inflados, principalmente na porção central do fuste.

Analisando os resultados da espécie PO, pode-se observar que as predições de diâmetro resultaram em um menor viés e uma maior estabilidade dos valores de BIAS ao longo do fuste para todas as alternativas de calibração. Além disso, os valores de BIAS, para a alternativa de calibração utilizando os diâmetros a 5,2 m resultaram em valores mais próximos de zero, principalmente nas seções de 30 a 80% e 100%. Para as predições de volume, a melhor alternativa de calibração, gerou um menor viés em quase toda a porção do fuste (20 a 90%).

3.3 Redes Neurais Artificiais

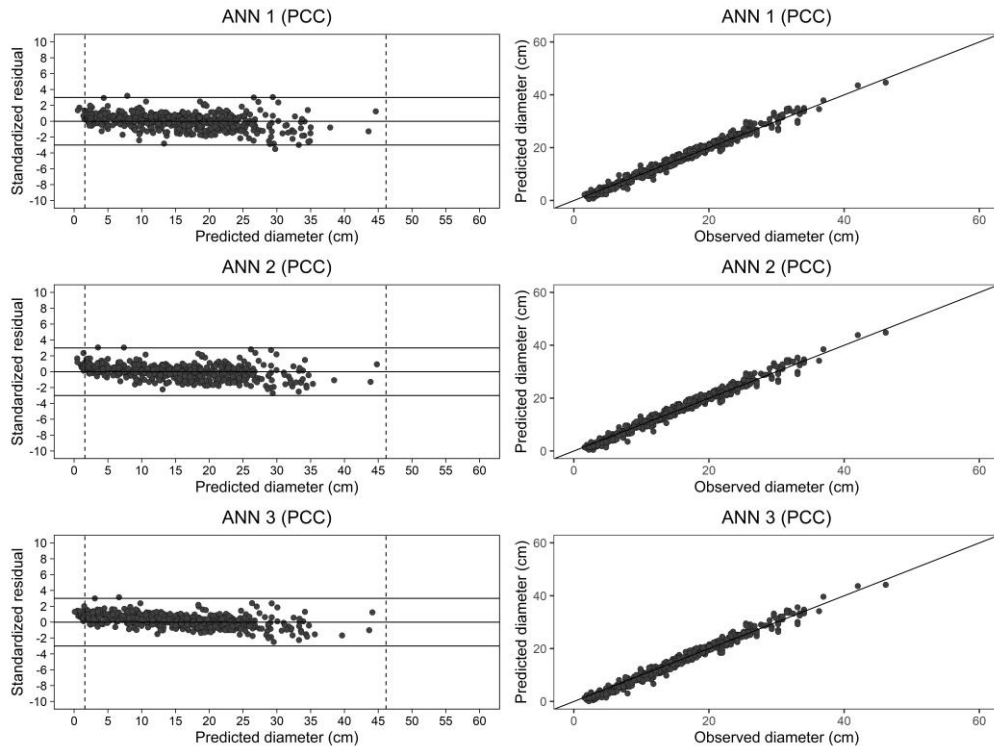
Como resultado das diferentes combinações testadas no presente estudo, 540 modelos de RNA foram submetidos ao treinamento, porém, 473 modelos de RNA convergiram, considerando as configurações utilizadas. Após o treinamento de cada RNA, as métricas estatísticas para a base de treino e teste foram extraídas e, por fim, foi calculado o rank geral contendo todas as RNA.

As RNA foram treinadas para a base de dados contendo as três espécies simultaneamente, um vez que, dentre as variáveis de entrada, variáveis dummy referentes às espécies estavam presentes. Sendo assim, as três RNA de melhor desempenho geral foram selecionadas e suas estatísticas, avaliadas por espécie, podem ser observadas na tabela 12. Para a espécie PCC, a RNA 1 com duas camadas ocultas, com sete neurônios cada, função de ativação logística e função de saída linear, em que 120 pesos foram ajustados, resultou nos melhores resultados de RMSE (1,22 cm e 0,0345 m³), MAPE (5,37 e 6,24 %), EF (0,9753 e 0,9721) e BIAS (-0,0014 cm e -0,0059 m³) para as predições de diâmetro e volume, respectivamente. Além disso, a RNA 1 resultou em uma distribuição de resíduos mais homogênea, enquanto que para as RNA 2 e 3, é possível observar uma tendência de subestimação para os menores valores de diâmetro (Figura 22). Analisando os resíduos do volumes seccionais (Figura 23), há uma tendência de superestimação para os maiores valores para todas as três RNA, enquanto que esse padrão não é verificado para as predições de volume total, onde os resíduos se mantiveram dentro do intervalo de -3 e 3 (Figura 24).

Tabela 12 – Resumo estatístico do desempenho das melhores RNA para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).

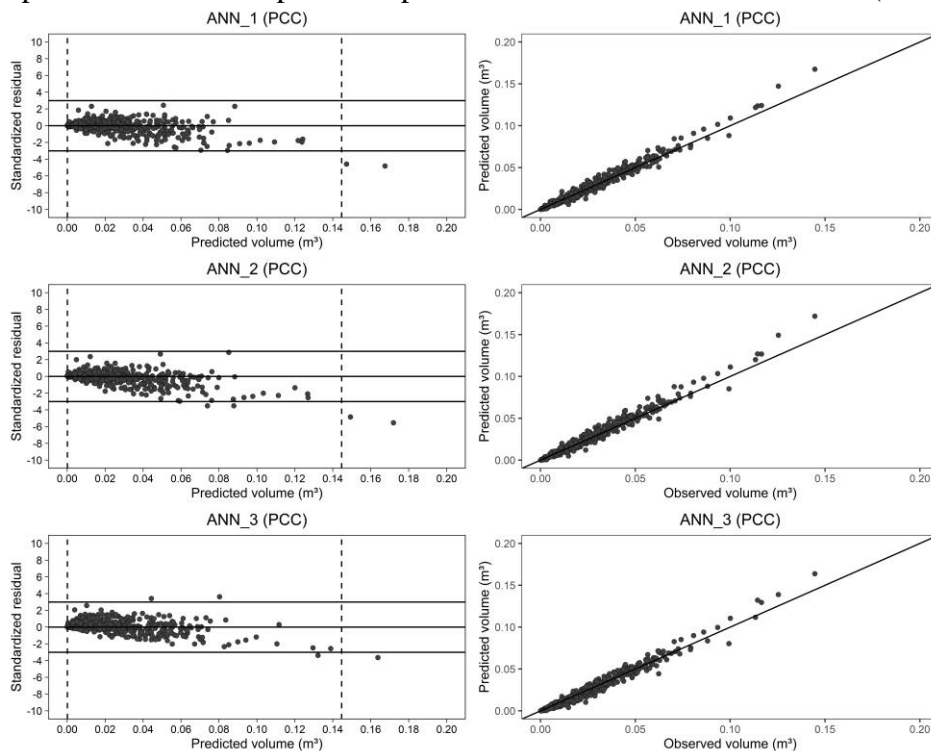
Model	Neurons	Activation function		Param.	Diameter					Volume					Summed ranks	Overall rank
		Hidden layer	Output layer		RMSE	MAPE	EF	BIAS	p value	RMSE	MAPE	EF	BIAS	p value		
PCC																
ANN1	7-7	tanh	Linear	120	1,22	5,37	0,9753	-0,0014	0,2880	0,0345	6,2410	0,9721	-0,0059	0,9387	8,00	1,00
ANN2	8-15	logistic	Linear	215	1,45	5,74	0,9653	-0,0081	0,9008	0,0417	6,5629	0,9592	-0,0107	0,3033	17,03	2,34
ANN3	11-11	tanh	tanh	232	1,63	6,96	0,9558	0,4089	0,0000	0,0386	7,7654	0,9649	0,0088	0,0132	21,49	3,00
PCH																
ANN1	7-7	tanh	Linear	120	1,20	5,40	0,9799	0,0300	0,0682	0,0424	6,5223	0,9754	-0,0003	0,5733	8,00	1,00
ANN2	8-15	logistic	Linear	215	1,41	5,97	0,9726	-0,0863	0,0720	0,0474	6,8838	0,9694	-0,0094	0,1419	18,20	2,30
ANN3	11-11	tanh	tanh	232	1,44	5,89	0,9712	-0,1952	0,0000	0,0524	7,5632	0,9625	-0,0123	0,0085	23,73	3,00
PO																
ANN1	7-7	tanh	Linear	120	1,60	6,26	0,9687	0,1085	0,0001	0,0788	8,7687	0,9389	0,0020	0,0348	12,77	1,16
ANN2	8-15	logistic	Linear	215	1,58	5,34	0,9697	-0,2433	0,0000	0,0792	7,6586	0,9383	-0,0204	0,0360	12,10	1,00
ANN3	11-11	tanh	tanh	232	1,70	5,70	0,9647	-0,1864	0,0001	0,0924	8,1187	0,9160	-0,0189	0,1673	20,60	3,00

Figura 22 – Distribuição dos resíduos padronizados em função dos diâmetros preditos e valores de diâmetro preditos das melhores RNA comparados com os valores observados para a base de teste para as espécies *Pinus caribaea* var. *caribaea* (PCC).



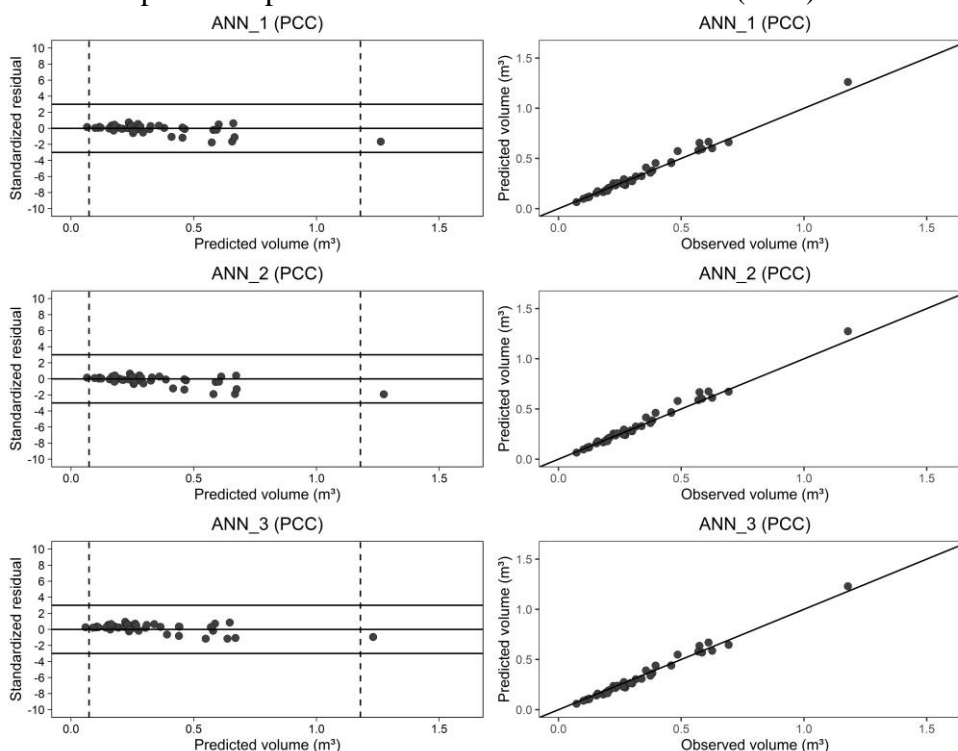
Fonte: Do autor (2024).

