

Artigos

Classificação da capacidade produtiva com alturas de árvores dominantes estimadas por RNA

Productive capacity classification with heights of dominant trees estimated by ANN

Marcos Vinicius Santana Leite^I 
Christian Dias Cabacinha^{II} 
Adriana Leandra de Assis^{II} 

^IVeracel Celulose, Eunápolis, BA, Brasil

^{II}Universidade Federal de Minas Gerais, Montes Claros, MG, Brasil

RESUMO

A construção de curvas de sítio a partir da modelagem da altura de árvores dominantes medidas em parcelas permanentes em diferentes idades, considerando uma idade de referência, se constitui no método mais prático e difundido no meio florestal para classificar a capacidade produtiva local. Dentro de um processo de Planejamento Florestal, o grau de assertividade das curvas geradas tem consequências quantitativas e qualitativas na alocação de recursos, de forma que a melhoria contínua dos métodos de classificação é de suma importância. Este estudo teve como objetivo propor o uso de modelos de Redes Neurais Artificiais (RNA) para estimar a altura de árvores dominantes de eucalipto, e aplicá-los na geração de curvas de sítio utilizando o método da curva guia, como uma alternativa aos modelos tradicionais de regressão não-linear, avaliando a precisão das estimativas e a estabilidade da classificação da capacidade produtiva local gerada por essas abordagens. Os dados utilizados foram provenientes das medições de 8.819 parcelas permanentes instaladas em povoamentos clonais de *Eucalyptus urophylla* × *Eucalyptus grandis*. Foram ajustados cinco modelos não lineares clássicos e as RNA foram treinadas com dois algoritmos: *Feed Forward Back Propagation Network* (FFBP) e *Cascade Forward Back Propagation Network* (CFBP). Os resultados mostraram que, de maneira geral, quando utilizada somente a idade das parcelas para estimativa da altura dominante, não houve diferença nos resultados entre as RNA treinadas com os dois algoritmos e os modelos não lineares. Contudo, com adição de novas variáveis do povoamento durante o treinamento das RNA, houve uma melhora nas estimativas das alturas dominantes e gerou uma classificação da capacidade produtiva 13% mais estável se comparada aos modelos de regressão não linear.

Palavras-chave: Inteligência artificial; Sítios florestais; *Feed Forward Back Propagation Network*; *Cascade Forward Back Propagation Network*

ABSTRACT

The generation of site curves built from modeling the height of dominant trees measured in permanent plots at different ages considering a reference age constitutes the most practical and widespread method to classify the local productive capacity. Within a Forest Planning process, the assertiveness of the generated curves has quantitative and qualitative consequences in the allocation of resources, so that the continuous improvement of classification methods is of paramount importance. This study aimed to propose models of Artificial Neural Networks (ANN) to estimate the height of dominant eucalyptus trees, and apply them in the generation of site curves using the guide curve method, as an alternative to non-linear and assess the accuracy of estimates and stability of the classification of local productive capacity generated by these approaches. The data used are from measurements of 8,819 permanent plots installed in clonal stands of *Eucalyptus urophylla* × *Eucalyptus grandis*. Five classical non-linear models were fitted and the ANN were trained with two algorithms: Feed Forward Back Propagation Network (FFBP) and Cascade Forward Back Propagation Network (CFBP). In general, when only the age of the plots was used to estimate the dominant height, there was no difference in the results between the ANN trained with the two algorithms and the non-linear models. However, with the addition of new stand variables during ANN training, there was an improvement in estimates of dominant heights and generated a 13% more stable productive capacity classification compared to non-linear regression models.

Keywords: Artificial intelligence; Forest sites; Feed Forward Back Propagation Network; Cascade Forward Back Propagation Network

1 INTRODUÇÃO

As pesquisas acerca da previsão da produtividade de um determinado local são cada vez mais importantes dentro do contexto dos processos de Planejamento Florestal, pois contribuem para melhorar a estimativa da colheita anual, para a obtenção da idade de rotação e para as decisões que envolvem o regime de manejo a ser adotado após a colheita, nos casos em que há a opção de condução das brotações. Tais resultados são fundamentais quando se deseja a racionalização e a otimização da utilização dos recursos florestais (SILVA *et al.*, 2013) e têm como primeira etapa a Classificação da Capacidade Produtiva, que pode ser feita utilizando-se equações que descrevem alguma variável dendrométrica que represente esse potencial local em função da idade do povoamento. A classificação da capacidade produtiva local é crucial para a tomada de decisões no manejo florestal, orientando sobre os locais onde serão realizados investimentos e implantados os povoamentos (SCHUCHOVSKI *et al.*, 2019).

A altura média das árvores dominantes é a principal variável dendrométrica empregada para indicar a capacidade produtiva de florestas, pois possui boa correlação com a produção volumétrica, indica a qualidade do sítio para todas as idades presentes e é menos dependente da densidade do povoamento. Todavia, sabe-se que vários outros fatores influenciam no desenvolvimento de uma floresta, e que, portanto, se utilizados para a sua modelagem, poderão trazer benefícios na aplicação da técnica.

As curvas de índice de local constituem-se atualmente no método mais prático e difundido para classificar a produtividade florestal. Essa produtividade é expressa por meio de um índice de sítio, que se refere a uma altura dominante ou a uma classe de produção relativa em uma idade de referência do povoamento (SANTOS *et al.*, 2018).

Na geração de tais curvas, são normalmente utilizados dados de altura dominante obtidos em parcelas permanentes de inventários contínuos, medidas em diferentes idades nos diferentes talhões que compõem um povoamento florestal. A relação altura-idade é ajustada a modelos matemáticos e posteriormente em função da idade de referência, diferentes classes de altura são definidas e curvas de sítio são geradas para representar os limites inferiores e superiores das classes nas diferentes idades.

Para a classificação da capacidade produtiva, nas diferentes idades, é avaliado em função da média das alturas dominantes, qual a classe de sítio que cada parcela pertence. Nesta etapa, verifica-se em algumas unidades amostrais, um problema sério da metodologia, que se trata daquelas parcelas que mudam de sítio nas diferentes idades, devido a erros de medição ou de estimativa ou por serem parcelas que apresentam alturas muito próximas ao limite de uma determinada classe de sítio. Por isso, no final da classificação, o modelo mais adequado deve ser aquele que gera uma classificação mais estável, ou seja, a família de curvas que caracterizará a capacidade produtiva dos talhões de um povoamento, deve ser aquela obtida pelo modelo que, em uma avaliação feita em parcela por parcela, nas diferentes idades de medição, não mudaram ou mudaram poucas vezes de sítio.

