



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS AGRÁRIAS E ENGENHARIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS FLORESTAIS

ISÁIRA LEITE E LOPES

**AVALIAÇÃO E SELEÇÃO DE VARIÁVEIS PREDITORAS NA ESTIMATIVA DA
DENSIDADE DA MADEIRA DE EUCALIPTO**

JERÔNIMO MONTEIRO – ES

2018

ISÁIRA LEITE E LOPES

**AVALIAÇÃO E SELEÇÃO DE VARIÁVEIS PREDITORAS NA ESTIMATIVA DA
DENSIDADE DA MADEIRA DE EUCALIPTO**

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Florestais do Centro de Ciências Agrárias e Engenharias da Universidade Federal do Espírito Santo, como parte das exigências para obtenção do Título de Mestre em Ciências Florestais na Área de Concentração Ciências Florestais.
Orientador: Prof^a. Dr^a. Mayra Luiza Marques da Silva
Coorientadora: Prof^a. Dr^a. Graziela Baptista Vidaurre

JERÔNIMO MONTEIRO – ES

2018

Dados Internacionais de Catalogação-na-publicação (CIP)
(Biblioteca Setorial Sul, Universidade Federal do Espírito Santo, ES, Brasil)
Bibliotecária: Lizzie de Almeida Chaves – CRB-6 ES-000871/O

L864a Lopes, Isáira Leite e, 1991-
Avaliação e seleção de variáveis preditoras na estimativa da densidade da madeira de eucalipto / Isáira Leite e Lopes. – 2018.
69 f. : il.

Orientador: Mayra Luiza Marques da Silva.
Coorientador: Graziela Baptista Vidaurre.
Dissertação (Mestrado em Ciências Florestais) – Universidade Federal do Espírito Santo, Centro de Ciências Agrárias e Engenharias.

1. Densidade da madeira. 2. Florestas – Medição. 3. Redes neurais (Computação). 4. Eucalipto. I. Silva, Mayra Luiza Marques da. II. Vidaurre, Graziela Baptista. III. Universidade Federal do Espírito Santo. Centro de Ciências Agrárias e Engenharias. IV. Título.

CDU: 630

**AVALIAÇÃO E SELEÇÃO DE VARIÁVEIS PREDITORAS NA ESTIMATIVA DA
DENSIDADE DA MADEIRA DE EUCALIPTO**

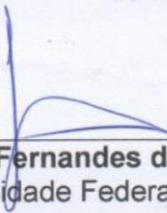
Isáira Leite e Lopes

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Florestais do Centro de Ciências Agrárias e Engenharias da Universidade Federal do Espírito Santo, como parte das exigências para obtenção do Título de Mestre em Ciências Florestais na Área de Concentração Ciências Florestais.

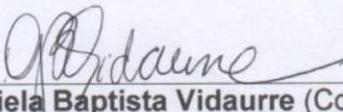
Aprovada em 28 de fevereiro de 2018.



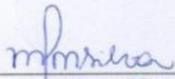
Dr^a. Aline Edwiges Mazon de Alcântara (Examinadora externa)
Cenibra



Prof. Dr. Gilson Fernandes da Silva (Examinador interno)
Universidade Federal do Espírito Santo



Profa. Dra. Graziela Baptista Vidaurre (Coorientadora)
Universidade Federal do Espírito Santo



Prof^a. Dr^a. Mayra Luiza Marques da Silva (Orientadora)
Universidade Federal de São João Del-Rei

DEDICATÓRIA

À minha mãe,
Maria Geralda Leite
À minha irmã,
Isadora Leite e Lopes
Ao meu pai,
Edson Fernando Oliveira Lopes
Dedico.

AGRADECIMENTOS

A Deus, agradeço e ofereço este trabalho a Deus, por nunca me desamparar e proporcionar ânimo e força nos momentos mais difíceis.

A Nossa Senhora pelas intercessões constantes, proteção e por passar na frente das situações, guiando meus passos.

A minha amada mãe, por reunir as virtudes de mulher batalhadora, forte, dedicada, que nunca mediu esforços para proporcionar a mim e a minha irmã, todas as oportunidades de estudo, sempre nos incentivando a dar o nosso melhor e apoiando nossas decisões.

Ao meu amado pai, por todo amor e sempre estar torcendo por mim

A minha irmã, por ser minha fiel companheira, melhor amiga, e por todo carinho.

À empresa, por ceder os dados necessários para a realização deste trabalho, pelo incentivo e disposição em esclarecer dúvidas. De modo especial, agradeço ao Mauro, Felipe, Antonilmar, Dilson e a Aline Mazon, esta última também agradeço pela pessoa acessível que é, e pela contribuição no trabalho.

A professora Mayra Luiza Marques da Silva, pela orientação, paciência, confiança depositada em mim, ensinamentos e pelo apoio para que eu obtivesse êxito neste trabalho.

A professora Graziela Baptista Vidaurre, por todas as contribuições e sempre me animar no desenvolvimento deste trabalho.

Ao professor Gilson Fernandes, pelos ensinamentos e contribuições, essenciais para melhoria deste trabalho.