Figura 23 – Distribuição dos resíduos padronizados em função dos volumes seccionais preditos e valores de volumes seccionais preditos das melhores RNA comparados com os valores observados para a base de teste para as espécies *Pinus caribaea* var. *caribaea* (PCC).



Fonte: Do autor (2024).

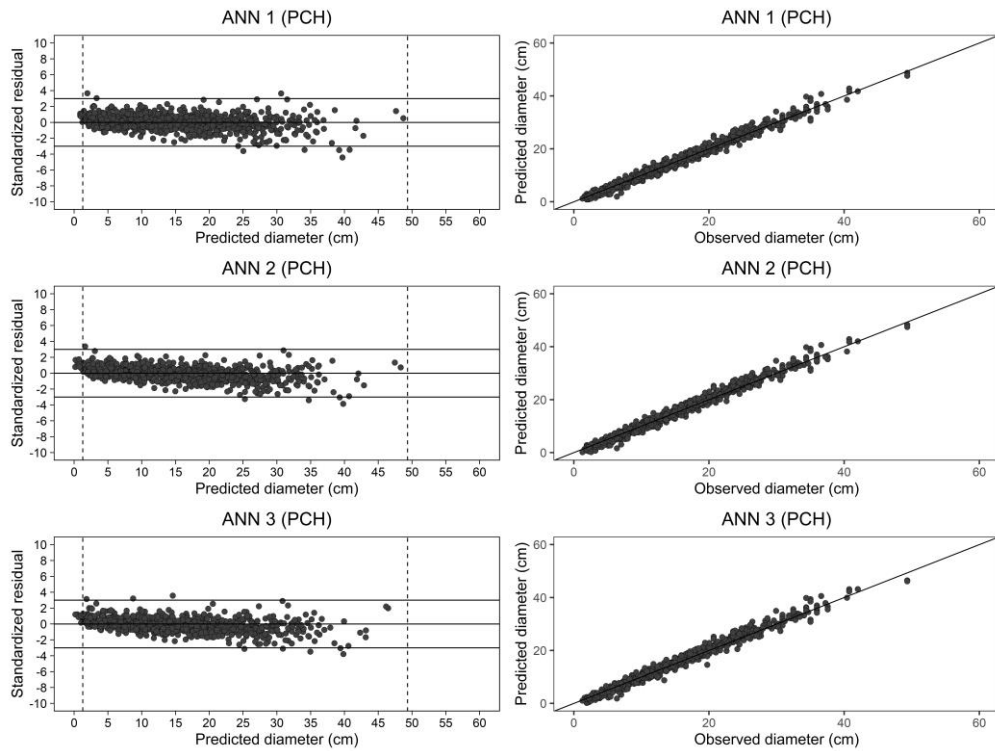
Figura 24 – Distribuição dos resíduos padronizados em função dos volumes totais preditos e valores de volumes totais preditos das melhores RNA comparados com os valores observados para a base de teste para as espécies *Pinus caribaea* var. *caribaea* (PCC).



Fonte: Do autor (2024).

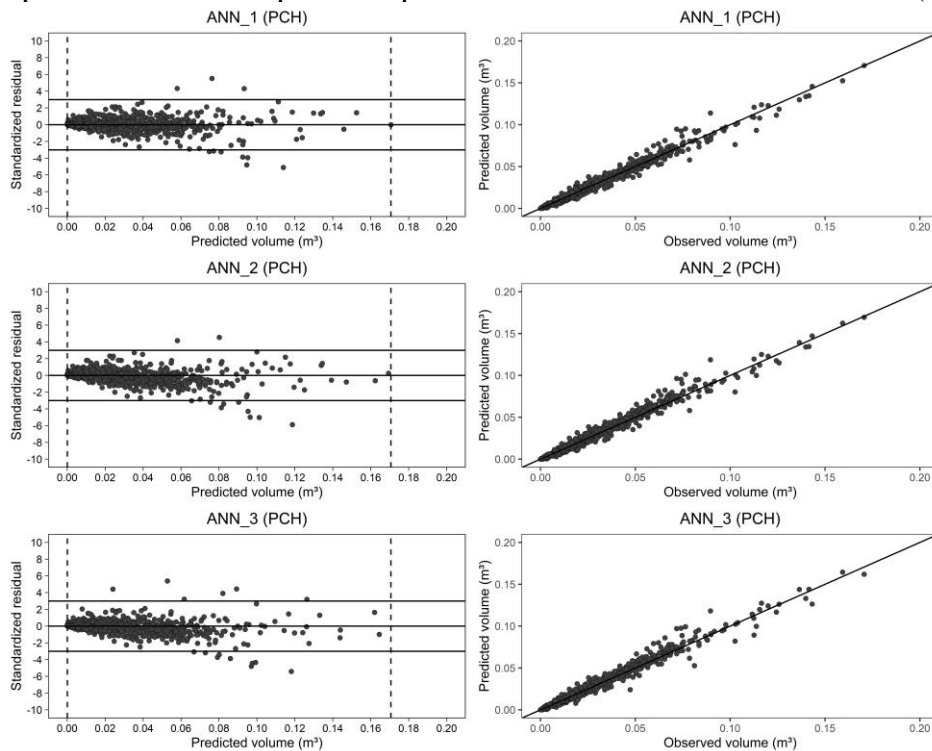
Para a espécie PCH, a RNA 1 também resultou em uma performance superior em relação aos valores de RMSE (1,20 cm e 0,0424 m³), MAPE (5,40 e 6,52 %), EF (0,9799 e 0,9754) e BIAS (0,0300 cm e -0,0003 m³) para as predições de diâmetro e volume, respectivamente. Já as distribuições de resíduos, corroborando com as métricas estatísticas, indicam uma distribuição mais simétrica para as predições de diâmetro da RNA 1, enquanto que para as RNA 2 e 3 há uma maior tendência de subestimação para os menores valores de diâmetro (Figura 25). As predições de volumes seccionais e total para todas as três RNA resultaram em resíduos sem tendências pronunciadas e quase todos os resíduos permaneceram no intervalo de -3 e 3 (Figura 26 e 27).

Figura 25 – Distribuição dos resíduos padronizados em função dos diâmetros preditos e valores de diâmetro preditos das melhores RNA comparados com os valores observados para a base de teste para as espécies *Pinus caribaea* var. *hondurensis* (PCH).



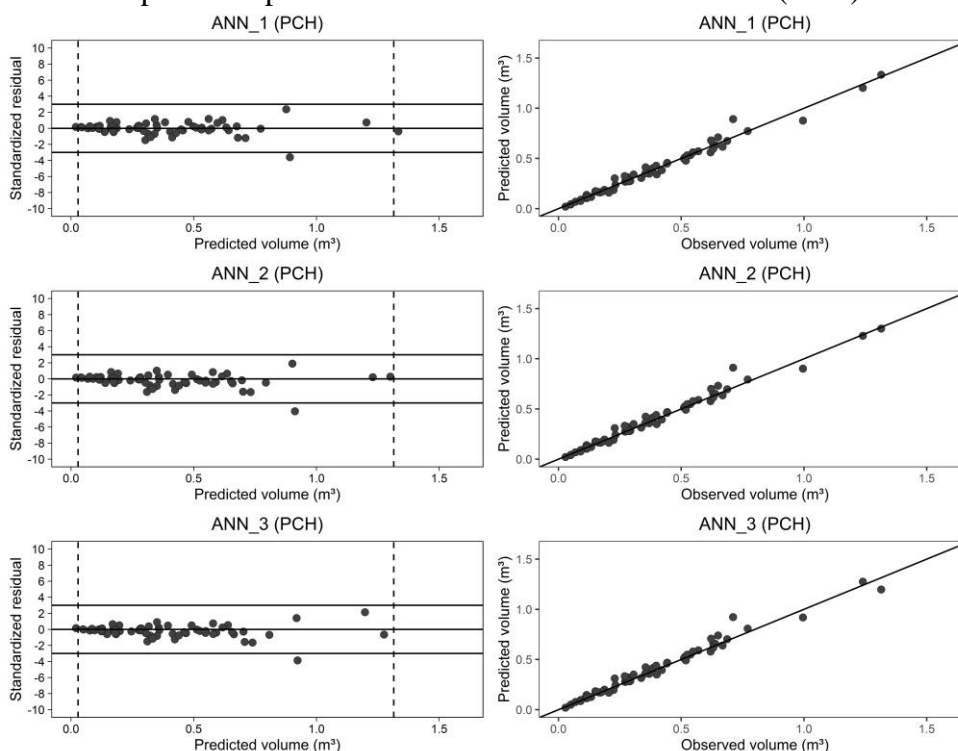
Fonte: Do autor (2024).

Figura 26 – Distribuição dos resíduos padronizados em função dos volumes seccionais preditos e valores de volumes seccionais preditos das melhores RNA comparados com os valores observados para a base de teste para as espécies *Pinus caribaea* var. *hondurensis* (PCH).



Fonte: Do autor (2024).

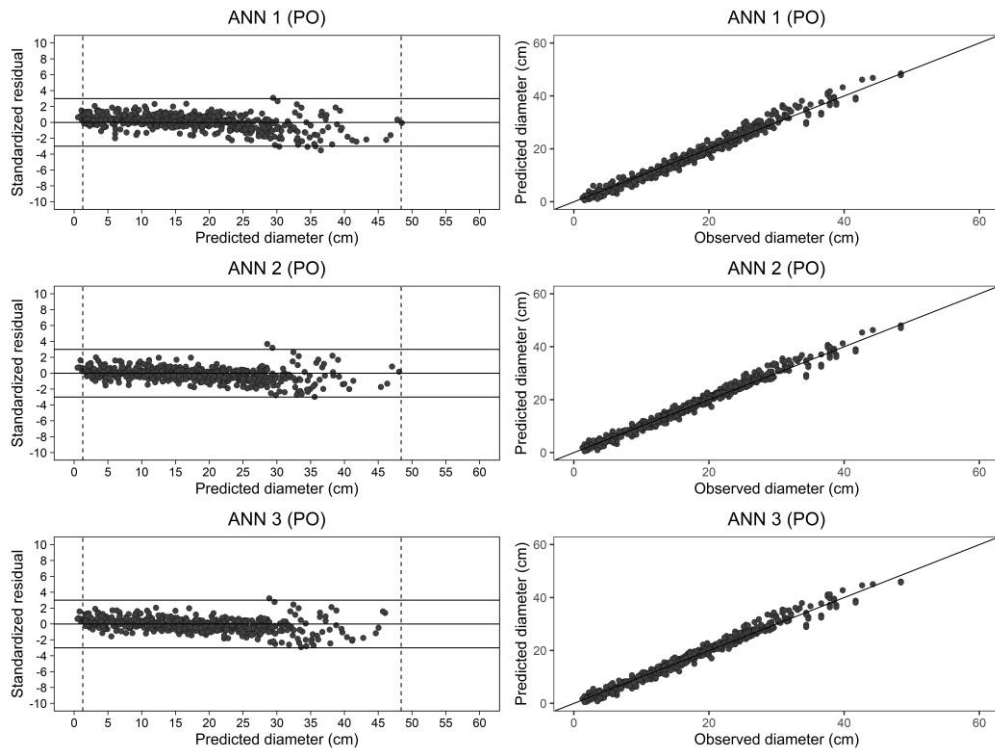
Figura 27 – Distribuição dos resíduos padronizados em função dos volumes totais preditos e valores de volumes totais preditos das melhores RNA comparados com os valores observados para a base de teste para as espécies *Pinus caribaea* var. *hondurensis* (PCH).