Técnicas de Inteligência Artificial estão sendo testadas nas ciências florestais para estimar variáveis dendrométricas e características de povoamentos florestais. As Redes Neurais Artificiais (RNA) se destacam devido as suas características de processamento semelhantes à informação do cérebro humano, tais como: não-linearidade, alto paralelismo, robustez, tolerância a falhas e ruídos, capacidade de lidar com informações imprecisas e distorcidas, além da sua capacidade de generalizar (LEITE *et al.*, 2011).

As RNA têm sido utilizadas com sucesso em diferentes situações, apresentando resultados equivalentes ou superiores aos modelos usuais de regressão quando aplicadas para aproximação de funções em modelagem florestal. Alguns estudos já foram realizados para estimativa de altura de árvores (DIAMANTOPOULOU, 2012; BINOTI *et al.*, 2013; ÖZÇELİK *et al.*, 2013; SOARES *et al.*, 2021) e classificação da capacidade produtiva. Cosenza *et al.* (2015) utilizaram RNA e Máquina de Vetor de Suporte (MVS) para definir classes de sítios florestais a partir de uma classificação pré-estabelecida por métodos tradicionais empregados. Nesse estudo, eram apresentadas as RNA e MVS, variáveis dendrométricas como entrada e a classe de sítio correspondente como saída. Porém, a estimativa de altura de árvores dominantes e posterior aplicação para classificação de sítios florestais utilizando essas abordagens ainda não foi encontrada na literatura científica atual.

Souza *et al.* (2021) avaliaram a performance da modelagem para classificação de sítios florestais em bases de dados com *outliers* e verificaram, ao comparar modelos de regressão linear com redes neurais de múltiplas camadas e de função de base radial na estimativa de alturas dominantes, que as redes de múltiplas camadas são indicadas para classificação de sítios florestais quando há presença de *outliers* na base de dados.

Nesse contexto, este trabalho teve como objetivo propor um modelo de Rede Neural Artificial (RNA) para estimar a altura dominante de árvores de eucalipto, em diferentes idades, aplicá-lo na classificação de sítios florestais utilizando o método

da curva guia e avaliar a estabilidade da classificação de sítios obtida como uma alternativa aos modelos de regressão não linear. A hipótese central que norteou esta pesquisa era que como as RNA melhoram as estimativas e especificamente no caso de estimativas de alturas dominantes, que é uma variável majoritariamente medida de forma indireta (utilizando hipsômetros) onde estamos sujeitos a imprecisões, as RNA poderiam gerar maior precisão e estabilidade para a classificação, sobretudo quando outras variáveis são fornecidas no treinamento das redes, e diminuir o empirismo na fase final da definição de classes de produtividade para as unidades de manejo.

2 MATERIAL E MÉTODOS

Os dados utilizados neste estudo são provenientes de plantios comerciais de híbridos de *Eucalyptus urophylla* x *Eucalyptus grandis*, implantados com espaçamento variando entre 9 e 12 m² por planta. Tais plantios estão localizados no extremo sul do estado da Bahia, abrangendo uma área de aproximadamente 96.000 ha, em primeira rotação, com idades de medição variando de 0,80 até 12,70 anos.

O clima da região é, segundo a classificação de Köppen, Tropical Equatorial (Af) e Tropical de Monção (Am). A região apresenta temperatura média anual de 24°C.

Os dados dendrométricos totalizaram 28.150 registros provenientes de 8.819 parcelas permanentes. As parcelas possuem, predominantemente, 260 m² de área, havendo uma minoria de parcelas com áreas que variam entre 200 e 400 m². Em cada parcela mediu-se a circunferência à altura do peito, tomado a 1,30 m do solo (CAP) de todas as árvores, a altura total das cinco primeiras árvores e a altura total das árvores dominantes, segundo o conceito de Assmann.

Para a construção das curvas de índices de local, foi empregado o método da curva guia, considerando-se as alturas dominantes estimadas pelo método tradicional (modelos de regressão não linear) e pelas RNA.

Foi assumida uma idade-índice de sete anos, o que corresponde à idade de rotação da espécie em estudo. A amplitude adotada entre as classes de sítio foi de cinco metros, exceto para a classe de sítio mais produtiva, que teve uma amplitude de seis metros, de maneira que a curva abrangesse todos os dados observados. Foram testados cinco modelos não lineares para estimativa da altura de árvores dominantes (Equações 1 a 5).

$$Hd = \beta_0 (1 - e^{-\beta_1 t})^{\beta_2} + \varepsilon \quad (1)$$

Chapman e Richards (MOD 1)

$$Hd = \beta_0 e^{-\beta_1 e^{-\beta_2 t}} + \varepsilon \quad (2)$$

Gompertz (MOD 2)

$$Hd = \frac{\beta_0}{1 + \beta_1 e^{\beta_2 t}} + \varepsilon \quad (3)$$

Logístico (MOD 3)

$$Hd = \beta_0 (1 - e^{-\beta_1 t \beta_2}) + \varepsilon \quad (4)$$

Weibull (MOD 4)

$$Hd = \beta_0 (1 - \beta_1 e^{-\beta_1 t}) + \varepsilon \quad (5)$$

Monomolecular (MOD 5)

Em que: Hd = Média das alturas das árvores dominantes; β'_s = parâmetros a serem estimados; e = número de Euler; ε = erro de estimativa; t = idade em anos.

Os parâmetros dos modelos foram estimados pelo método de *Levenberg-Marquardt* com o auxílio do software *R Statistical Computing* (versão 3.2.1).

Foram utilizadas redes neurais de múltiplas camadas (MLP), treinadas com dois algoritmos: *Feed Forward Back Propagation Network* (FFBP) e *Cascade Forward Back Propagation Network* (CFBP).

O FFBP treina uma rede MLP em que as camadas são interligadas por sinapses completas, ou seja, cada neurônio da camada i é ligado a todos os neurônios da camada seguinte ($i+1$), sendo as sinapses orientadas somente para frente (tipo *feed forward*). Sendo assim, durante o treinamento com esse algoritmo, os cálculos são realizados a partir da camada de entrada da rede em direção a camada de saída e os valores de erro são então propagados para camadas anteriores.