Ao professor Adriano Mendonça, pelos ensinamentos e contribuição com a minha formação.

Ao programa de pós-graduação, pela oportunidade de crescimento profissional.

Aos funcionários do NEDTEC, Jocely, Patrícia, Seu Sebastião, Elisângela e a todos aqueles que sempre me acolheram com um bom dia e um sorriso no rosto.

Aos amigos do laboratório de Mensuração e Manejo Florestal (Manejão), Márcia Fernandes, Livia Figueiredo, Anny Ataide, Jeferson Martins, Jeangelis Santos, Julyana Viana, Catherine Leite, Sandra Pires, Evandro Ferreira, Giovanni

Vieira, Luandson Araújo, Taíse Aozani, Eth Rocha, Leonardo Job, Janiel Silva, Daniel Binoti, Vanessa Valentino, Quétilla Souza, Letícia Paschoa, Mariana Aragão, Sandro Barros, Rodrigo Silva, Marcelo Otone, Antônio Almeida. Agradeço a esta família, por todos os momentos que passamos juntos, pela convivência harmônica, amizade, união e todos os ensinamentos compartilhados.

A minha família e amigos, por todo incentivo e pela torcida.

Aos amigos da pós-graduação, em especial, Emanuel Araújo, Elayne Galvão e Sandra Bezerra

Às irmãs de república, Anny Ataíde, Ana Paula Câmara, Sofia Maria e Maria Naruna, por serem a minha família em Jerôminho, meu aconchego, e por toda amizade, risadas e até mesmo choros compartilhados.

Aos amigos de Jerônimo Monteiro (Jerôminho), Lucas Bitencourt, Alam Rosa (*in memoriam*), Kenya Simões, Denise Degenar, Lúbia Gaspar, Lucas Dardengo, Ricardo (Chiclete), Berdan Lacerda, Seu Sebastião, Seu João e família, Néri Fossi e família, pela amizade, acolhimento e momentos de gordices.

RESUMO

LOPES, Isáira Leite. **Avaliação e seleção de variáveis preditoras na estimativa da densidade da madeira de eucalipto**. 2018. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal do Espírito Santo, Jerônimo Monteiro, ES. Orientadora: Prof^a. Dr^a. Mayra Luiza Marques da Silva. Coorientadora: Prof^a. Dr^a. Graziela Baptista Vidaurre.

Este trabalho teve como objetivo avaliar e selecionar as variáveis preditoras mais relevantes para estimação da densidade básica da madeira de árvores de eucalipto. Foram avaliadas as variáveis qualitativas obtidas em informações cadastrais (clone, sub-região e relevo), quantitativas obtidas de Inventário Florestal Contínuo – IFC (volume total com casca, diâmetro a altura do peito e altura total) e quantitativas referentes às informações climáticas da área em estudo (velocidade do vento, temperatura média, precipitação total média, déficit de pressão de vapor, déficit hídrico e altitude), para a estimação da densidade da madeira de 386 árvores. Os métodos de avaliação e seleção de variáveis utilizados foram: força bruta com aplicação de Redes Neurais Artificiais (RNA) testando todas as possíveis combinações entre as variáveis; algoritmo de *Garson* e *Random Forest*, que quantificam a importância individual das variáveis preditoras. A classificação das variáveis preditoras variou entre os métodos, o que pode ser atribuído às suas diferentes abordagens matemáticas. A variável clone destacou-se das demais, em todos os métodos. Para o método da força bruta, a simplificação da RNA com o uso de 5 variáveis resultou em maior grau de exatidão das estimativas de densidade básica, em que a combinação ótima consistiu nas variáveis clone, idade, volume total com casca, temperatura média e déficit hídrico. Quanto ao algoritmo de *Garson*, as 5 variáveis com maior valor de importância foram: clone, sub-região, relevo, idade e déficit hídrico. Já o *Random Forest*, apresentou dentre as 5 variáveis com maior importância, o clone, idade, altura total, precipitação total média e temperatura média. Entretanto, diante do esforço computacional para aplicação do método da força bruta, uma alternativa é o uso do *Random forest* ou algoritmo de *Garson*, visto que as variáveis selecionadas nestes métodos também proporcionaram boas estimativas de densidade básica da madeira.

Palavras-chave: *Random forest*, algoritmo de *Garson*, Redes Neurais Artificiais, madeira, mensuração florestal.

ABSTRACT

LOPES, Isáira Leite. **Evaluation and selection of predictive variables in the estimation of wood density of *Eucalyptus***. 2018. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal do Espírito Santo, Jerônimo Monteiro, ES. Orientadora: Prof^a. Dr^a. Mayra Luiza Marques da Silva. Coorientadora: Prof^a. Dr^a. Graziela Baptista Vidaurre.