Fonte: Do autor (2024).

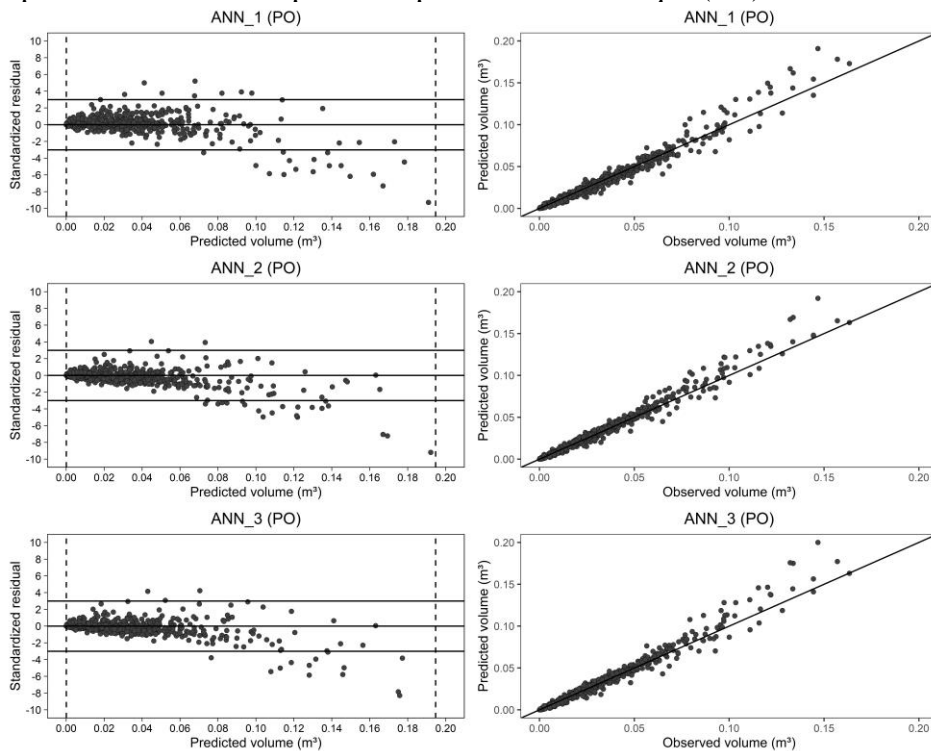
Para a espécie PO, a RNA 1 resultou nos melhores valores de BIAS (0,0020 cm) para as predições de diâmetro e nos melhores valores de RMSE (0,0788 m³), EF (0,9389) e BIAS (0,0020 m³) para as predições de volume. Já a RNA 2 resultou em performances superiores de RMSE (1,58 cm), MAPE (5,34 %) e EF (0,9697) para as predições de diâmetro e em melhores valores de MAPE (7,66 %) para as predições de volume, de modo que a RNA 2 resultou na melhor performance para a espécie PO. Os gráficos de resíduos de diâmetro (Figura 28), para as três RNA, indicam resíduos normalmente distribuídos e próximos de zero. Porém, quando analisando os volumes seccionais (Figura 29), pode-se observar uma tendência de superestimação acentuada para os maiores valores. Já para o volume total (Figura 30) os resíduos se comportaram de forma simétrica, sem a presença de tendências.

Figura 28. Distribuição dos resíduos padronizados em função dos diâmetros preditos e valores de diâmetro preditos das melhores RNA comparados com os valores observados para a base de treino para as espécies *Pinus oocarpa* (PO).



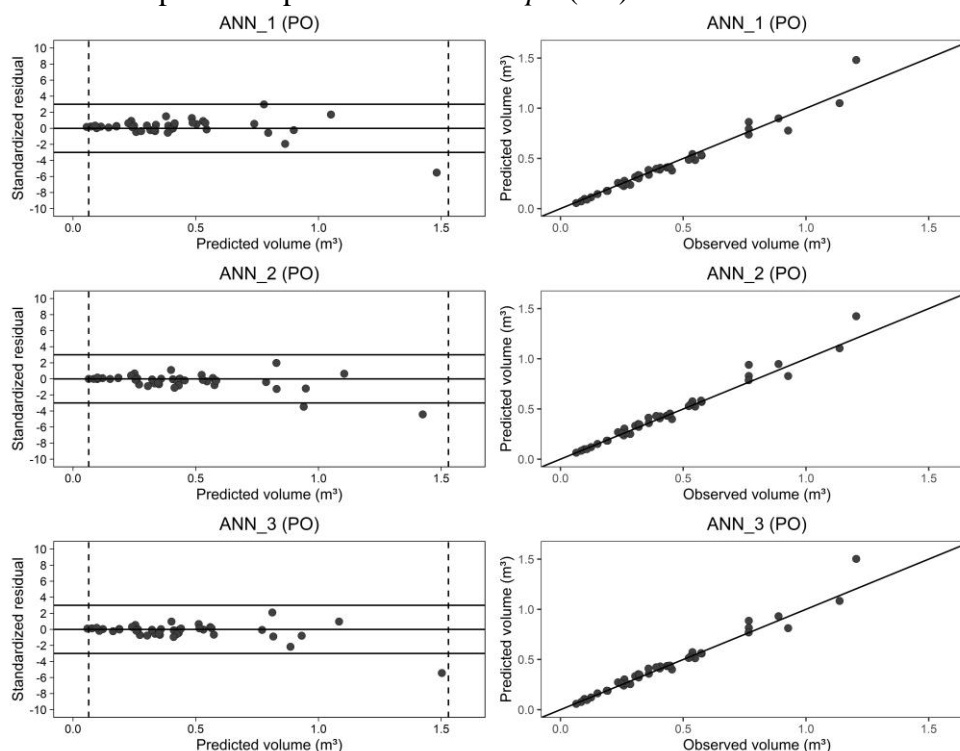
Fonte: Do autor (2024).

Figura 29 – Distribuição dos resíduos padronizados em função dos volumes seccionais preditos das melhores RNA e valores de volumes seccionais preditos comparados com os valores observados para a base de treino para as espécies *Pinus oocarpa* (PO).



Fonte: Do autor (2024).

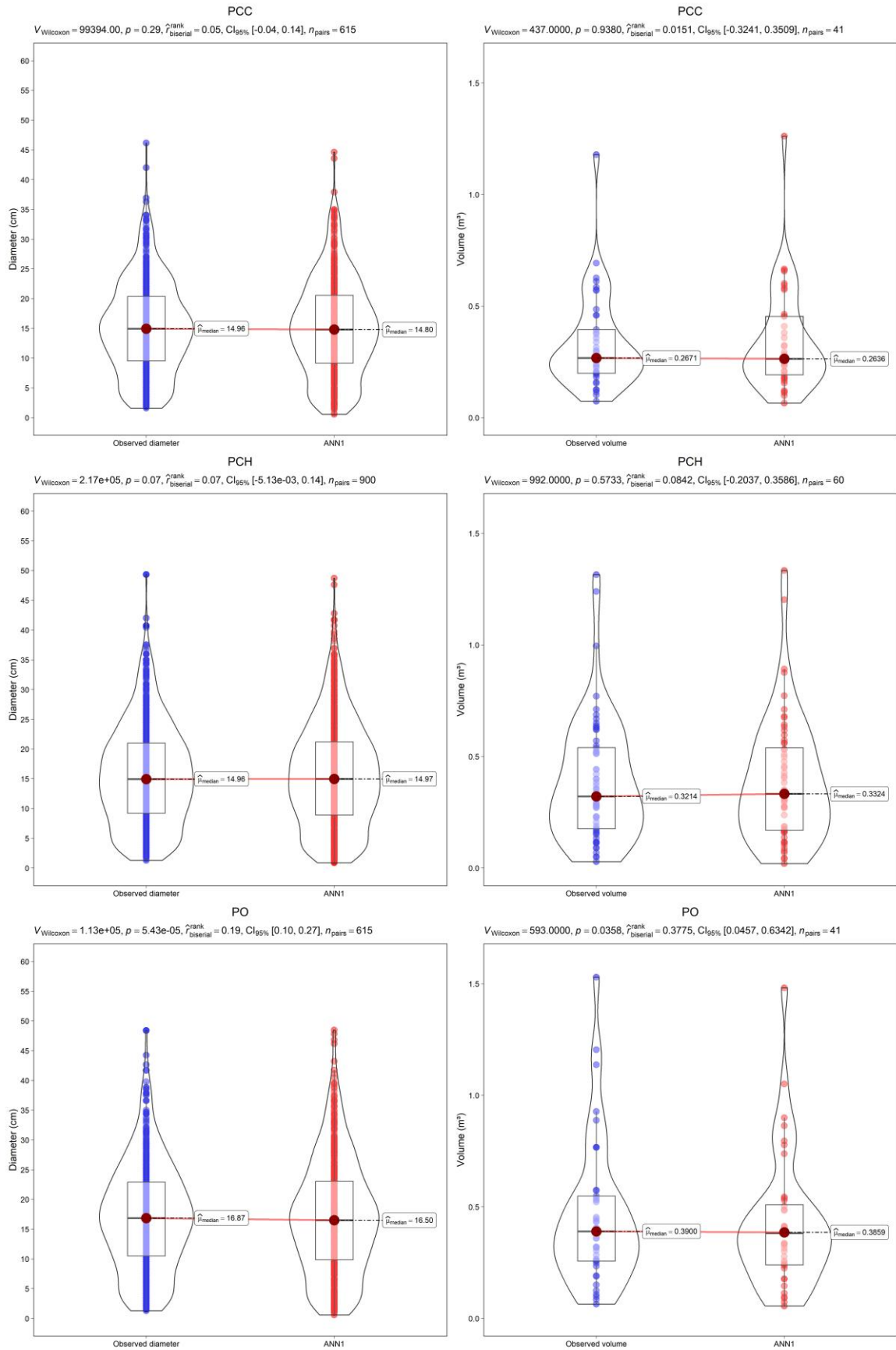
Figura 30 – Distribuição dos resíduos padronizados em função dos volumes totais preditos das melhores RNA e valores de volumes totais preditos comparados com os valores observados para a base de treino para as espécies *Pinus oocarpa* (PO).