Já as redes treinadas com o CFBP são semelhantes às redes FFBP, mas incluem uma ligação dos pesos a partir da entrada para cada camada e a partir de cada camada para as camadas sucessivas. Enquanto a alimentação de duas camadas de redes para a frente pode potencialmente aprender virtualmente qualquer relação de entrada e saída, as redes *Feed Forward* com mais camadas podem aprender relacionamentos complexos mais rapidamente.

Inicialmente, as RNA foram treinadas considerando-se somente a idade como variável de entrada, como ocorre nos modelos de regressão não linear. Isso foi necessário para permitir a justa comparação de desempenho das duas técnicas.

Posteriormente, e considerando que vários fatores afetam diretamente a produtividade local, avaliou-se a utilização de variáveis edafoclimáticas (tipo de solo e precipitação média) e do povoamento (área basal e diâmetro médio quadrático) na camada de entrada das RNA juntamente com a variável idade. Foram consideradas 19 classes de precipitação (variando entre 850 e 1550 mm) e 15 tipos de solo. Assim, 13 diferentes RNA foram treinadas, cada uma correspondendo à uma configuração em termos de quantidade e de tipo de variáveis de entrada (Tabela 1).

Todas as RNA treinadas possuíam uma única camada oculta e uma quantidade de neurônios em tal camada igual a quatro para cada variável de entrada. A altura dominante foi considerada como variável de saída. Utilizou-se a função de transferência sigmoide para a camada oculta e linear para a camada de saída, e os dados foram normalizados.

Tabela 1 – Arquitetura e variáveis de entrada das RNA treinadas para estimar altura dominante (Hd)

RNA	Arquitetura	Algoritmo de treinamento	Variáveis de Entrada				
			Idade	Área Basal	Diâmetro Médio Quadrático	Solo	Precipitação
1	01 - 04 - 01	FFBP	X				
	01 - 04 - 01	CFBP	X				
2	02 - 08 - 01	FFBP	X	-	-	X	-
	02 - 08 - 01	CFBP	X	-	-	X	-
3	02 - 08 - 01	FFBP	X	-	-	-	X
	02 - 08 - 01	CFBP	X	-	-	-	X
4	03 - 12 - 01	FFBP	X	-	-	X	X
	03 - 12 - 01	CFBP	X	-	-	X	X
5	02 - 08 - 01	FFBP	X	X	-	-	-
	02 - 08 - 01	CFBP	X	X	-	-	-
6	03 - 12 - 01	FFBP	X	X	X	-	-
	03 - 12 - 01	CFBP	X	X	X	-	-
7	03 - 12 - 01	FFBP	X	X	-	X	-
	03 - 12 - 01	CFBP	X	X	-	X	-
8	04 - 16 - 01	FFBP	X	X	-	X	X
	04 - 16 - 01	CFBP	X	X	-	X	X
9	04 - 16 - 01	FFBP	X	X	X	X	-
	04 - 16 - 01	CFBP	X	X	X	X	-
10	05 - 20 - 01	FFBP	X	X	X	X	X
	05 - 20 - 01	CFBP	X	X	X	X	X
11	02 - 08 - 01	FFBP	X	-	X	-	-
	02 - 08 - 01	CFBP	X	-	X	-	-
12	03 - 12 - 01	FFBP	X	-	X	X	-
	03 - 12 - 01	CFBP	X	-	X	X	-
13	04 - 16 - 01	FFBP	X	-	X	X	X
	04 - 16 - 01	CFBP	X	-	X	X	X

Fonte: Autores (2021)

Em que: FFBP = RNA obtida com o algoritmo *Feed Forward Back Propagation Network* e CFBP = RNA obtida com o algoritmo *Cascade Forward Back propagation Network*.

A base de dados foi dividida aleatoriamente em dois conjuntos: 70% para treinamento e 30% para generalização (20% para teste e 10% para validação cruzada). Como critério de parada dos algoritmos de treinamento, utilizou-se o número total de ciclos igual a 1.000 ou o erro quadrático médio inferior a 1% (LEITE *et al.*, 2016). Os modelos de RNA foram construídos usando a caixa de ferramentas do *Neural Network Toolbox* do software MatLab (versão 7.10).

As estimativas foram avaliadas considerando-se o coeficiente de correlação (r) entre alturas dominantes observadas e estimadas, a raiz quadrada do erro médio em metros (RMSE) e a raiz quadrada do erro médio em porcentagem (RMSE%), conforme as Equações (6) e (7) (BINOTI *et al.*, 2013). Ainda, foram gerados gráficos de dispersão dos erros percentuais (Equação 8).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Hd - \widehat{Hd})^2}{n}} \quad (6)$$

$$RMSE\% = \left(\frac{RMSE}{\underline{Hd}} \right) * 100 \quad (7)$$

$$Erro (\%) = \left(\frac{(Hd - \widehat{Hd})}{\underline{Hd}} \right) * 100 \quad (8)$$

Em que: Hd = Altura dominante observada; \widehat{Hd} = Altura média das árvores dominantes estimada; \underline{Hd} é a média das alturas dominantes observadas em todas as parcelas; n = número total de observações.

O procedimento para a avaliação da estabilidade consistiu-se em interpretar a manutenção da classificação dos sítios para os modelos testados. Assim, um modelo que apresenta uma classificação mais estável pode ser definido como aquele capaz de manter uma parcela classificada numa mesma classe de sítio do primeiro ao último ano observado na maior parte das parcelas.

Para testar a estabilidade da classificação da capacidade produtiva ao longo das remedições, o índice de sítio foi estimado e então foi contabilizado o número de vezes que as parcelas mudavam de classe em cada abordagem avaliada. A abordagem considerada mais estável foi aquela que apresentou um maior número de parcelas que não mudaram de classe nas suas observações.

Comparou-se os resultados de estabilidade a partir do teste de Qui-Quadrado (χ^2), em tabelas de contingência. Na aplicação do teste, foram considerados três experimentos, testando todas as combinações, duas a duas para as opções avaliadas conforme segue:

a) Experimento 1: Comparou-se os resultados de estabilidade para a classificação gerada pelos modelos de regressão não lineares, com o intuito de selecionar o modelo que gerou a melhor classificação de acordo com a estabilidade para ser comparado com a classificação gerada por RNA;

b) Experimento 2: O modelo selecionado no primeiro experimento foi comparado aos modelos de RNA gerados pelos dois algoritmos (FFBP e CFBP), considerando apenas a idade como variável de entrada;

c) Experimento 3: O melhor modelo selecionado no experimento 1 e os modelos de RNA geradas, considerando apenas a idade como variável de entrada, foram comparados com os melhores modelos de RNA gerados pelos dois algoritmos ((FFBP e CFBP), considerando como entrada além da idade, as variáveis edafoclimáticas e do povoamento.