The objective of this work was to evaluate and select the most relevant predictor variables for estimating the basic density of eucalyptus wood. The qualitative variables obtained from cadastral data (clone, sub-region and relief), the quantitative variables obtained from the Continuous Forest Inventory - IFC (total volume with bark, diameter at breast height and total height) and quantitative data from the climatic information of the study area (wind speed, mean temperature, mean total precipitation, vapor pressure deficit, water deficit and altitude) were used to estimate the wood density of 386 trees. The methods of evaluation and selection of variables used were: brute force with application of Artificial Neural Networks (RNA) testing all possible combinations between variables; algorithm of Garson and Random Forest, that quantify the individual importance of the predictor variables. The classification of the predictor variables varied among the methods, which can be attributed to their different mathematical approaches. The clone variable stood out from the others, in all methods. For the brute force method, the simplification of the artificial neural network with the use of 5 variables resulted in a higher degree of accuracy of the basic density estimates, where the optimal combination consisted of clone, age, total volume with bark, mean temperature and water deficit. As for the Garson algorithm, the 5 variables with the highest importance were: clone, subregion, relief, age and water deficit. Random Forest presented, among the 5 most important variables, clone, age, total height, mean total precipitation and mean temperature. However, in the face of computational effort to apply the brute force method, an alternative is the use of Random forest or Garson algorithm, since the variables selected in these methods also provided good estimates of the basic density of wood.

Keywords: Random forest, Garson algorithm, Artificial Neural Networks, wood, forest measurement.

SUMÁRIO

1 INTRODUÇÃO	11
2 OBJETIVOS	13
2.1 Objetivo Geral	13
2.2 Objetivos Específicos	13
3 REVISÃO DE LITERATURA	14
3.1 Densidade básica da madeira	14
3.2 Fatores que influenciam o crescimento e a densidade da madeira	16
3.3 Redes Neurais Artificiais	19
3.4 Seleção de variáveis	24
3.4.1 Força bruta com aplicação de Redes Neurais Artificiais	24
3.4.2 <i>Random Forests</i>	25
3.4.3 Algoritmo de <i>Garson</i>	33
4 MATERIAL E MÉTODOS	36
4.1 Descrição dos dados	36
4.2 Seleção de variáveis	38
4.2.1 Preparação do banco de dados	38
4.2.2 Processamento das Redes Neurais Artificiais	39
4.2.3 Aplicação do método força bruta com base em Redes Neurais Artificiais	40
4.2.4 Aplicação do algoritmo de <i>Garson</i>	41
4.2.5 Aplicação do método <i>Random Forest</i>	42
4.3 Estimação da Densidade Básica da Madeira	42
4.4 Avaliação das estimativas geradas pelos métodos aplicados	43
5 RESULTADOS E DISCUSSÃO	45
5.1 Seleção das variáveis preditoras e estimativa da densidade com o método da força bruta	45
5.2 Seleção das variáveis preditoras e estimativa da densidade com o algoritmo de <i>Garson</i>	48
5.3 Seleção das variáveis preditoras e estimativa da densidade com a <i>Random Forest</i>	52
5.4 Resultados para estimação da densidade básica da madeira de árvores de <i>Eucalyptus</i>	54
5.4.1 Redes Neurais Artificiais	55
6 CONCLUSÕES	61
REFERÊNCIAS	62

LISTA DE FIGURAS

Figura 1 – Neurônio biológico.	20
Figura 2 – Modelo de um neurônio artificial.....	21
Figura 3 – Tipos de função de ativação.....	21
Figura 4 – Exemplificação da arquitetura <i>feedforward</i> de camadas múltiplas.	22
Figura 5 – Divisão da árvore de regressão.....	26
Figura 6 – Funcionamento da <i>Random Forest</i> para classificação e regressão.....	29
Figura 7 – Demonstração da obtenção da importância da variável baseada na permutação.....	32
Figura 8 – Demonstração dos pesos resultantes de ajustes com aplicação de RNA.	34
Figura 9 – Correlação entre a densidade básica da madeira e as variáveis quantitativas do estudo.	38
Figura 10 – Média da importância relativa das variáveis preditoras (IR%) na predição da densidade básica da madeira, pelo algoritmo de <i>Garson</i>	50
Figura 11 – Média da importância das variáveis preditoras baseada na permutação (pVI) para predição da densidade básica da madeira, pelo <i>Random Forest</i>	53
Figura 12 – Densidade básica estimada versus densidade básica observada, distribuição dos erros percentuais das estimativas de densidade básica e histograma de erros das estimativas de densidade básica, obtidas por Redes Neurais Artificiais (RNA), para os dados de treinamento.	58
Figura 13 – Densidade básica estimada versus densidade básica observada, distribuição dos erros percentuais das estimativas de densidade básica e histograma de erros das estimativas de densidade básica, obtidas por Redes Neurais Artificiais (RNA), para os dados de validação.....	59