Fonte: Do autor (2024).

A RNA 1 resultou em predições estatisticamente semelhantes aos valores observados para as espécies PCC e PCH, de acordo o teste de médias pareadas para os valores de diâmetro e volume, além da distribuição dos valores preditos continuar similar a dos dados observados (Figura 31). No entanto, as predições foram significativamente diferentes para a espécie PO.

Figura 31 – Resultados dos testes de médias pareadas comparando as predições de diâmetro e volume total, realizadas pelas melhores RNA, com os dados observados para a base de treino para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).



Fonte: Do autor (2024).

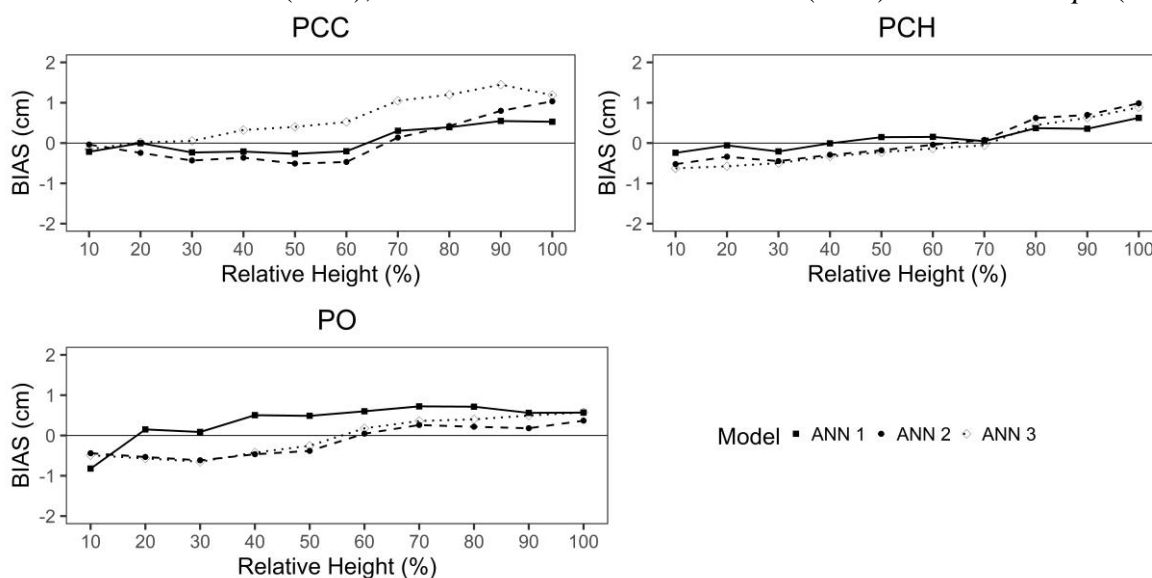
Os valores de BIAS resultantes das RNA, para cada espécie ao longo do fuste podem ser observados na tabela 13. A variação dos valores de BIAS resultantes das predições de diâmetro (Figura 32), para a espécie PCC, indicam menos viés para as porções de 40 a 70 % e 80 a 100 % do fuste, confirmando a performance superior da RNA 1. Avaliando as predições de volume, as três RNA resultaram em um comportamento instável do BIAS (Figura 33), mas este foi menos pronunciado para a RNA 1. Esses resultados evidenciam o melhor desempenho da RNA 1.

Tabela 13 – Valores de BIAS em porções relativas resultantes das predições de diâmetros e volumes seccionais realizadas pelas melhores RNA avaliadas para a base de teste para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).

		BIAS					
RH	n	Diameter (cm)			Volume (m ³)		
		ANN1	ANN2	ANN3	ANN1	ANN2	ANN3
PCC							
10	154	-0,2165	-0,0378	-0,1490	-0,0006	-0,0005	-0,0006
20	79	-0,0039	-0,2453	0,0179	-0,0005	-0,0012	-0,0003
30	82	-0,2337	-0,4317	0,0518	-0,0010	-0,0017	0,0002
40	41	-0,2110	-0,3636	0,3232	-0,0017	-0,0029	0,0007
50	43	-0,2671	-0,5086	0,4019	-0,0016	-0,0027	0,0010
60	38	-0,2051	-0,4706	0,5219	-0,0007	-0,0014	0,0018
70	39	0,3036	0,1353	1,0495	0,0014	0,0013	0,0035
80	46	0,3925	0,4232	1,1990	0,0011	0,0014	0,0027
90	45	0,5473	0,7979	1,4500	0,0005	0,0007	0,0011
100	48	0,5280	1,0340	1,1894	0,0001	0,0002	0,0002
PCH							
10	224	-0,2420	-0,5227	-0,6288	-0,0005	-0,0014	-0,0018
20	112	-0,0586	-0,3394	-0,5759	-0,0005	-0,0014	-0,0020
30	120	-0,2093	-0,4500	-0,4983	-0,0006	-0,0017	-0,0017
40	61	-0,0077	-0,2925	-0,3312	0,0002	-0,0016	-0,0014
50	62	0,1459	-0,1797	-0,2289	0,0007	-0,0007	-0,0006
60	57	0,1520	-0,0412	-0,1342	0,0008	0,0004	0,0004
70	59	0,0459	0,0770	-0,0545	0,0002	0,0007	0,0005
80	67	0,3696	0,6200	0,4447	0,0011	0,0018	0,0017
90	67	0,3554	0,6968	0,6253	0,0003	0,0006	0,0007
100	71	0,6247	0,9857	0,8883	0,0001	0,0002	0,0002
PO							
10	158	-0,8220	-0,4390	-0,4902	-0,0020	-0,0014	-0,0017
20	75	0,1507	-0,5330	-0,5687	-0,0004	-0,0027	-0,0027
30	82	0,0870	-0,6164	-0,6502	-0,0003	-0,0036	-0,0034
40	41	0,5044	-0,4648	-0,4224	0,0015	-0,0038	-0,0031
50	45	0,4882	-0,3813	-0,2539	0,0016	-0,0020	-0,0014
60	40	0,6008	0,0454	0,1772	0,0026	0,0005	0,0008
70	34	0,7226	0,2601	0,3622	0,0030	0,0015	0,0019

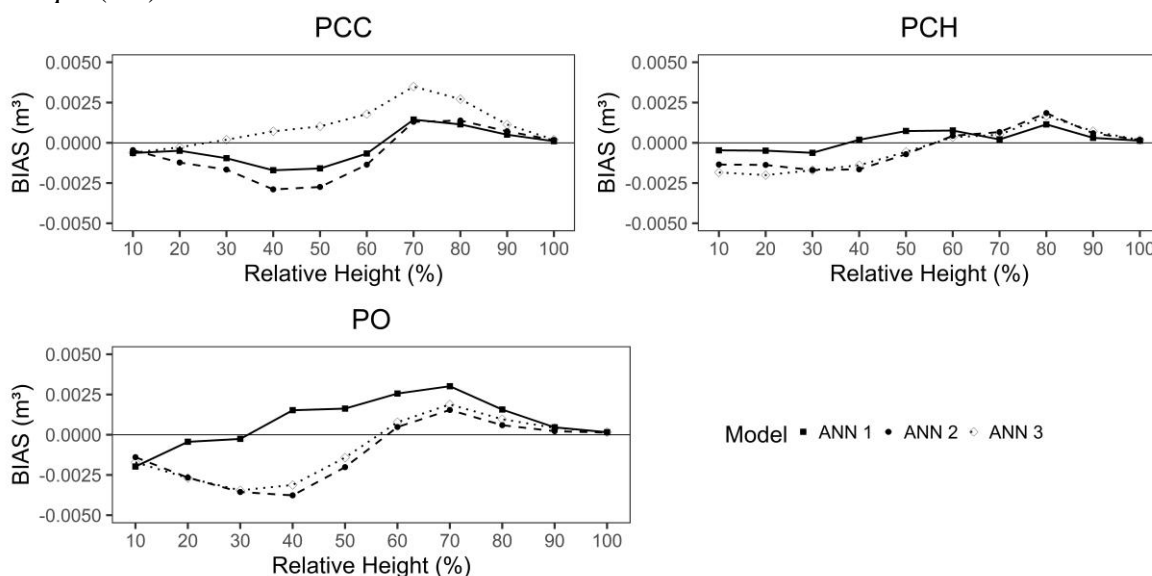
80	52	0,7141	0,2165	0,4010	0,0016	0,0006	0,0010
90	42	0,5616	0,1803	0,4968	0,0005	0,0002	0,0004
100	46	0,5687	0,3690	0,5736	0,0002	0,0001	0,0002

Figura 32 – Comportamento dos valores de BIAS em porções relativas resultantes das predições de diâmetro realizadas pelas melhores RNA avaliadas para a base de teste para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).



Fonte: Do autor (2024).

Figura 33 – Comportamento dos valores de BIAS em porções relativas resultantes das predições de volumes seccionais realizadas pelas melhores RNA avaliadas para a base de teste para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).



Fonte: Do autor (2024).

Para a espécie PCH, as predições de diâmetro da RNA 1 resultaram em valores de BIAS menores para quase toda porção do fuste, com exceção da porção de 60%. Já para as predições de volume, os valores de BIAS só não foram menores para a porção de 50 a 60%. Ambos os resultados apontam a RNA 1 como a mais adequada.

Considerando a espécie PO, as predições geradas pela RNA 2 resultaram em valores menores de BIAS apenas nas porções de 10 e 60 a 100% para as predições de diâmetro e volume. No entanto, por ter resultado nas melhores predições para as espécies PCC e PCH, a RNA 1 foi selecionada para posterior comparação com os modelos de efeitos mistos.

O coeficientes obtidos a partir do treinamento da RNA 1 podem ser observados na tabela 14.

Tabela 14 – Coeficientes ajustados para a RNA 1.

Hidden Layer	Node	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
1	Neuron 1	0,0349	0,0967	-0,2073	-15,9580	1,2035	-1,8037	-1,9580	5,3630
	Neuron 2	-0,7387	0,4191	0,3145	1,0779	-0,5704	-0,7613	-0,7824	1,0969
	Neuron 3	-1,1788	0,4806	0,1975	0,9681	0,7107	-0,0729	-0,8243	-0,7598
	Neuron 4	-1,7186	-0,4378	0,1002	0,0846	1,8080	-0,0139	-1,1736	0,6624
	Neuron 5	0,2065	-0,6553	0,5558	0,1250	0,7489	0,2760	0,3876	-0,4454
	Neuron 6	1,0871	-0,8230	-0,0365	1,0184	-0,2997	0,3111	0,2840	-0,7056
	Neuron 7	0,5251	0,5846	0,0793	-0,7560	-0,0819	0,6367	-0,7639	-0,6637
2	Neuron 1	0,0013	0,0342	0,9974	-1,0306	-1,3955	0,3493	0,2699	-0,6708
	Neuron 2	0,0595	0,2806	1,7065	-0,5717	1,3034	-0,7661	2,3290	-0,0239
	Neuron 3	0,1245	-0,6271	-1,3315	1,7314	0,5745	-0,3037	-0,6327	-0,9446
	Neuron 4	1,9945	-0,3303	-0,8496	1,4017	-1,1549	0,5233	-0,3288	1,5568
	Neuron 5	1,2970	-0,6656	-1,1041	1,9495	2,4440	0,6512	-0,3761	2,2832
	Neuron 6	-0,1519	0,0936	-3,2754	-1,5385	-2,5183	3,1229	3,3293	3,3545
	Neuron 7	-0,3503	0,1220	-0,2100	-0,5549	0,7557	-0,4574	1,5636	0,0269
	Output	0,2074	-0,7842	-0,5140	-1,1336	-0,3375	-0,2526	0,2675	-1,0330

3.4 Avaliação geral

Os modelos mistos e as RNAs selecionadas foram comparados de acordo sua capacidade preditiva para diâmetro e volume e os valores de rank para cada métrica pode ser verificado na tabela 15. Os modelos mistos foram superiores às RNAs para as espécies PCC e PO

Tabela 15. Ranks para os desempenhos das alternativas de calibração selecionadas comparadas à melhor RNA selecionada, considerando as predições de diâmetros e volumes totais para a base de teste para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).