Ainda, para avaliar a precisão, comparou-se o melhor modelo de regressão não linear, as RNA apenas com a variável idade como entrada e as melhores RNA com entrada da idade, variáveis edafoclimáticas e do povoamento, com as alturas dominantes observadas, aplicando-se o teste F de Graybill.

Todas as análises foram realizadas considerando um nível de significância $\alpha=0,05$.

3 RESULTADOS E DISCUSSÃO

A elaboração das curvas de índice de sítio depende de uma curva que descreva o comportamento médio dos dados, especialmente para o caso de curvas anamórficas. Para os dados considerados, todos os modelos apresentaram resultados satisfatórios para estimativa de altura das árvores dominantes do povoamento em estudo (Tabela 2).

Tabela 2 – Estatísticas de exatidão para estimativa da altura dominante (Hd) e estabilidade da classificação de sítios para cada modelo avaliado

Modelo	Estatísticas					
	Correlação	RMSE (m)		RMSE (%)		
1	0,9201	2,430		9,452		
2	0,9184	2,455		9,551		
3	0,9160	2,489		9,683		
4	0,9202	2,431		9,582		
5	0,9203	2,427		9,441		
	Percentual que não mudaram de sítio	Número de vezes que as parcelas mudaram de sítio				
		0	1	2	3	≥ 4
		Frequência observada				
1	45,05	3976	3397	1094	302	50
2	43,30	3822	3517	1121	304	55
3	40,72	3594	3663	1203	307	52
4	45,08	3979	3413	1084	298	45
5	45,54	4019	3390	1064	293	53

Fonte: Autores (2021)

Em que: RMSE (m) = raiz quadrada do erro médio em metros; RMSE (%) = raiz quadrada do erro médio em porcentagem.

Verificou-se um coeficiente de correlação variando de 0,9160 a 0,9203 e raiz do erro médio quadrático (RMSE%) entre 9,683 e 9,441% para os modelos 3 e 5 (destacado em negrito), respectivamente. Todos os modelos, ao analisar a dispersão residual, não apresentaram tendências e em relação as estatísticas de precisão, performance similar. Fiorentin *et al.* (2017), quando avaliaram a aplicação de modelos não lineares para a classificação de sítios florestais em *Pinus elliotti* e *Pinus taeda* na região de

Caçador-SC, também verificaram performance similar para a maioria dos modelos avaliados, exceto para o modelo de Clutter e Jones, onde os autores verificaram resultados ligeiramente inferiores. O desempenho dos modelos está intimamente relacionado a consistência dos dados e como as relações entre as variáveis são estabelecidas, portanto é comum em certos casos, alguns modelos não retornarem bons resultados. Contudo, verifica-se que o modelo 5 foi o que apresentou a melhor dispersão de resíduos. Tal resultado, em relação a esse modelo, é semelhante ao encontrado por Téo *et al.* (2011).

Os modelos testados apresentaram resultados semelhantes para a estabilidade (Tabela 2). A porcentagem de parcelas estáveis variou entre 40,72% a 45,54% para os modelos 3 e 5 (destacado em negrito), respectivamente. Tal avaliação revela que o comportamento das parcelas ao longo das remediações segue o mesmo padrão do desempenho de precisão estatística dos modelos.

Os dois tipos de RNA treinadas com as mesmas variáveis utilizadas nos modelos de regressão não linear apresentaram resultados de precisão estatística e de estabilidade bastante próximos aos modelos de regressão testados (Tabela 3). As redes neurais treinadas com o CFBP (destacado em negrito) apresentaram resultados um pouco superiores aos apresentados pelas redes com o FFBP, resultados bem próximos aos encontrados para o modelo 5.

Os dois tipos de RNA apresentaram visualmente uma dispersão dos resíduos sem tendenciosidades, semelhantes e com o mesmo comportamento ao encontrado pelos modelos de regressão não linear, com maiores erros nas menores idades.

Já para o teste de estabilidade, as RNA apresentaram 46,71 % e 46,76 % de parcelas que permaneceram estáveis na classificação, com os algoritmos FFBP e CFBP, respectivamente. Esses valores em porcentagem representam um total de 4.119 e 4.124 parcelas estáveis, ao passo que o modelo de regressão não linear (modelo 5) alcançou uma estabilidade de 45,54 %, ou seja, um total de 4.019 parcelas estáveis. As RNA conseguiram, portanto, classificar os sítios florestais de forma satisfatória.

Tabela 3 – Estatísticas de precisão e estabilidade para as diferentes RNA treinadas com variáveis semelhantes aos modelos de regressão não linear para estimativa da altura de árvores dominantes (Hd)

RNA		Correlação				RMSE	RMSE	
		Trein.	Val.	Teste	Total	(m)	(%)	
1	FFBP	0,9240	0,9224	0,9248	0,9239	2,416	9,405	
	CFBP	0,9259	0,9250	0,9228	0,9251	2,413	9,392	
		Percentual que não mudaram de sítio	Número de vezes que as parcelas mudaram de sítio					
			0	1	2	3	≥ 4	
		Frequência observada						
1	FFBP	46,71	4119	3241	1090	301	68	
	CFBP	46,76	4124	3245	1082	297	71	

Fonte: Autores (2021)

Em que: Trein. = treinamento; Val. = validação; RMSE (m) = raiz quadrada do erro médio em metros; RMSE (%) = raiz quadrada do erro médio em porcentagem; FFBP = RNA obtida com o algoritmo *Feed Forward Back Propagation Network* e CFBP = RNA obtida com o algoritmo *Cascade Forward Back propagation Network*.

As RNA treinadas com adição de variáveis edafoclimáticas e do povoamento tendem a apresentar resultados superiores aos modelos de regressão não linear e das RNA treinadas para estimar altura dominante em função da idade (Tabela 4). A adição de mais variáveis de entrada resulta em um aumento da precisão nas estimativas, podendo gerar benefícios para a técnica da classificação da capacidade produtiva em florestas.

Todas as RNA propostas foram capazes de estimar a altura dominante sem tendenciosidade de resíduos.