LISTA DE TABELAS

Tabela 1 – Número de observações da densidade básica da madeira e estatísticas descritivas das variáveis quantitativas do Inventário florestal contínuo, das variáveis climáticas, da altitude e da densidade básica da madeira	37
Tabela 2 – Quantidade total de combinações de acordo com o número de variáveis utilizadas para os ajustes	40
Tabela 3 – Estatísticas de avaliação das estimativas da densidade básica da madeira durante o treinamento e da generalização das Redes Neurais Artificiais ...	43
Tabela 4 – Combinações ótimas das variáveis preditoras conforme o número de entradas, tempo de processamento (T), arquitetura e estatísticas utilizadas para avaliar o desempenho das Redes Neurais Artificiais.....	46
Tabela 5 – Estatísticas das Redes Neurais Artificiais utilizadas para cálculo da importância relativa das variáveis preditoras da densidade básica da madeira pelo algoritmo de <i>Garson</i>	49
Tabela 6 – Importância relativa das variáveis preditoras (IR%) na predição da densidade básica da madeira, pelo algoritmo de <i>Garson</i>	50
Tabela 7 – Estatísticas utilizadas para avaliar o desempenho do <i>Random forest</i> na estimação da densidade básica da madeira, para os dados de treino e validação...	52
Tabela 8 – Ranking das variáveis preditoras da densidade básica da madeira pelo <i>Random Forest</i>	53
Tabela 9 – Variáveis de entrada para as Redes Neurais Artificiais com uso de cinco variáveis selecionadas a partir dos métodos, força bruta, algoritmo de <i>Garson</i> , <i>Random Forest</i> e do Inventário florestal contínuo.	55
Tabela 10 – Estatísticas utilizadas para avaliar o desempenho das Redes Neurais Artificiais (RNAs) na estimação da densidade básica da madeira de árvores de <i>Eucalyptus</i>	56

1 INTRODUÇÃO

O setor brasileiro de florestas, nos últimos anos, vem se destacando no cenário mundial por contemplar uma extensa área de árvores plantadas (7,84 milhões de hectares), que contribui com 91% de toda a madeira produzida para fins industriais e 6,2% do PIB Industrial no país. A cadeia produtiva do setor brasileiro de florestas plantadas abrange uma gama de produtos, desde a produção, a colheita e o transporte de madeira, até a obtenção de produtos finais nos segmentos industriais de celulose, papel, painéis de madeira, pisos laminados, carvão vegetal e biomassa (IBÁ, 2017).

No Brasil, os plantios de eucalipto representam 72% do total da área de árvores plantadas com 5,67 milhões de hectares (IBÁ, 2017) pois, o gênero *Eucalyptus* adaptou-se bem às condições climáticas do país, consistindo em uma fonte de madeira de rápido crescimento, que vem sendo amplamente utilizada em todos os segmentos do setor florestal (COUTO et al., 2013).

O interesse na qualidade da madeira de plantios comerciais é tão importante quanto a produção volumétrica (MILAGRES, 2009; MAURI et al., 2015). Desta forma, é válido abordar que para o uso da madeira de modo otimizado, torna-se essencial conhecer suas propriedades físicas, anatômicas, químicas e mecânicas (COUTO et al., 2013), de modo a avaliar suas variações entre espécies, povoamentos, locais, árvores, dentro de árvores ou anéis, visto que a madeira é produzida por meio de interações complexas entre os fatores genéticos e do ambiente local (DOWNES; DREW, 2009).

A densidade básica é uma propriedade relevante para qualidade da madeira, uma vez que associa-se a qualidade do produto final nas diversas aplicações da madeira (CAIXETA et al., 2003; DOWNES; DREW, 2009). Esta propriedade é definida pela razão entre a massa de madeira seca e o volume saturado, podendo ser expressa em g.cm^{-3} ou kg.m^{-3} (LOPES; GARCIA, 2002).

Por outro lado, a determinação da densidade básica demanda muito tempo e elevados custos (DIAS et al, 2017). Neste sentido, a modelagem da densidade em função de parâmetros do povoamento e informações climáticas do local de plantio pode contribuir para obtenção de boas estimativas.

Diferentes ferramentas computacionais e matemáticas são adotadas para a modelagem da densidade básica dos povoamentos, em especial as Redes Neurais Artificiais (RNA) (LEITE et al., 2016), que tem proporcionado resultados satisfatórios (STEIN, 2010). Estas consistem em sistemas computacionais paralelamente distribuídos e constituídos de processadores simples, os quais se encontram massivamente conectados, que podem ser aplicados para auxiliar na tomada de decisões em aplicações de classificação de padrões, agrupamento de dados, aproximação de funções, recuperação de conteúdo, otimização, predição e controle (JAIN; MAO; MOHIUDDIN, 1996).

É almejada a simplificação de modelos baseados em RNA, por meio de uma seleção das variáveis entrada ou de um procedimento pelo qual as variáveis irrelevantes ou redundantes possam ser identificadas e eliminadas, para um melhor desempenho do modelo (OLIVEIRA; PEDROLLO; CASTRO, 2015).

Entretanto, apesar dos resultados promissores com aplicação de RNA, pouca atenção é dada a seleção de variáveis para posterior estimação da densidade básica, o que torna uma tarefa fundamental visto que por meio da identificação das variáveis que tem maior influência podem ser alcançadas melhores estimativas da densidade da madeira

Assim, a seleção de variáveis pode dar suporte aos gestores florestais na adoção de melhorias para o controle da qualidade da madeira destinada ao processo industrial, minimizar o custo e tempo de obtenção de dados, além de assessorar no direcionamento das práticas silviculturais.