Model	Diameter ranks				Total volume ranks				Summed ranks	Overall rank
	RMSE	MAPE	EF	BIAS	RMSE	MAPE	EF	BIAS		
PCC										
Fixed effects	1	1	1	2	1	1	1	1	9	1
ANN 1	2	2	2	1	2	2	2	2	15	2
PCH										
Calibrated (DBH and D5.2)	2	1	2	2	2	2	2	2	15	2
ANN 1	1	2	1	1	1	1	1	1	9	1
PO										
Calibrated (D5.2)	2	1	2	1	1	1	1	2	11	1
ANN 1	1	2	1	2	2	2	2	1	13	2

Em uma comparação conjunta, envolvendo todas as espécies (Tabela 16), o desempenho dos modelos mistos e da RNA foram semelhantes para a predição de diâmetro, enquanto que, para a predição de volume, os modelos mistos foram superiores.

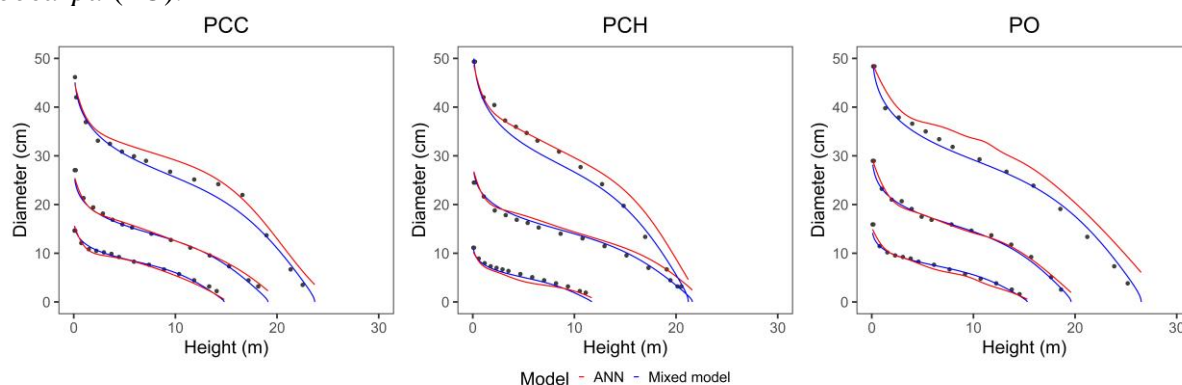
Tabela 16 – Combinação dos ranks para os desempenhos das alternativas de calibração selecionadas comparadas à RNA selecionada, considerando as predições de diâmetros e volumes totais para a base de teste.

Model	RMSE	MAPE	EF	BIAS	Summed ranks	Overall rank
Diameter ranks						
Mixed model	5	3	5	5	18	1
ANN	4	6	4	4	18	1
Total volume ranks						
Mixed model	4	4	4	5	17	1
ANN	5	5	5	4	19	2

Após a seleção dos modelos mistos e da RNA que melhor representassem a forma do fuste das três espécies, gráficos com valores das árvores menor, média e maior da base de teste

foram gerados para verificar a capacidade preditiva dos modelos (Figura 34). Para a espécie PCC, as duas técnicas resultaram em curvas semelhantes para a menor árvore, um comportamento semelhante foi observado para a árvore média, porém houve uma superestimação da RNA e subestimação do modelo misto no topo do fuste. Além disso, para a árvore maior a RNA superestimou os diâmetros na porção média do fuste enquanto que o modelo misto subestimou.

Figura 34 – Comportamento das curvas de afilamento ajustas pelas técnicas de modelagem mista e RNA aplicadas a árvores de diferentes tamanhos (pequeno, médio e grande) para as espécies *Pinus caribaea* var. *caribaea* (PCC), *Pinus caribaea* var. *hondurensis* (PCH) e *Pinus oocarpa* (PO).



Fonte: Do autor (2024).

Considerando a espécie PCH, houve uma subestimação dos diâmetros estimados pela RNA na porção média do fuste para a menor árvore, para a árvore média houve uma superestimação dos valores do topo do fuste. Já para a maior árvore, o modelo misto subestimou os diâmetros na porção baixa do fuste, enquanto que a RNA superestimou os valores na parte alta do fuste.

Para a espécie PO, o comportamento das técnicas para a árvore menor foi semelhante, sem tendências para ambas. Para a árvore média o comportamento também foi semelhante, resultando em previsões precisas. Por fim, para a árvore maior, o modelo misto descreveu o perfil da árvore sem tendência evidente, enquanto que a RNA superestimou os diâmetros em quase toda a porção do fuste. Por fim, percebe-se que, de forma geral, houve uma melhor representação da forma dos fustes pela modelagem mista, onde as curvas foram mais suaves e próximas das medidas reais para os três tamanhos de árvores.

4 DISCUSSÃO

A compreensão da forma do fuste é uma ferramenta importante, uma vez que permite definir o sortimento de madeira das florestas e, conseqüentemente, o seu valor econômico (Kohler *et al.*, 2016). Por meio do ajuste de equações matemáticas é possível descrever o perfil do fuste das árvores, porém, a seleção de um modelo adequado para diferentes espécies e condições pode ser uma tarefa complexa devido às variações na forma dos fustes e as diversas alternativas disponíveis (Li; Weiskittel, 2010). A fim contribuir com informações sobre técnicas para a modelagem do perfil do fuste de árvores comerciais no Brasil, o presente estudo desenvolveu e analisou funções baseadas em técnicas de modelagem mista e redes neurais artificiais para povoamentos de *Pinus caribaea* var. *caribaea*, *Pinus caribaea* var. *hondurensis* e *Pinus oocarpa* localizados na região sudeste do país.

Dentre os diversos tipos de equações de regressão testadas, os resultados do presente estudo indicaram a escolha de uma equação de forma variável (Kozak, 2004). Na literatura, pode-se encontrar a aplicação eficiente dessas equações para diversas condições e espécies (Kozak 1988; Pérez *et al.* 1990; Muhairwe 1999; Sharma; Zhang. 2004; Oliveira *et al.*, 2022; Hansen *et al.*, 2023), uma vez que são capazes de descrever a forma do fuste a partir de mudanças no expoente ou variável da base para o topo, representando as formas do neilóide, parabolóide, cone e diversas formas intermediárias (Kozak 1997). Além disso, utilizando essa técnica, assume-se que a forma do fuste muda continuamente ao longo do fuste (Lee *et al.*, 2003). No presente estudo, assim como em estudos anteriores ao redor do mundo (Rojo *et al.*, 2005; Corral-Rivas *et al.*, 2007; Li; Weiskittel, 2010; Heiðarsson; Pukkala, 2011), o modelo de Kozak (2004) foi capaz de resultar em menores erros e maior eficiência das predições para espécies do gênero *Pinus*.

Em um levantamento realizado por Terra *et al.* (2020), dos trabalhos realizados em plantios brasileiros de *Pinus*, os modelos predominantemente utilizados são os modelos de Hradetzky (1976) e Schöepfer (1966), evidenciando a necessidade de estudos considerando outros tipos de modelos, como as equações de forma variável. Analisando pesquisas ainda mais recentes, é possível verificar estudos aplicando o modelo de Kozak (2004) para plantios florestais no Brasil (Andrade *et al.*, 2021; Andrade *et al.*, 2022; Bernardi *et al.*, 2021; Rocha *et al.*, 2022; Lacerda *et al.*, 2022; Santos *et al.*, 2023), porém, poucos estudos envolvem espécies do gênero *Pinus*, como o trabalho realizado por Stefanello *et al.* (2021), onde o modelo proporcionou as predições mais acuradas de volume comercial para plantios comerciais de *Pinus taeda* na região sul do país.

Nem todos os parâmetros dos modelos foram significativos, sendo assim removidos e as equações foram reajustadas para preservar a parcimônia dos modelos. Posteriormente, parâmetros de efeito aleatório a nível de árvore foram incluídos em dois parâmetros do modelo, para cada espécie. Embora seja possível expandir todos os parâmetros com o efeito aleatório, isso pode dificultar a convergência durante o ajuste. Nos trabalhos de Arias-Rodil *et al.* (2015) e Bouriaud *et al.* (2019), onde consideraram a expansão de até dois e três parâmetros, respectivamente, também foram reportados problemas de convergência.

A inclusão do efeito aleatório proporcionou uma melhora significativa considerando as três espécies de *Pinus*. Sharma (2020), avaliando a aplicação de equações de forma variável para modelar o afilamento do fuste em plantios *Pinus resinosa* no Canadá, também constatou que a adição de efeitos aleatórios a nível de árvore resultou em uma melhora significativa das métricas estatísticas. Além disso, Schröder *et al.*, (2015), estudando plantios de *Pinus elliottii* Engelm., no sul do Brasil, também reportaram melhora significativa dos parâmetros estatísticos de modelos de afilamento por meio de modelagem mista, porém os autores consideraram o efeito aleatório a nível de classe diamétrica. Já, Santos *et al.*, (2021) também aplicando modelagem mista em plantios de *Pinus taeda*, com 17 anos de idade, constataram um aumento significativo na precisão dos modelos, utilizando o nível árvore como efeito aleatório, incluindo a comparação com outros níveis como sítio e classe diamétrica. Esses resultados, juntamente com os encontrados no presente estudo, enaltecem a relevância da utilização do efeito aleatório a nível individual de árvore que, por sua vez, é capaz de conferir generalidade e especificidade aos modelos, simultaneamente, uma vez que atributos individuais como forma e afilamento são essenciais para predições de volume precisas ao longo do fuste. Apesar da existência de trabalhos envolvendo modelagem mista para o gênero *Pinus* no Brasil, o presente trabalho é o primeiro estudo a desenvolver equações de forma variável com adição de efeitos aleatórios a nível de árvore para as espécies *Pinus caribaea* var. *caribaea*, *Pinus caribaea* var. *hondurensis* e *Pinus oocarpa*, na região sudeste do país, sendo uma contribuição extremamente relevante para a modelagem do fuste de pinus.