De modo geral, as redes neurais treinadas com o CFBP apresentaram uma leve superioridade em relação às redes com o FFBP. Tal resultado é semelhante ao encontrado em trabalhos realizados em outras áreas da ciência, tais como em Diamantopoulou (2012), Tengelen e Armand (2014) e Jarndal *et al.* (2021).

Isso pode ter ocorrido devido as redes CFBP possuírem um efeito em cascata durante o seu treinamento: as suas camadas são ligadas a todas as camadas posteriores,

transmitindo os valores de pesos calculados a partir dos dados de entrada, fazendo com que o seu aprendizado seja superior às redes treinadas com o FFBP, que tem apenas a ligação de uma camada com a sua posterior.

A adição de variáveis que descrevem o comportamento do povoamento (área basal e o diâmetro médio quadrático) melhoraram mais a precisão das redes do que a adição de variáveis categóricas (solo e precipitação), posto que os valores de correlação foram superiores a 0,95 e o RMSE% variou entre 6,40 e 7,16%. As RNA 2, 3 e 4, que tiveram como entrada a adição de somente variáveis categóricas, apresentaram valores de coeficiente de correlação e RMSE% próximos aos alcançados pelo modelo de regressão não linear e pelas redes que foram treinadas com apenas a idade como variável de entrada. Özçelik *et al.* (2013) também observaram uma redução nos valores de RMSE para as estimativas geradas por RNA em relação aos modelos de regressão não linear.

Cosenza *et al.* (2015) mencionam que a inclusão de muitas variáveis pouco significativas pode prejudicar ou ser indiferente no desempenho de redes neurais artificiais. De acordo com autores, para que a classificação ocorra de forma satisfatória, o uso conjunto das variáveis de tipo de solo, espaçamento do plantio, idade e altura dominante foram suficientes para classificar os sítios. A inclusão da idade, nesse caso, foi fundamental, por dar sentido ao valor da altura dominante, de forma que um talhão com idade avançada e baixa altura dominante indicaria baixo potencial produtivo e vice-versa.

As RNA com variáveis edafoclimáticas (tipo de solo e precipitação média) apresentaram um percentual de parcelas estáveis entre 46,14% e 48,24% (Tabela 4). Tais resultados representam um ganho em quantidade de parcelas estáveis obtidas na classificação a partir da RNA 2 do tipo CFBP igual a 200 parcelas em relação aos modelos que estimavam altura dominante somente com a variável idade.

Tabela 4 – Estatísticas de precisão e estabilidade para as diferentes RNA treinadas com as demais variáveis para estimativa da altura de árvores dominantes (Hd)

RNA	Correlação				RMSE (m)	RMSE (%)
	Trein.	Val.	Teste	Total		
2	0,9240	0,9224	0,9248	0,9239	2,374	9,24
	0,9259	0,9250	0,9228	0,9251	2,355	9,17
3	0,9209	0,9245	0,9210	0,9216	2,407	9,37
	0,9351	0,9314	0,9343	0,9342	2,214	8,62
4	0,9246	0,9208	0,9251	0,9240	2,374	9,24
	0,9383	0,9357	0,9351	0,9372	2,164	8,42
5	0,9554	0,9549	0,9539	0,9550	1,840	7,16
	0,9565	0,9554	0,9533	0,9556	1,827	7,11
6	0,9600	0,9583	0,9592	0,9595	1,748	6,81
	0,9593	0,9604	0,9600	0,9596	1,745	6,79
7	0,9575	0,9573	0,9549	0,9569	1,802	7,01
	0,9579	0,9588	0,9559	0,9577	1,785	6,95
8	0,9601	0,9610	0,9569	0,9597	1,745	6,79
	0,9611	0,9592	0,9600	0,9605	1,729	6,73
9	0,9613	0,9610	0,9602	0,9611	1,715	6,68
	0,9611	0,9618	0,9610	0,9612	1,710	6,66
10	0,9627	0,9641	0,9613	0,9627	1,678	6,53
	0,9651	0,9630	0,9638	0,9644	1,643	6,40
11	0,9564	0,9547	0,9556	0,9559	1,824	7,10
	0,9571	0,9544	0,9588	0,9569	1,801	7,01
12	0,9586	0,9590	0,9568	0,9583	1,771	6,90
	0,9590	0,9591	0,9574	0,9587	1,765	6,87
13	0,9610	0,9603	0,9613	0,9609	1,721	6,70
	0,9617	0,9607	0,9625	0,9616	1,703	6,63
	Percentual que não mudaram de sítio	Número de vezes que as parcelas mudaram de sítio				
		0	1	2	3	≥ 4
		Frequência observada				
2	46,42	4094	3309	1043	309	64
	46,14	4069	3376	1015	288	72
3	46,79	4126	3241	1094	296	62
	48,24	4254	3149	1052	314	50
4	46,07	4063	3345	1037	319	55
	47,05	4149	3272	1018	328	52
5	55,40	4886	2604	1013	256	60
	56,41	4975	2545	997	251	51
6	57,13	5038	2501	982	241	57
	57,47	5068	2477	979	246	49

Continua ...

Tabela 4 – Conclusão

	Percentual que não mudaram de sítio	Número de vezes que as parcelas mudaram de sítio				
		0	1	2	3	≥ 4
		Frequência observada				
7	57,49	5070	2479	965	262	43
	57,76	5094	2447	988	251	39
8	57,89	5105	2459	983	237	35
	57,67	5086	2465	988	241	39
9	58,10	5124	2458	971	229	37
	58,05	5119	2467	955	230	48
10	58,61	5169	2431	949	225	45
	58,68	5175	2414	959	223	48
11	56,00	4939	2573	987	256	64
	57,00	5027	2489	980	264	59
12	56,97	5024	2544	950	253	48
	57,15	5040	2539	947	246	47
13	57,41	5063	2496	971	228	61
	58,00	5115	2427	982	243	52

Fonte: Autores (2021)

Em que: Trein. = treinamento; Val. = validação; RMSE (m) = raiz quadrada do erro médio em metros; RMSE (%) = raiz quadrada do erro médio em porcentagem.

Já para as redes que tiveram a adição de variáveis dendrométricas durante o seu treinamento, a estabilidade de parcelas foi sempre superior a 55%, o que significa um aumento de 6,36% no número de parcelas estáveis, quando comparada a estabilidade do modelo não linear de Chapman e Richards (MOD 1). Na melhor configuração para as redes, na RNA 10, verificou-se uma estabilidade de 58,61% e 58,68%, respectivamente, para as redes treinadas com o FFBP e CFBP (destacados em negrito), o que significa um aumento de 13,06% e 13,14% em estabilidade. Esse ganho em percentual de parcelas que não mudaram de sítio pode parecer pequeno a princípio, contudo ao se considerar o total de parcelas envolvidas neste estudo, verificamos que 1.152 e 1.159 parcelas, respectivamente, passaram a não mudar de sítio com o uso da rede neural para estimar a altura de dominantes.