2 OBJETIVOS

2.1 Objetivo Geral

Avaliar e selecionar variáveis preditoras mais relevantes para estimação da densidade básica da madeira de árvores de eucalipto.

2.2 Objetivos Específicos

- Obter a importância das variáveis preditoras (inventário florestal contínuo e climáticas) com base nos métodos da força bruta com aplicação de Redes Neurais Artificiais, do algoritmo de *Garson* e do *Random Forest*;
- comparar a viabilidade dos métodos na determinação da importância das variáveis preditoras;
- modelar a densidade básica por meio de Redes Neurais Artificiais;
- avaliar o desempenho das Redes Neurais Artificiais para predição da densidade básica da madeira em função das variáveis selecionadas em cada método.

3 REVISÃO DE LITERATURA

3.1 Densidade básica da madeira

A densidade básica é uma das características de maior relevância a ser estudada na madeira. A mesma é considerada complexa por estar associada a várias variáveis, como exemplo, dimensões das fibras, espessura da parede, volume dos vasos e parênquimas, proporção entre madeira primaveril e outonal, e arranjo dos elementos anatômicos (FOELKEL; BRASIL; BARRICHELO, 1971). Pode ser definida conforme a Norma Brasileira – NBR 11941 da Associação Brasileira de Normas Técnicas – ABNT (2003) como a relação da massa seca em estufa a (105 ± 2) °C pelo respectivo volume da madeira acima do ponto de saturação das fibras, resultando em uma informação segura da quantidade de matéria seca contida em um certo volume de madeira (FOELKEL, 1978), expressa em g.cm^{-3} ou kg.m^{-3} (CAMPOS; LEITE, 2009). De fácil obtenção e alta herdabilidade, além de ser passível de controle genético (BÁEZ, 1993 citado por LOPES; GARCIA, 2002), e com isto tornou-se um índice prestigiado por pesquisadores e melhoristas (BRASIL; VEIGA; MELLO, 1979).

Esta característica é abordada em diferentes áreas do setor florestal. Em relação à tecnologia da madeira, sua importância é dada de modo a obter um produto final que atenda às especificações desejadas de rendimento em celulose, resistências físico-mecânicas do papel, produção e qualidade de carvão. No melhoramento florestal, verifica-se o potencial de seleção das espécies. No manejo, permite estabelecer o tipo de prática a ser adotada de acordo com o produto final e no inventário florestal é inerente à produtividade da floresta apresentada em quantidade de madeira seca por hectare (SHIMOYAMA; BARRICHELO, 1989).

Frequentemente são realizadas pesquisas de densidade básica vinculada à qualidade da madeira. Isto também pode ser atribuído, segundo Brasil, Veiga e Mello (1979) a sua relação com as propriedades mecânicas da madeira, assim permitindo analisar o uso adequado para diversos fins (PANSHIN; DE ZEEUW, 1970 citado por LOPES; GARCIA, 2002). Diante disto, vale destacar o potencial da densidade básica na seleção de espécies conforme o produto almejado, como por

exemplo, a demanda por espécies de alta densidade para produção de carvão vegetal para siderurgia ou dormentes para estradas de ferro, diferente das exigências para produção de celulose de fibra curta, em que geralmente, prefere-se espécies de densidade básica intermediárias (RIBEIRO; ZANI FILHO, 1993).

É evidente a tendência do uso de massa de madeira seca por área nas empresas florestais de produção de celulose e carvão visando prever a quantidade dos produtos resultantes por material genético em determinado local por área. (TRUGILHO, 2009; TRUGILHO et al., 2010).

Conforme Scolforo e Thiersch (2004), a partir da obtenção da densidade básica média da árvore por meio do inventário florestal permite que, no manejo florestal, as projeções do desenvolvimento da floresta estejam associadas ao produto final, como exemplo, celulose e energia, o que afeta diretamente o planejamento otimizado da empresa florestal, implicando em tomadas de decisões mais viáveis.

Quanto a obtenção de amostras para determinação da densidade básica, essa pode ser realizada por métodos destrutivos, em que é necessário a derrubada de árvores para a retirada de discos de madeira ou cunhas opostas a 0, 25, 50, 75 e 100% da altura comercial da árvore, ou também pela retirada de cavacos (SCOLFORO, 2005). Podem ser utilizados como procedimentos para estimar a densidade, o método de atenuação da radiação gama (BRUDER, 2012), imersão em água (BRUDER; REZENDE; COSTA, 2016) e uso de balança hidrostática (ANDRADE, 2007).

Já os métodos não destrutivos, consistem em realizar pequenas inserções ou na obtenção de poucas amostras sem que haja necessidade de derrubar a árvore (SCOLFORO; THIERSCH, 2004), como por exemplo: pilodyn, resistógrafo (SCOLFORO, 2005; GOUVÊA et al., 2011; COUTO et al., 2013) e a espectroscopia no infravermelho próximo – NIRS (GOUVÊA et al., 2011). Ainda conforme esse autor, o resistógrafo é mais adequado para estimar a densidade básica da madeira de espécies de *Eucalyptus* sp, quando comparado ao pilodyn e ao NIRS, o que pode ser atribuído a obtenção da densidade com maior precisão visto que a broca do resistógrafo perfura todo o diâmetro do tronco da árvore.