Em seguida, considerou-se a heterocedasticidade e autocorrelação residual. Essas propriedades devem ser consideradas na modelagem do afilamento do fuste e, de acordo Leites e Robinson (2004) e Sharma e Parton (2009), a modelagem mista, por meio da inclusão de efeitos aleatório, pode remover a heterocedasticidade e a autocorrelação dos resíduos do modelo. No entanto, em contraste essas afirmações, no presente estudo, a inclusão dos efeitos aleatórios não foi suficiente para eliminar a autocorrelação dos resíduos, de modo que, a inclusão da estrutura de erro autoregressiva contínua de primeira ordem CAR(1) aos modelos

de Kozak (2004), promoveu uma melhora significativa das métricas estatísticas (Tabela 7). Estudos recentes obtiveram resultados semelhantes, evidenciando que em certos casos a autocorrelação pode ser reduzida, mas não completamente removida apenas com a adição dos parâmetros aleatórios (Liu *et al.*, 2020; Özçelik; Alkan, 2020). Essa informação deve ser enfatizada, uma vez que, ainda que os parâmetros estimados pelos mínimos quadrados permaneçam não viesados e com distribuição normal, há uma redução de sua eficiência na presença de autocorrelação (Gujarati; Porter, 2008). Além disso, a modelagem mista tornou a variância dos resíduos ainda mais homogênea.

Utilizando a modelagem mista, parâmetros de efeito aleatório podem ser preditos a partir de, no mínimo, uma medida de diâmetro. Esse processo, chamado de calibração, permite a obtenção de curvas calibradas individualmente para novas árvores. Com esse objetivo, foram avaliadas três alternativas de calibração que utilizaram o DAP, $D_{5,2}$ e DAP e $D_{5,2}$ de cada árvore da base de teste como medidas adicionais. Além disso, a alternativa sem calibração também foi avaliada, uma vez que as equações de efeitos mistos, além das curvas específicas por indivíduos, podem fornecer uma curva média, a partir apenas dos parâmetros de efeito fixo (Westfall, 2016). No presente estudo, ausência de calibração foi a alternativa mais recomendada para a espécie *Pinus caribaea* var. *hondurensis*. Arias-Rodil *et al.* (2015), em contraste com os resultados do presente estudo, concluíram que a ausência de dados de calibração limitaria a aplicabilidade da equação mista ajustada para *Pinus radiata* na Espanha.

Ao utilizar a curva média para as predições, esta está condicionada aos efeitos aleatório de modo que é capaz de capturar parcialmente a dependência entre observações, baseada na autocorrelação individual em sua estrutura de covariância (Verbeke *et al.*, 2014; Lansanova *et al.*, 2020).

Para *Pinus caribaea* var. *hondurensis* e *Pinus oocarpa*, as medidas de diâmetro adicionais para calibração dos modelos foi recomendada, sendo $D_{5,2}$ e DAP e $D_{5,2}$, respectivamente. Trincado e Burkhart (2006), estudando modelos de afilamento para *Pinus taeda* L., constataram que a inclusão de efeitos aleatórios e calibração para novas árvores, utilizando dois diâmetros, resultou em medidas mais precisas. Já os resultados encontrados por Lansanova *et al.* (2020) indicaram o DAP como medida adicional para a calibração de modelos mistos de afilamento para plantios de *Tectona grandis*. O autor também destaca para o fato de que, com a utilização do DAP como medida adicional, não há a necessidade de coletas adicionais de diâmetros ao longo do fuste durante as operações de inventário. No presente estudo, modelos com boa capacidade de predição de $D_{5,2}$ foram desenvolvidos individualmente por espécie, resultando em bons resultados na calibração e tornando ainda dispensável a coleta

adicional de diâmetros no campo. Fatores como esse são indispensáveis para a redução de tempo e custo na coleta de dados, tornando a técnica de calibração viável.

Para *Pinus caribaea* var. *caribaea*, houve uma maior precisão das predições nas porções mais baixas do fuste, tanto de diâmetro como de volume. Téó *et al.* (2018), aplicando o modelo de Schöpfer (1966) para *Pinus taeda*, encontraram um padrão semelhante. Esse comportamento pode ter seu impacto negativo reduzido, uma vez que, as porções mais baixa são que as constituem maior quantidade e são a matéria-prima dos produtos mais nobres.

Hussain *et al.* (2020), aplicando o mesmo modelo com efeitos aleatórios e calibração utilizando os diâmetros a 5,3 m para as espécies *Larix gmelinii* Rupr. e *Abies nephrolepis* Maxim, encontraram maiores erros para os diâmetros na porção mais alta do fuste. Estes resultados são semelhantes ao encontrado no presente estudo para a espécie *Pinus caribaea* var. *hondurensis*. Como consequência, as predições dos volumes nas porções mais baixas do fuste também resultaram em maiores erros. He *et al.* 2021 ressaltam o fato de que a transformação dos valores preditos de diâmetro em predições de volume pode resultar em maior erro dessas predições. Esse comportamento pode ser atribuído à maior variação de forma que ocorre na base do fuste e isso pode ter um maior impacto negativo nas maiores árvores, pois estas normalmente carregam um maior valor econômico. Dessa forma, é recomendável que se busque estratégias para aumentar a precisão do volume, principalmente na base dos fustes.

Considerando as RNA, resultados satisfatórios foram encontrados para configurações contendo duas camadas ocultas, com sete neurônios cada. Contrastando com resultados como o de Bonete *et al.* (2019), em que, aplicando RNA para explicar o afilamento do fuste de plantios de *Pinus taeda* L. não desbastados e com menor amplitude de idade, encontraram resultados satisfatório a partir da utilização de quatro camadas ocultas. Os autores também encontraram valores de RMSE e MAPE maiores, quando comparados com os valores obtidos de forma simultânea para três espécies do presente estudo. Diferentemente dos autores citados, no presente trabalho, a transformação $\frac{H_i - h_{ij}}{H_i - 1,3}$ foi utilizada como entrada das RNA e, mesmo não tendo sido realizados testes sobre o efeito da inclusão dessa variável, isso pode sugerir que esta seja considerada para futuros estudos sobre o afilamento de *Pinus*. Também vale ressaltar que, além dos valores mais baixos de RMSE, no presente estudo, foi incluída a essa medida uma penalização relacionada ao número de parâmetros presentes nas redes.

Seki (2023), estudando plantios de *Pinus sylvestris* L. na Turquia, encontrou valores de RMSE maiores, porém melhores resultados para EF e BIAS, onde as métricas EF e RMSE consideraram o número de parâmetros. Diferentemente do presente estudo, o autor utilizou

RNA com apenas uma camada oculta com cinco neurônios, mas o desempenho das RNA foi inferior aos dos modelos de regressão avaliados em seu estudo. E, além disso, os resíduos alcançaram uma grande amplitude e variância menos homogênea.

De forma geral, com o avanço da aplicação das RNA, diversas configurações vêm sendo disponibilizadas para aplicação na modelagem de variáveis florestais (Costa Filho *et al.*, 2019). No entanto, a seleção da configuração adequada e das variáveis de entrada de uma RNA é um processo que pode demandar grande tempo e diversos testes, uma vez que é necessário escolher as variáveis preditoras, quantidade de camadas ocultas, número de neurônios por camada, função de ativação para as camadas ocultas e de saída, algoritmo de aprendizagem, dentre outras. Dessa forma, estudos como esse são essenciais, pois podem contribuir para a escolha correta de configurações dentre as diversas possibilidades existentes a fim de se obter resultados mais precisos e confiáveis.

Uma das grandes vantagens de utilizar os modelos mistos e as RNA está relacionada ao fato de que, além das predições acuradas, ambas as técnicas permitem a aplicação em dados que não precisam necessariamente obedecer os pressupostos estatísticos exigidos na regressão clássica (Ozçelik *et al.*, 2011; Schikowski *et al.*, 2015), a exemplo dos dados de afilamento que são medidas repetidas e correlacionadas para o mesmo indivíduo. Além disso, essas técnicas permitem modelar dados complexos provenientes de diferentes locais e material genético, uma vez que são capazes de considerar os diversos níveis presentes nos dados.

Predições ainda mais acuradas poderiam ser alcançadas por meio do fornecimento de atributos a nível de parcela, como idade, área basal, diâmetro médio quadrático e altura dominante que, infelizmente, não estavam presentes na base de dados em estudo. A utilização dessas variáveis como preditoras nos modelos se torna ainda mais viável por serem variáveis comumente obtidas nas operações de inventário das empresas. Outras variáveis a nível individual como informações sobre a copa ou diâmetros adicionais podem ser aplicadas no estudo do comportamento da forma do fuste para obtenção de informações mais precisas mas, do ponto de vista prático, sua inclusão nas operações rotineiras das empresas poderia resultar em um aumento de custo que anularia a viabilidade dessa prática.

5 CONCLUSÃO

Modelos de efeito mistos e redes neurais artificiais foram eficientes na predição do diâmetro ao longo do fuste e volume total para espécies do gênero *Pinus*.

A equação de forma variável de Kozak (2004) foi a alternativa mais adequada dentre as técnicas de regressão avaliadas e, com a adição de parâmetros de efeito aleatório e estrutura de erro autoregressiva contínua de primeira ordem houve melhora significativa, tornando os modelos mais coerentes com a natureza dos dados. Após a calibração das equações, os modelos resultaram em predições de diâmetro ao longo do fuste e volume total estatisticamente iguais aos valores observados para as espécies *Pinus caribaea* var. *caribaea* e *Pinus oocarpa*.

Os diâmetros à 1,3 m e 5,2 m do solo são recomendados para a calibração da equação mista de afilamento para *Pinus caribaea* var. *caribaea*, enquanto que, a calibração com apenas diâmetros à 1,3 m é recomendada para *Pinus oocarpa*. Além disso, recomenda-se a utilização da equação de predição do diâmetro a 5,2 m para realização da calibração, dispensando-se assim a necessidade de medidas adicionais de diâmetros e tornando a aplicação desta metodologia viável.

As redes neurais artificiais promoveram resultados satisfatórios de diâmetro ao longo do fuste e volume total para *Pinus caribaea* var. *caribaea* (PCC) e *Pinus caribaea* var. *hondurensis*, simultaneamente, com valores preditos não diferindo estatisticamente dos valores observados. A melhor configuração foi composta por duas camadas ocultas, com sete neurônios cada e funções de ativação tangente hiperbólica e linear nas camadas ocultas e de saída, respectivamente.

De forma geral, considerando as predições de diâmetro e volume total para três espécies em estudo, a modelagem mista resultou nos melhores resultados, não sendo capaz apenas de prever diâmetros significamente iguais aos valores observados para a espécie *Pinus caribaea* var. *hondurensis*, mas mesmo assim os valores de volume total foram satisfatórios.

Embora as equações mais simples de afilamento ofereçam mais praticidade e sejam de mais simples entendimento, a acurácia têm sido superada por modelos mais complexos e, uma vez que o poder computacional não é mais um fator limitante, seu uso em diversas ocasiões se torna injustificável.