Assim como nos resultados das estatísticas, foi possível também perceber uma superioridade das redes com o CFBP em relação ao FFBP também no teste de estabilidade: as redes treinadas com o primeiro algoritmo, na maioria das vezes, apresentaram um maior número de parcelas estáveis ao longo do tempo e, conseqüentemente, um menor número de parcelas que mudaram de sítio nos anos de remediações.

A Tabela 5 mostra o resultado do teste de Qui-quadrado dos três experimentos propostos para comparação da estabilidade da classificação para as abordagens estudadas.

Tabela 5 – Teste de Qui-Quadrado para os experimentos de comparação da estabilidade da classificação de sítios (Exp 1, 2 e 3)

		MOD 1	MOD 2	MOD 3	MOD 4	MOD 5	RNA 1 FFBP	RNA 1 CFBP	RNA 10 FFBP	RNA 10 CFBP
Exp 1	MOD 1	-	-	-	-	-	-	-	-	-
	MOD 2	5,698	-	-	-	-	-	-	-	-
	MOD 3	34,551*	12,971*	-	-	-	-	-	-	-
	MOD 4	0,374	6,401	35,236*	-	-	-	-	-	-
	MOD 5	0,879	9,011	43,152*	1,159	-	-	-	-	-
Exp 2	RNA 1 FFBP	-	-	-	-	6,858	-	-	-	-
	RNA 1 CFBP	-	-	-	-	7,314	0,126	-	-	-
Exp 3	RNA 10 FFBP	-	-	-	-	318,081*	259,788*	258,715*	-	-
	RNA 10 CFBP	-	-	-	-	324,666*	264,361*	263,205*	0,221	-

Fonte: Autores (2021)

Em que: *valores em negrito são significativos para $\alpha=0,05$; Exp 1 = Experimento 1; Exp 2 = Experimento 2; Exp 3 = Experimento 3; MOD1 = Modelo de Chapman e Richards; MOD2 = Modelo de Gompertz; MOD3 = Modelo Logístico; MOD4 = Modelo de Weibull; MOD5 = Modelo Monomolecular; RNA 1 FFBP = Rede neural artificial 1 obtida com o algoritmo *Feed Forward Back Propagation Network*; RNA 1 CFBP = Rede neural artificial 1 obtida com o algoritmo *Cascade Forward Back propagation Network*; RNA 10 FFBP = Rede neural artificial 10 obtida com o algoritmo *Feed Forward Back Propagation Network*; RNA 10 CFBP = Rede neural artificial 10 obtida com o algoritmo *Cascade Forward Back propagation Network*.

Analisando os resultados para o Experimento 1, verifica-se que o único modelo de regressão linear que se diferenciou dos demais modelos avaliados foi o MOD 3 (Logístico). O MOD 5 (Monomolecular) foi o que melhor se ajustou (maior correlação e menor RMSE%) e o que apresentou melhor estabilidade, sendo este então comparado com os resultados de estabilidade obtidos com as RNA treinadas apenas com a variável de entrada idade (Experimento 2).

Os resultados do Experimento 2 indicam não haver diferenças entre o MOD 5 e a RNA 1, treinada com os algoritmos FFBP e CFBP. Esse resultado sugere que embora se observe uma ligeira melhora nas estatísticas de precisão e também na estabilidade da classificação com as alturas estimadas pela rede, a estabilidade foi a mesma obtida com o modelo de regressão não linear. Portanto, modelos de redes neurais com apenas a variável de entrada idade e compatíveis aos modelos clássicos de regressão não linear para estimativa de altura dominante não geram uma classificação da capacidade produtiva mais estável.

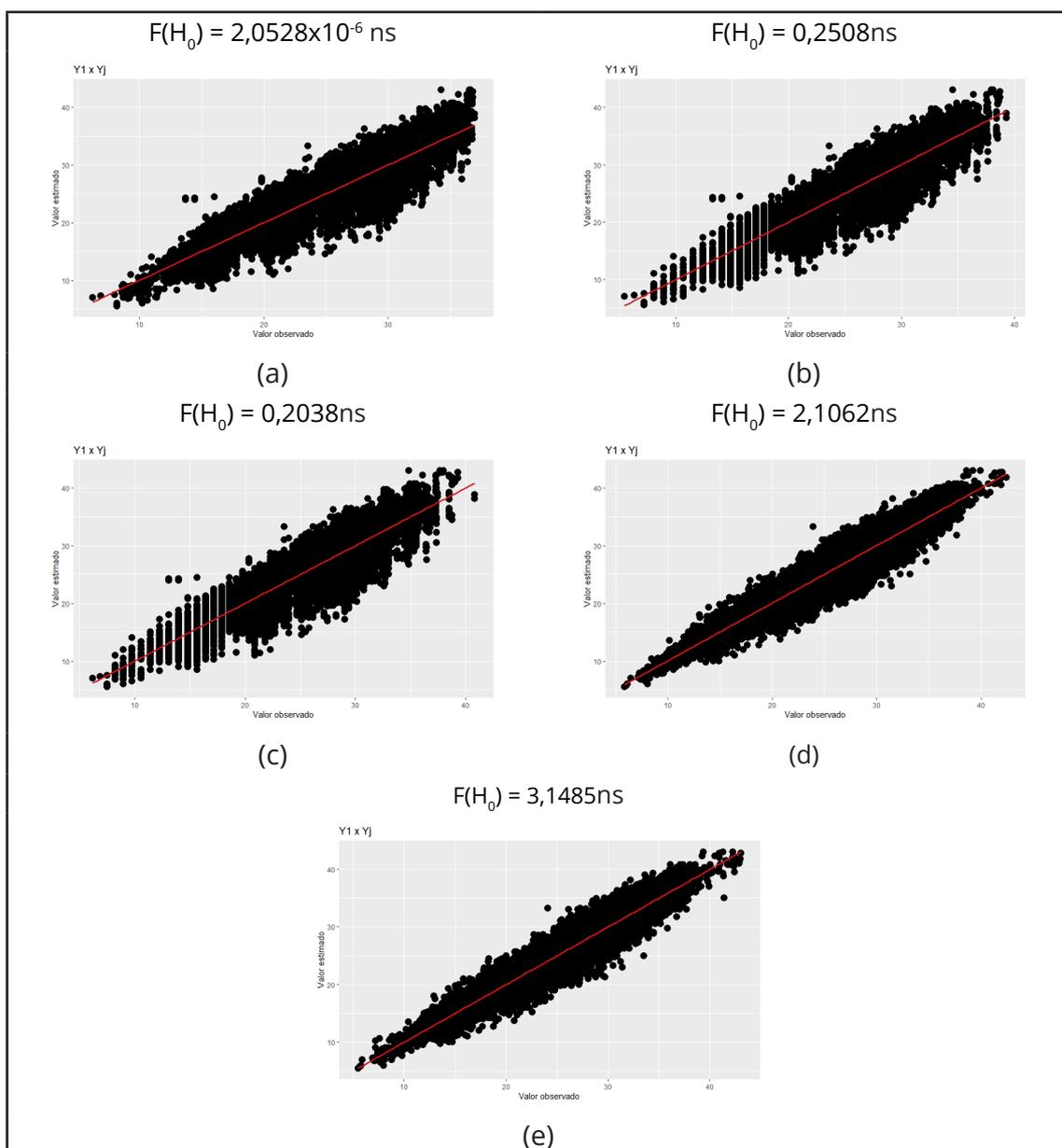
Contudo, quando se comparou no Experimento 3, o MOD 5, a RNA 1 (treinada com ambos algoritmos) e a RNA 10 (treinada com FFBP e CFBP) com a introdução de variáveis do povoamento e edafoclimáticas somadas à idade das parcelas no modelo de rede neural, verificou-se que a RNA 10 diferiu estatisticamente das demais abordagens. Esse resultado sugere que a introdução de tais variáveis e a idade para estimar a altura de dominantes geram uma classificação mais estável.