O uso de métodos não destrutivos tem sido difundido em diversos setores florestais e industriais, visto que consiste em ferramenta adequada para compreensão da variação entre indivíduos e também para caracterização de

material no campo. Além disto, esses são vantajosos por proporcionarem agilidade na coleta de dados, demandarem baixo custo e pela possibilidade de avaliação de um amplo número de amostras (GOUVÊA et al., 2011). Já a determinação da densidade básica da madeira por métodos destrutivos pode demandar dias para alcançar os resultados.

Vale abordar que as variações da densidade de madeira de eucalipto ocorrem de acordo com a espécie, localidade, idade do povoamento, procedência de sementes, taxa de crescimento, entre árvores e dentro da mesma árvore (FERREIRA; KAGEYAMA, 1978). Este mesmo autor ainda reforça a relevância da obtenção de informações quanto aos aspectos de qualidade e melhoramento para as indústrias que tem a madeira de eucalipto como matéria prima.

Visto que a formação da madeira pode ser influenciada por variações genéticas, silviculturais, edáficas e geográficas, previamente à adoção de tratamento ou melhoramento, é importante analisar a relação entre estes fatores e as características de crescimento que influenciam na densidade (SHIMOYAMA; BARRICHELO, 1989). Logo, a partir das relações entre densidade e outras características, como por exemplo, diâmetro, altura, idade, capacidade produtivas dos locais, o manejo, materiais genéticos, e também características mecânicas da madeira, pode-se realizar sua modelagem (SCOLFORO, 2005).

A elaboração de metodologias que permitam estimar com rapidez e precisão a densidade auxiliam o responsável pelo inventário na obtenção de estimativas de massa e rendimento industrial dos povoamentos, assim interferindo em atividades da empresa florestal, sendo estas a pesquisa e o planejamento florestal, mais especificamente na formulação de planos de suprimento (SCOLFORO, 2005; DEMUNER, 2011).

Assim, pode-se atrelar o manejo florestal e, ou melhoramento com a tecnologia, otimizando a produtividade com a qualidade da madeira e também a qualidade do produto final, resultando assim em maior rentabilidade (SHIMOYAMA; BARRICHELO, 1989).

3.2 Fatores que influenciam o crescimento e a densidade da madeira

As propriedades da madeira são resultantes de fatores genéticos, ambientais e suas interações (BRAZ et al., 2014). A interação genótipo x ambiente influenciam na expressão fenotípica dos clones (ROSADO et al., 2012), logo as diferentes respostas dos genótipos conforme as peculiaridades dos ambientes podem auxiliar o gestor florestal na destinação apropriada dos clones com a finalidade de obter maior produção (GONÇALVES, 2015) alinhada a melhoria da qualidade da madeira.

Pois, como as condições climáticas, características do solo e material genético interferem diretamente na produtividade, também podem afetar a qualidade da madeira (FERNANDES et al., 2011). A exemplo disto, tem-se densidade da madeira que pode ser influenciada, em uma mesma espécie, pela idade, genótipo, índice de sítio, clima, localização geográfica, tratos silviculturais (VITAL, 1984), dentre outros fatores.

A densidade da madeira também é uma resposta a diversos processos morfológicos e fisiológicos, estreitamente ligados ao funcionamento das árvores. A disponibilidade de água é fundamental para o funcionamento das árvores, visto que é utilizada para atender à demanda de transpiração, manter os tecidos funcionais, translocar nutrientes, hormônios e apoiar processos bioquímicos (POORTER et al., 2010; PFAUTSCH et al., 2016).

Estudos sobre correlações entre taxa de crescimento e densidade da madeira denotam correlações positivas, negativas ou nulas. Entretanto, a negativa é a mais aceitável (maior taxa de crescimento proporciona menor densidade) (FERREIRA; KAGEYAMA, 1978). Neste sentido, ambientes com condições favoráveis ao crescimento estimulam a atividade fisiológica e cambial das plantas de modo a propiciar uma intensa divisão celular, sem que haja espessamento da parede celular (OLIVEIRA et al., 2012), o que acarreta em uma menor densidade da madeira (BERGÈS; NEPVEU; FRANC, 2008).

A alta precipitação estimula a frequência da atividade cambial que resulta no crescimento das árvores e influencia negativamente na densidade básica da madeira (MENESES et al., 2015).

A precipitação também é associada ao crescimento em diâmetro. Para Sette Junior et al. (2010) ao observarem o incremento em diâmetro do tronco de árvores

de *Eucalyptus grandis*, verificaram como condições favoráveis ao máximo crescimento do tronco das árvores a alta precipitação, baixo déficit de pressão de vapor, disponibilidade de água no solo e horas de luz/dia, que resultou no aumento da taxa de fotossíntese. Logo, foi induzida a maior taxa de divisão das células cambiais com o conseqüente aumento do crescimento do tronco. Além do diâmetro, a altura é influenciada pela precipitação (GIVNISH et al., 2014) mas não exclusivamente, visto que é correspondente a idade da árvore (MIFSUD, 2002).