O comportamento dos resíduos ao longo do fuste e a avaliação da propagação do erro para as predições volumétricas devem ser considerados na escolha do modelo de afilamento para conduzir a uma seleção mais adequada.

Recomenda-se a adição de outros atributos a nível de parcela, a fim de melhorar a representação do afilamento dos fustes das árvores, desde que a coleta destas não comprometam a viabilidade das operações de inventário florestal.

REFERÊNCIAS

- AKAIKE, H. Maximum likelihood identification of Gaussian autoregressive moving average models. **Biometrika**, v. 60, n. 2, p. 255-265, 1973.
- AMARIOAREI, A.; PAUN, M.; STRIMBU, B. Development of nonlinear parsimonious forest models using efficient expansion of the Taylor series: Applications to site productivity and taper. **Forests**, v. 11, n. 4, p. 458, 2020.
- AMATEIS, R. L.; BURKHART, H. E. Cubic-foot volume equations for loblolly pine trees in cutover, site-prepared plantations. **Southern Journal of Applied Forestry**, v. 11, n. 4, p. 190-192, 1987.
- ANDRADE, V. C. L.; FREITAS, B. C.; TERRA, D. L. C. V. Equações de perfil do tronco avaliados para *Corymbia citriodora* Hill & Johnson. **Agrarian**, v. 14, n. 51, p. 82-94, 2021.
- ANDRADE, V. C. L.; TERRA, D. L. C. V.; CARVALHO, S. P. C. Análise de regressão do perfil do fuste de *Corymbia citriodora* formado em área de Cerrado. **Ciência Florestal**, v. 32, p. 1500-1527, 2022.
- ARIAS-RODIL, M. *et al.* Fitting and calibrating a multilevel mixed-effects stem taper model for maritime pine in NW Spain. **PloS one**, v. 10, n. 12, p. e0143521, 2015.
- ARIAS-RODIL, M. *et al.* Modelling and localizing a stem taper function for *Pinus radiata* in Spain. **Canadian Journal of Forest Research**, v. 45, n. 6, p. 647-658, 2015.
- BECKER, R. A.; CHAMBERS, J. M.; WILKS, A. R. The New S Language. Wadsworth & Brooks/Cole. **Computer Science Series, Pacific Grove, CA**, 1988.
- BERNARDI, L. K. *et al.* Diferentes modelos para o afilamento do tronco de *Eucalyptus* sp. para o cenário florestal brasileiro. **Ciência Florestal**, v. 31, p. 1364-1382, 2021.
- BI, H. Trigonometric variable-form taper equations for Australian eucalypts. **Forest Science**, v. 46, n. 3, p. 397-409, 2000.
- BISHOP, C. M. *et al.* **Neural networks for pattern recognition**. Oxford university press, 1995.
- BONETE, I. P. *et al.* Artificial neural networks and mixed-effects modeling to describe the stem profile of *Pinus taeda* L. **Floresta**, v. 50, n. 1, p. 1123-1132, 2019.
- BRUCE, D.; CURTIS, R. O.; VANCOEVERING, C. Development of a system of taper and volume tables for red alder. **Forest science**, v. 14, n. 3, p. 339-350, 1968.
- CLARK, A.; SOUTER, R. A.; SCHLAEGEL, B. E. **Stem profile equations for southern tree species**. USDA Forest Service, Research Paper SE-282, 1991.
- CORRAL-RIVAS, J. J. *et al.* A merchantable volume system for major pine species in El Salto, Durango (Mexico). **Forest Ecology and Management**, v. 238, n. 1-3, p. 118-129, 2007.

COSTA-FILHO, S. V. S. *et al.* Configuração de algoritmos de aprendizado de máquina na modelagem florestal: um estudo de caso na modelagem da relação hipsométrica. **Ciência Florestal**, v. 29, p. 1501-1515, 2019.

DEMAERSCHALK, J. P. Converting volume equations to compatible taper equations. **Forest Science**, v. 18, n. 3, p. 241-245, 1972.

DE-MIGUEL, S. *et al.* Evaluating marginal and conditional predictions of taper models in the absence of calibration data. **Canadian Journal of Forest Research**, v. 42, n. 7, p. 1383-1394, 2012.

FANG, Z.; BAILEY, R. L. Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. **Forest science**, v. 47, n. 3, p. 287-300, 2001.

FANG, Z.; BORDERS, B. E.; BAILEY, R. L. Compatible volume-taper models for loblolly and slash pine based on a system with segmented-stem form factors. **Forest Science**, v. 46, n. 1, p. 1-12, 2000.

FLORIANO, E. P. *et al.* Fitting and selecting traditional models for tree's height time series data. **Ciência Florestal**, v. 16, n. 2, p. 177-199, 2006.

FONWEBAN, J. *et al.* Taper functions for Scots pine (*Pinus sylvestris* L.) and Sitka spruce (*Picea sitchensis* (Bong.) Carr.) in northern Britain. **Forestry**, v. 84, n. 1, p. 49-60, 2011.

GUIMARÃES, D. P.; LEITE, H. G. Um novo modelo para descrever o perfil do tronco. **Revista Árvore**, v. 16, n. 2, p. 170-180, 1992.

GUJARATI, D. N.; PORTER, D. C. **Basic Econometrics**. McGraw-Hill: New York. 2008.

GÜNTHER, F.; FRITSCH, S. Neuralnet: training of neural networks. **The R Journal**, v. 2, n. 1, p. 30, 2010.

HANSEN, E. *et al.* Taper, volume, and bark thickness models for spruce, pine, and birch in Norway. **Scandinavian Journal of Forest Research**, v. 38, n. 6, p. 413-428, 2023.

HE, P. *et al.* Evaluation of four regression techniques for stem taper modeling of Dahurian larch (*Larix gmelinii*) in Northeastern China. **Forest Ecology and Management**, v. 494, p. 119336, 2021.

HEIDARSSON, L.; PUKKALA, T. Taper functions for lodgepole pine (*Pinus contorta*) and Siberian larch (*Larix sibirica*) in Iceland. **Icelandic Agricultural Sciences**, n. 24, p. 3-11, 2011.

HIRIGOYEN, A. *et al.* Modelling taper and stem volume considering stand density in *Eucalyptus grandis* and *Eucalyptus dunnii*. **iForest-Biogeosciences and Forestry**, v. 14, n. 2, p. 127, 2021.

HOLLANDER, M.; WOLFE, D. A. **Nonparametric Statistical Methods**. John Wiley and Sons: New York, NY, USA, 1973; p. 503.

HORLE, D. C. *et al.* Modelagem não linear do perfil longitudinal de fustes de *Pinus oocarpa*. **Cerne**, v. 16, p. 177-184, 2010.

HRADETZKY, J. **Analyse und interpretation statistischer abränger keiten.** (Biometrische Beiträge zu aktuellen forschungs projekten). Baden: Württemberg Mitteilungen der FVA, 146p, 1976.

HUSSAIN, A. *et al.* Stem taper equations for three major conifer species of Northeast China. **Scandinavian Journal of Forest Research**, v. 35, n. 8, p. 562-576, 2020.

JAYALAKSHMI, T.; SANTHAKUMARAN, A. Statistical normalization and back propagation for classification. **International Journal of Computer Theory and Engineering**, v. 3, n. 1, p. 1793-8201, 2011.

JIANG, L.; BROOKS, J. R.; WANG, J. Compatible taper and volume equations for yellow-poplar in West Virginia. **Forest ecology and management**, v. 213, n. 1-3, p. 399-409, 2005.

JÚNIOR, I. S. T. *et al.* Machine learning and regression models to predict multiple tree stem volumes for teak. **Southern Forests**, v. 83, n. 4, p. 01-09, 2021.

KOHLER, S. V. *et al.* Evolution of tree stem taper in *Pinus taeda* stands. **Ciência Rural**, v. 46, p. 1185-1191, 2016.

KOZAK A. My last words on taper equations. **The Forestry Chronicle**. v. 80, n. 4, p. 507-515, 2004.

KOZAK, A. A variable-exponent taper equation. **Canadian Journal of Forest Research**, v. 18, n. 11, p. 1363-1368, 1988.

KOZAK, A. Effects of multicollinearity and autocorrelation on the variable-exponent taper functions. **Canadian Journal of Forest Research**, v. 27, n. 5, p. 619-629, 1997.

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, 2022.

LANSSANOVA, R. L. *et al.* Calibration of a mixed-effect stem taper model for *Tectona grandis*. **Journal of Tropical Forest Science**, v. 32, n. 4, p. 341-348, 2020.

LANTZ, B. **Machine learning with R: expert techniques for predictive modeling.** Packt publishing ltd, 2019.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436-444, 2015.

LEE, W. K. *et al.* Modeling stem profiles for *Pinus densiflora* in Korea. **Forest Ecology and Management**, v. 172, n. 1, p. 69-77, 2003.

LEITES, L. P.; ROBINSON, A. P. Improving taper equations of loblolly pine with crown dimensions in a mixed-effects modeling framework. **Forest Science**, v. 50, n. 2, p. 204-212, 2004.

LI, R. *et al.* Regional stem taper equations for eleven conifer species in the Acadian region of North America: development and assessment. **Northern Journal of Applied Forestry**, v. 29, n. 1, p. 5-14, 2012.

LI, R.; WEISKITTEL, A. R. Comparison of model forms for estimating stem taper and volume in the primary conifer species of the North American Acadian Region. **Annals of Forest Science**, v. 67, n. 3, p. 302, 2010.

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, 2020.

LIU, Y. *et al.* Tree profile equations are significantly improved when adding tree age and stocking degree: an example for *Larix gmelinii* in the Greater Khingan Mountains of Inner Mongolia, northeast China. **European Journal of Forest Research**, v. 139, p. 443-458, 2020.

MAX, T. A.; BURKHART, H. E. Segmented polynomial regression applied to taper equations. **Forest Science**, v. 22, n. 3, p. 283-289, 1976.

MUHAIRWE, C. K. Taper equations for *Eucalyptus pilularis* and *Eucalyptus grandis* for the north coast in New South Wales, Australia. **Forest Ecology and Management**, v. 113, n. 2-3, p. 251-269, 1999.

OGANA, F. N. A mixed-effects height-diameter model for *Gmelina arborea* Roxb stands in Southwest Nigeria. **Journal of Forest Research**, v. 27, n. 1, p. 1-7, 2021.

OGANA, F. N.; ERCANLI, I. Modelling height-diameter relationships in complex tropical rain forest ecosystems using deep learning algorithm. **Journal of Forestry Research**, v. 33, n. 3, p. 883-898, 2022.

OLIVEIRA, X. M. *et al.* Impact of the inclusion of the sociological position and upper-stem diameter in the taper modeling of *Pinus taeda* trees, subjected to different thinning regimes. **Trees, Forests and People**, v. 8, p. 100251, 2022.

ORMEROD, D. W. A simple bole model. **The Forestry Chronicle**, v. 49, n. 3, p. 136-138, 1973.