A Figura 1 mostra, para as melhores abordagens, a dispersão entre valores observados e estimados e o valor da estatística $F(H_0)$ para o teste F de Graybill.

Na Figura 1, observa-se que para todas as abordagens apontadas como melhores em análises anteriores, a estatística $F(H_0)$ foi não significativa, portanto, aceita-se a hipótese H_0 , isto é, estatisticamente $\beta_0=0$ e $\beta_1=1$, pelo teste F Graybill, a 5% de significância. Pode-se afirmar que as estimativas de altura dominante obtidas pelo modelo Monomolecular (MOD5), RNA1 FFBP, RNA1 CFBP, RNA10 FFBP e RNA10 CFBP são estatisticamente iguais às alturas dominantes observadas. Contudo, observa-se

que a dispersão é claramente mais uniforme (com menores erros) para as estimativas com as RNA10 FFBP e RNA10 CFBP, que consideram como entrada além da idade, as variáveis edafoclimáticas e do povoamento. A introdução de tais variáveis na RNA melhorou as estimativas de altura dominante.

Figura 1 – Dispersão entre valores observados e estimados e estatística $F(H_0)$ para o teste F de Graybill de (a) modelo Monomolecular, (b) RNA1 FFBP, (c) RNA1 CFBP, (d) RNA10 FFBP (e) RNA10 CFBP



Fonte: Autores (2021)

Neste estudo, assim como no estudo realizado por Cosenza *et al.* (2017), nenhuma distinção entre genótipos e parcelas foi feita durante o processo de ajuste, de forma que a curva para a altura dominante foi ajustada usando o conjunto de dados completo. Os autores supracitados indicam que técnicas de regressão avançadas, como modelos mistos, podem ser aplicadas para melhorar a estimativa do índice do local, uma vez que podem incorporar esses fatores e podem corrigir a autocorrelação serial residual causada pelo uso de dados longitudinais de gráficos permanentes e conseqüentemente gerar classificações mais estáveis. Os resultados encontrados pelos autores supracitados e neste trabalho reforçam o potencial das redes neurais artificiais para serem aplicadas à modelagem florestal. A classificação de sítios com base em variáveis ambientais e dendrométricas mostrou-se uma boa técnica de mapeamento de sítios e, portanto, pode ser utilizada como ferramenta de apoio aos gestores florestais.

Com isso, recomenda-se a classificação da capacidade produtiva de povoamentos não desbastados, a partir da estimativa da altura de dominantes com modelos tendo como entrada a idade das parcelas, combinadas com área basal e diâmetro médio quadrático, e variáveis edafoclimáticas de solo e precipitação. Especificamente, pela facilidade de introdução de variáveis e flexibilidade da modelagem, recomenda-se ainda que tais modelos, a exemplo deste estudo, sejam obtidos a partir do treinamento de redes neurais de múltiplas camadas (MLP), com os algoritmos *Feed Forward Back Propagation Network* (FFBP) e *Cascade Forward Back Propagation Network* (CFBP).

4 CONCLUSÃO

Nas condições em que este estudo foi conduzido, os resultados permitiram as seguintes conclusões:

Não há diferença na precisão das estimativas e estabilidade da classificação de sítios entre as redes neurais artificiais treinadas com os algoritmos FFBP e CFBP e o melhor modelo de regressão não linear, quando é usada apenas a idade como variável de entrada da RNA;

As redes neurais treinadas com o algoritmo CFBP apresentam resultados ligeiramente superiores às redes treinadas com o FFBP;

A introdução de variáveis do povoamento como área basal, diâmetro médio quadrático, combinado com a idade das parcelas e variáveis categóricas de solo e precipitação no treinamento das redes como variáveis de entrada, melhora a precisão das estimativas das alturas dominantes e gera uma classificação da capacidade produtiva mais estável que os modelos de regressão não linear.

AGRADECIMENTOS

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG, pelo financiamento deste estudo.