Quanto ao crescimento em volume, para Fernandes et al. (2011) ao avaliarem a influência da produtividade de 15 clones de eucalipto em 4 locais com diferentes precipitações pluviométricas na densidade básica da madeira, constataram uma correlação negativa entre estas variáveis (menor produtividade e maior densidade) em relação ao aspecto ambiental. Assim, foi sugerido a destinação de clones com baixa densidade básica e alta produtividade para regiões de menor potencial de crescimento de modo a melhorar a produtividade e a densidade atender as exigências para determinado produto.

A temperatura também afeta a densidade da madeira, pois seu aumento ocasiona a diminuição da viscosidade da água que demanda uma menor quantidade e tamanho dos vasos para o transporte de água do fuste até as folhas. Para isto, há diminuição da expansão celular, aumento da espessura da parede celular e conseqüentemente, o aumento da densidade da madeira que está negativamente relacionado a condutividade hidráulica (THOMAS; MONTAGU; CONROY, 2004; THOMAS; MONTAGU; CONROY, 2007).

A condutividade hidráulica depende da área transversal do fuste ocupada pelos vasos, tamanho e número desses vasos. Em locais secos, as árvores apresentam o decréscimo dessa condutividade associado ao aumento da densidade da madeira, ambos em resposta ao estreitamento dos diâmetros dos vasos e aumento de suas frequências. Por outro lado, em locais úmidos, as árvores possuem vasos mais largos e com isso o transporte de água é mais eficiente, o qual permite maior condutância estomática e taxas fotossintéticas, de modo a induzir o crescimento das árvores (POORTER et al., 2010; PFAUTSCH et al., 2016).

Outros fatores como a ocorrência de ventos, copa irregular ou inclinação do terreno (encostas) também afetam o desenvolvimento das árvores, que quando

submetidas a estas situações tornam-se inclinadas. Assim, o incremento de madeira ocorre de modo anormal com formação de madeira de reação, devido a tentativa dessas árvores retornarem à posição ereta (MONTEIRO et al., 2010; BOSCHETTI et al., 2015).

Um dos indicativos da madeira de reação é a alta densidade básica da madeira (MONTEIRO et al., 2010) que pode ser explicado pelo maior espessamento das paredes de fibras e conseqüentemente, em uma maior resistência da madeira. Estas são características de árvores com maior tolerância a ação dos ventos (BRAZ et al., 2014). Áreas de baixada (menor altitude) devido não ocorrência de ventos climatológicos podem apresentar árvores mais susceptíveis a danos (BRAZ et al., 2017).

3.3 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNA) são utilizadas em diversas áreas de conhecimento, pelo bom desempenho na resolução de problemas como aproximação de funções, classificação e predição. Entretanto, além de considerar o desempenho obtido, é primordial que o usuário tenha habilidade para interpretar os resultados, de modo a auxiliar na tomada de decisões (BRAGA; CARVALHO; LUDERMIR, 2007).

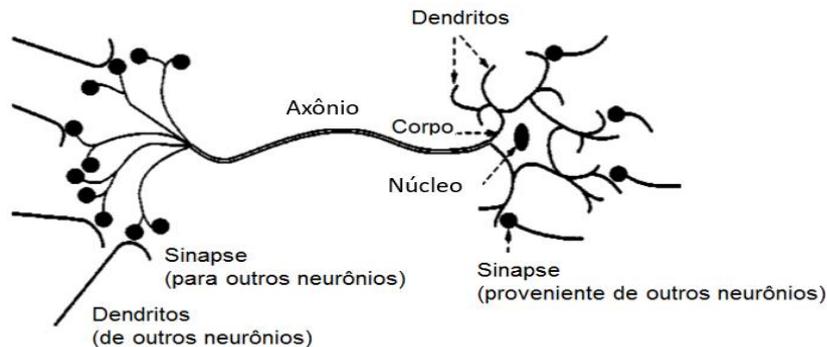
As RNA são sistemas computacionais paralelamente distribuídos, os quais são formados por unidades de processamento simples, também denominadas de neurônios artificiais, que efetuam certas funções matemáticas, em geral, não-lineares. Estas unidades são arranjadas em uma ou mais camadas interligadas por uma grande quantidade de conexões, que podem estar associadas a pesos. Estes pesos armazenam o conhecimento demonstrado no modelo e também ponderam as entradas referentes a cada neurônio da rede. A estrutura e o funcionamento das redes baseiam-se no cérebro humano (BRAGA; CARVALHO; LUDERMIR, 2007).

Haykin (2001) ressalta duas características que aproximam as Redes Neurais Artificiais ao cérebro humano:

- O conhecimento da rede é adquirido por meio do processo de aprendizagem;
- O conhecimento adquirido é armazenado em função das forças de conexão entre os neurônios, denominadas de pesos sinápticos.