ÖZÇELİK, R. *et al.* Estimating tree bole volume using artificial neural network models for four species in Turkey. **Journal of environmental management**, v. 91, n. 3, p. 742-753, 2010.

ÖZÇELİK, R.; ALKAN, O. Fitting and calibrating a mixed-effects segmented taper model for Brutian pine. **Cerne**, v. 26, p. 464-473, 2020.

ÖZÇELİK, R.; BROOKS, J. R.; JIANG, L. Modeling stem profile of Lebanon cedar, Brutian pine, and Cilicica fir in Southern Turkey using nonlinear mixed-effects models. **European Journal of Forest Research**, v. 130, p. 613-621, 2011.

PARRESOL, B. R.; HOTVEDT, J. E.; CAO, Q. V. A volume and taper prediction system for bald cypress. **Canadian Journal of Forest Research**, v. 17, n. 3, p. 250-259, 1987.

PEREZ, D. N.; BURKHART, H. E.; STIFF, C. T. A variable-form taper function for *Pinus oocarpa* Schiede in Central Honduras. **Forest Science**, v. 36, n. 1, p. 186-191, 1990.

PINHEIRO, J. C.; BATES, D. M. **Mixed-effects models in S and S-PLUS**. New York: Springer; 2000.

POUDEL, K. P.; CAO, Q. V. Evaluation of methods to predict Weibull parameters for characterizing diameter distributions. **Forest Science**, v. 59, n. 2, p. 243-252, 2013.

RAPTIS, D. I. *et al.* Mixed-effects height–diameter models for black pine (*Pinus nigra* Arn.) forest management. **Trees**, v. 35, n. 4, p. 1167-1183, 2021.

RESENDE, M. D. V.; SILVA, F. F.; AZEVEDO, C. F. **Estatística matemática, biométrica e computacional: Modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão aleatória, seleção genômica, QTL-GWAS, estatística espacial e temporal, competição, sobrevivência**. Viçosa: Suprema gráfica e Editora Ltda, 2014, 882p.

RIEDMILLER, M. Advanced supervised learning in multi-layer perceptrons—from backpropagation to adaptive learning algorithms. **Computer Standards & Interfaces**, v. 16, n. 3, p. 265-278, 1994.

ROCHA, K. J. *et al.* Form and Volume of the Stem of *Tectona grandis* Lf in the Central-WESTERN Region of Brazil. **Forests**, v. 13, n. 11, p. 1818, 2022.

ROJO, A. *et al.* Stem taper functions for maritime pine (*Pinus pinaster* Ait.) in Galicia (Northwestern Spain). **European Journal of Forest Research**, v. 124, p. 177-186, 2005.

SAHIN, A. *et al.* Estimation of tree height with machine learning techniques in coppice-originated pure sessile oak (*Quercus petraea* (Matt.) Liebl.) stands. **Scandinavian Journal of Forest Research**, v. 38, n. 1-2, p. 87-96, 2023.

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, 2018.

SALEKIN, S. *et al.* Global tree taper modelling: a review of applications, methods, functions, and their parameters. **Forests**, v. 12, n. 7, p. 913, 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, 2022.

SANTOS, G. M. *et al.* Modelagem mista generalizada para estimar afileamento do fuste de árvores de *Pinus taeda* em diferentes espaçamentos de plantio. **Advances in Forestry Science**, v. 8, n. 1, p. 1261-1269, 2021.

SANTOS, M. L. *et al.* The Effect of Age on the Evolution of the Stem Profile and Heartwood Proportion of Teak Clonal Trees in the Brazilian Amazon. **Forests**, v. 14, n. 10, p. 1962, 2023.

SCHABENBERGER, O.; PIERCE, F.J. **Contemporary Statistical Models for the Plant and Soil Sciences**; CRC Press: Boca Raton, FL, USA, 2002; p. 730.

SCHIKOWSKI, A. B. *et al.* Modeling of stem form and volume through machine learning. **Anais da Academia Brasileira de Ciências**, v. 90, p. 3389-3401, 2018.

SCHIKOWSKI, A. B.; DALLA CORTE, A. P.; SANQUETTA, C. R. Estudo da forma do fuste utilizando redes neurais artificiais e funções de afilamento. **Pesquisa Florestal Brasileira**, v. 35, n. 82, p. 119-127, 2015.

SCHÖEPFER, W. Automatisierung des massen, sorten und wertberechnung stenender waldbestände schriftenreihe Bad. **Berlin: Wurt-Forstl**, 1966.

SCHRÖDER, T. *et al.* Taper equations for *Pinus elliottii* Engelm. in southern Paraná, Brazil. **Forest Science**, v. 61, n. 2, p. 311-319, 2015.

SEKI, M. Predicting stem taper using artificial neural network and regression models for Scots pine (*Pinus sylvestris* L.) in northwestern Türkiye. **Scandinavian Journal of Forest Research**, v. 38, n. 1-2, p. 97-104, 2023.

SHARMA, M. Incorporating stand density effects in modeling the taper of red pine plantations. **Canadian Journal of Forest Research**, v. 50, n. 8, p. 751-759, 2020.

SHARMA, M.; PARTON, J. Modeling stand density effects on taper for jack pine and black spruce plantations using dimensional analysis. **Forest science**, v. 55, n. 3, p. 268-282, 2009.

SHARMA, M.; ZHANG, S. Y. Variable-exponent taper equations for jack pine, black spruce, and balsam fir in eastern Canada. **Forest ecology and management**, v. 198, n. 1-3, p. 39-53, 2004.

SOARES, F. A. A. M. N. *et al.* Recursive diameter prediction and volume calculation of eucalyptus trees using Multilayer Perceptron Networks. **Computers and electronics in agriculture**, v. 78, n. 1, p. 19-27, 2011.

STEFANELLO, F. R. *et al.* Preußner functions for volume estimation of *Pinus taeda* L. in Southern Brazil. **New Zealand Journal of Forestry Science**, v. 51, 2021.

STEFANELLO, F. R. *et al.* Preußner functions for volume estimation of *Pinus taeda* L. in Southern Brazil. **New Zealand Journal of Forestry Science**, v. 51, 2021.

TÉO, S. J. *et al.* Stem taper equation with extensive applicability to several age classes of *Pinus taeda* L. **Floresta**, v. 48, n. 4, p. 471-482, 2018.

TERRA, D. L. C. V.; ANDRADE, V. C. L.; SILVA, H. D. Histórico Sobre Modelos De Afilamento Em Plantios Brasileiros Com O Gênero *Pinus*. **Advances in Forestry Science**, v. 7, n. 4, p. 1241-1246, 2020.

TRINCADO, G.; BURKHART, H. E. A generalized approach for modeling and localizing stem profile curves. **Forest Science**, v. 52, n. 6, p. 670-682, 2006.

TRINCADO, G.; BURKHART, H. E. A generalized approach for modeling and localizing stem profile curves. **Forest Science**, v. 52, n. 6, p. 670-682, 2006.

VALENÇA, M. **Fundamentos das redes neurais: exemplos em Java**. Olinda, Pernambuco: Editora Livro Rápido, 2007, 382p.

VERBEKE, G. *et al.* The analysis of multivariate longitudinal data: a review. **Statistical methods in medical research**, v. 23, n. 1, p. 42-59, 2014.

WESTFALL, J. A. Strategies for the use of mixed-effects models in continuous forest inventories. **Environmental monitoring and assessment**, v. 188, p. 1-11, 2016.

YANG, Y. *et al.* Nonlinear mixed-effects modeling of variable-exponent taper equations for lodgepole pine in Alberta, Canada. **European Journal of Forest Research**, v. 128, p. 415-429, 2009.

ZHANG, S. *et al.* Variable-exponent taper equation based on multilevel nonlinear mixed effect for Chinese fir in China. **Forests**, v. 12, n. 2, p. 126, 2021.

TERCEIRA PARTE

1 CONSIDERAÇÕES FINAIS

Diante dos objetivos propostos, este estudo buscou avaliar, por meio de métricas estatísticas mundialmente difundidas, a precisão de modelos tradicionais de regressão, modelos mistos e suas alternativas de calibrações e Redes Neurais Artificiais (RNA) para predições de altura total, diâmetros ao longo do fuste e volume do fuste de árvores de *Pinus* spp. A análise detalhada foi capaz de fornecer *insights* valiosos acerca da eficácia de diferentes abordagens de modelagem para etapas essenciais da mensuração florestal. Os principais resultados encontrados foram:

- A equação de Ratkowsky (1990) destacou-se como a mais adequada para explicar a interação entre altura e diâmetro das árvores de *Pinus caribaea* var. *hondurensis*.
- A inclusão do diâmetro dominante e altura dominante resultou em melhorias significativas nas predições da altura total das árvores em diferentes sítios e com diferentes idades, indicando a relevância dessas variáveis na modelagem hipsométrica.
- A introdução de efeitos aleatórios e calibração por unidade amostral demonstrou grande eficácia na melhoria do desempenho do modelo a nível local. Além disso, as alternativas de calibrações mais adequadas foram identificadas, sendo necessário nove árvores distribuídas igualmente em diferentes classes diamétricas para o modelo local na forma mista e, para o modelo que já possuía variáveis a nível de parcela, a calibração não foi necessária.
- A equação de forma variável de Kozak (2004) foi a alternativa mais adequada para a predição do diâmetro ao longo do fuste e volume total de *Pinus caribaea* var. *caribaea*, *Pinus caribaea* var. *hondurensis* e *Pinus oocarpa*.
- A modelagem com efeitos aleatórios e a função de autocorrelação contínua de primeira ordem proporcionaram melhorias significativas, tornando os modelos mais coerentes com os dados e resultou em valores significamente iguais aos reais para *Pinus caribaea* var. *caribaea* e *Pinus oocarpa*.
- As Redes Neurais Artificiais demonstraram eficiência na predição de diâmetro ao longo do fuste e volume total para *Pinus caribaea* var. *caribaea* e *Pinus caribaea* var. *hondurensis*, sendo capaz de gerar predições estatisticamente

iguais aos valores observados de forma simultânea para as duas espécies por meio de variáveis *dummy*.

Os resultados encontrados nesse trabalho apontam para a necessidade de estudos mais aprofundados sobre a otimização de estratégias para a seleção da estrutura mais apropriada das redes neurais artificiais e interpretação de seus parâmetros ajustados.

Embora as equações mais simples de afilamento ofereçam mais praticidade e sejam menos complexas, suas capacidades de predição têm sido superadas por técnicas mais complexas e, com o avanço computacional e maior difusão do conhecimento, justifica-se a aplicação de técnicas mais avançadas e eficientes.

O comportamento residual em toda a extensão do fuste e a avaliação da propagação da imprecisão para as estimativas de volume devem ser analisados para a escolha de um modelo de afilamento mais adequado.

Além disso, recomenda-se a inclusão de outros atributos a nível de unidade amostral para melhorar a representação do afilamento dos fustes das árvores, sempre respeitando viabilidade operacional nas rotinas inventário florestal. Essas melhoras podem ser alcançadas a partir de tecnologias mais avançadas como sensores remotos de alta precisão.