REFERÊNCIAS

BINOTI, M. L. M. S.; BINOTI, D. H. B.; LEITE, H. G. Aplicação de redes neurais artificiais para estimação da altura de povoamentos equiâneos de eucalipto. **Revista Árvore**, Viçosa, v. 37, n. 4, p. 639-645, ago. 2013. DOI: 10.1590/S0100-67622013000400007

COSENZA, D. N. Site classification for eucalypt stands using artificial neural network based on environmental and management features. **Cerne**, Lavras, v. 23, n. 3, p. 310-320, jul./set. 2017. DOI: 10.1590/01047760201723032352

COSENZA, D. N.; LEITE, H. G.; MARCATTI, G. E.; BINOTI, D. H. B.; ALCÂNTARA, A. E. M. de; RODE, R. Classificação da capacidade produtiva de sítios florestais utilizando máquina de vetor de suporte e rede neural artificial. **Scientia Forestalis**, Piracicaba, v. 43, n. 108, p. 955-963, dez. 2015. DOI: 10.18671/scifor.v43n108.19

DIAMANTOPOULOU, M. J. Assessing a reliable modeling approach of features of trees through neural network models for sustainable forests. **Sustainable Computing: Informatics and Systems**, [s. l.], v. 2, p. 190-197, dez. 2012. DOI: 10.1016/j.suscom.2012.10.002

FIORENTIN, L. D. *et al.* Modelos não lineares aplicados para classificação de sítios florestais de *Pinus elliottii* e *Pinus taeda*. **Revista Brasileira de Biometria**, Lavras, v. 35, n. 3, p. 474-496, set. 2017. Disponível em: <https://biometria.ufla.br/index.php/BBJ/article/view/70/115>. Acesso em: 09 ago. 2021.

JARNDAL, A.; HUSAIN, S.; HASHMI, M. Genetic algorithm initialized artificial neural network based temperature dependent small-signal modeling technique for GaN high electron mobility transistors. **International Journal RF and Microwave Computer-Aided Engineering**, [s. l.], v. 31, n. 3, p. 1-18, mar. 2021. DOI: 10.1002/mmce.22542

LEITE, H. G.; NETO, R. R. de O.; CASTRO, R. R. de; BINOTI, M. L. M. da S.; BINOTI, D. H. B.; PAULINO, E. J.; LOPES, P. F.; COLODETTE, J. L. Redes neurais artificiais para a estimação da densidade básica da madeira. **Scientia Forestalis**, Piracicaba, v. 44, n. 109, p. 149-154, mar. 2016. DOI: 10.18671/scifor.v43n109.14

LEITE, H. G.; SILVA, M. L. M.; BINOTI, D. H. B.; FARDIN, L.; TAKIZAWA, F. H. Estimation of inside-bark diameter and heartwood diameter for *Tectona grandis* Linn. trees using artificial neural networks. **European Journal of Forest Research**, [s. l.], v.130, n.2, p.263-269, mar. 2011. DOI: 10.1007/s10342-010-0427-7

ÖZÇELİK, R.; DIAMANTOPOULOU, M. J.; CRECENTE-CAMPO, F.; ELER, U. Estimating *Crimean juniper* tree height using nonlinear regression and artificial neural network models. **Forest Ecology and Management**, [s. l.] v. 306, p. 52–60, out. 2013. DOI: 10.1016/j.foreco.2013.06.009

SANTOS, C. J. Regressão não linear quantílica para classificação de sítio em povoamentos de *Eucalyptus* sp. **Scientia Forestalis**, Piracicaba, v. 46, n. 120, p. 551-561, dez. 2018. DOI: 10.18671/scifor.v46n120.04

SCHUCHOVSKI, M. S. *et al.* Altura dominante e índice de sítio para *Pinus taeda* em dois polos de desenvolvimento florestal no sul do Brasil. **Scientia Forestalis**, Piracicaba, v. 47, n. 121, p. 92-104, mar. 2019. DOI: dx.doi.org/10.18671/scifor.v47n121.09

SILVA, F. DA.; PIMENTEL, A.; CORTE, A. P. D.; SANQUETTA, C. R. Classificação de sítio para *Pinus caribaea* var. *hondurensis* na região do Triângulo Mineiro. **Revista Acadêmica: Ciências Agrárias e Ambientais**, Curitiba, v. 11, Supl. 1, p. S105-S112, 2013. DOI: 10.7213/academica.10.S01.AO12

SOARES, G. M.; SILVA, L. D.; HIGA, A. R.; SIMON, A. A; SÃO JOSÉ, J. F. B de. Redes neurais artificiais (RNA) para estimativa de altura em um plantio misto de *Eucalyptus globulus* LABILL e *Acacia mearnsii* de WILD. **Revista Árvore**, Viçosa, v. 45:e4512, s.n., p. 1-9, ago. 2021. DOI: 10.1590/1806-908820210000012

SOUZA, P. D. *et al.* Performance da modelagem para classificação de sítios florestais em bases de dados com outliers. **Nativa**, Sinop, v. 9, n. 1, p. 54-61, jan./fev. 2021. DOI: 10.31413/nativa.v9i1.11202

TENGELENG, S.; ARMAND, N. Performance of using cascade forward back propagation neural networks for estimating rain parameters with rain drop size distribution. **Atmosphere**, [s. l.], v. 5, s.n., p. 454-472, jun. 2014. DOI: 10.3390/atmos5020454

TÉO, S. J.; BRESSAN, D. R.; COSTA, R. H. da. Uso de modelos estatísticos para classificação de sítios em povoamentos de *Pinus taeda* na região de Caçador – SC. **Revista Floresta**, Curitiba, PR, v. 41, n. 1, p. 179-188, jan./mar. 2011. DOI: 10.5380/rf.v41i1.21201

Contribuição de Autoria

1 – Marcos Vinicius Santana Leite

Engenheiro Florestal, Me., Especialista de Planejamento Florestal

<https://orcid.org/0000-0001-7777-9276> • santanav.marcos@gmail.com

Contribuição: : Investigação, Metodologia, Visualização de dados (tabela, gráfico), Escrita – primeira redação

2 – Christian Dias Cabacinha

Engenheiro Florestal, Dr., Professor

<https://orcid.org/0000-0002-8148-083X> • cabacinha@ufmg.br

Contribuição: Curadoria de dados, Análise formal, Obtenção de financiamento, Administração do projeto, Recursos, Supervisão, Escrita – revisão e edição

3 – Adriana Leandra de Assis

Engenheira Florestal, Dra., Professora

<https://orcid.org/0000-0002-0723-6935> • alassis@ufmg.br

Contribuição: Análise formal, Supervisão, Escrita – revisão e edição

Como citar este artigo

Leite, M. V. S.; Cabacinha, C. D.; Assis, A. L. Classificação da capacidade produtiva com alturas de árvores dominantes estimadas por RNA. *Ciência Florestal*, Santa Maria, v. 32, n. 3, p. 1552-1574, 2022. DOI 10.5902/1980509867120. Disponível em: <https://doi.org/10.5902/1980509867120>.