A principal célula do cérebro humano é o neurônio biológico (Figura 1), tipicamente constituído por um corpo celular, axônio tubular e um emaranhado de dendritos (MEHROTRA; MOHAN; RANKA, 1997). Os dendritos desempenham a função de receber informações provenientes de outros neurônios ou de outras partes do corpo na forma de impulsos nervosos, conduzindo-as até o corpo celular. No corpo celular são processadas as informações, gerando-se novos impulsos ou não (conforme o limiar de ativação ou bias), os quais são transmitidos pelo contato entre o terminal axônico de um neurônio e os dendritos de outros neurônios, denotando assim a sinapse. As sinapses permitem a união dos neurônios resultando na formação de redes neurais (BRAGA; CARVALHO; LUDERMIR, 2007).

Figura 1 – Neurônio biológico.

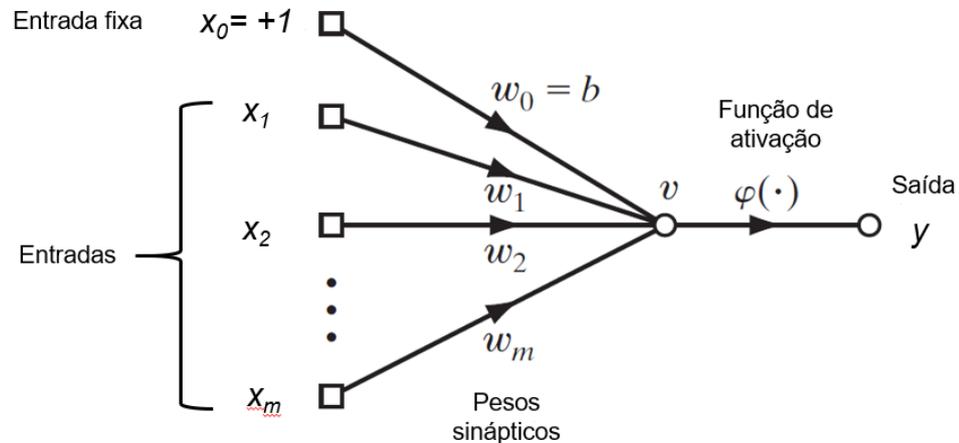


Fonte: Adaptado de Mehrotra, Mohan e Ranka (1997).

A partir do conhecimento da estrutura e funcionamento do neurônio biológico, tem-se a sua representação matemática por meio de um neurônio artificial (Figura 2). Os sinais de entradas da rede (similar aos impulsos nervosos) representados por X são ponderados pelos pesos sinápticos (w) que podem ser positivos ou negativos em função das sinapses correspondentes serem excitatórias ou inibitórias, respectivamente; o *bias* (b) é tratado como um peso sináptico que pondera uma entrada fixa igual a +1. Ocorre o somatório de cada entrada multiplicada pelo seu respectivo peso, resultando no valor de v (combinação linear dos sinais de entrada),

este valor é transformado pela função de ativação ($\varphi(\cdot)$) que gera a saída única (y) do neurônio artificial (SILVA; SCHIMIDT, 2016; HAYKIN, 2009).

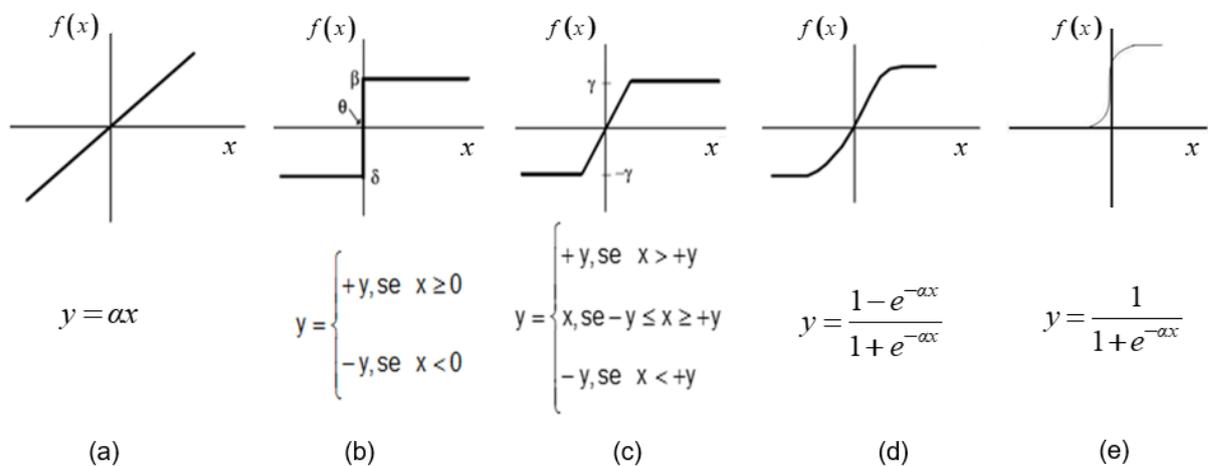
Figura 2 – Modelo de um neurônio artificial.



Fonte: Adaptado de Haykin (2009).

De acordo com Braga, Carvalho e Ludermir (2007) há diferentes tipos de funções de ativação, as quais são exemplificadas graficamente: a função linear (3a), degrau (3b), rampa ou step (3c) e sigmoideal, cujos tipos são tangente hiperbólica (3d) e logística (3e).

Figura 3 – Tipos de função de ativação.

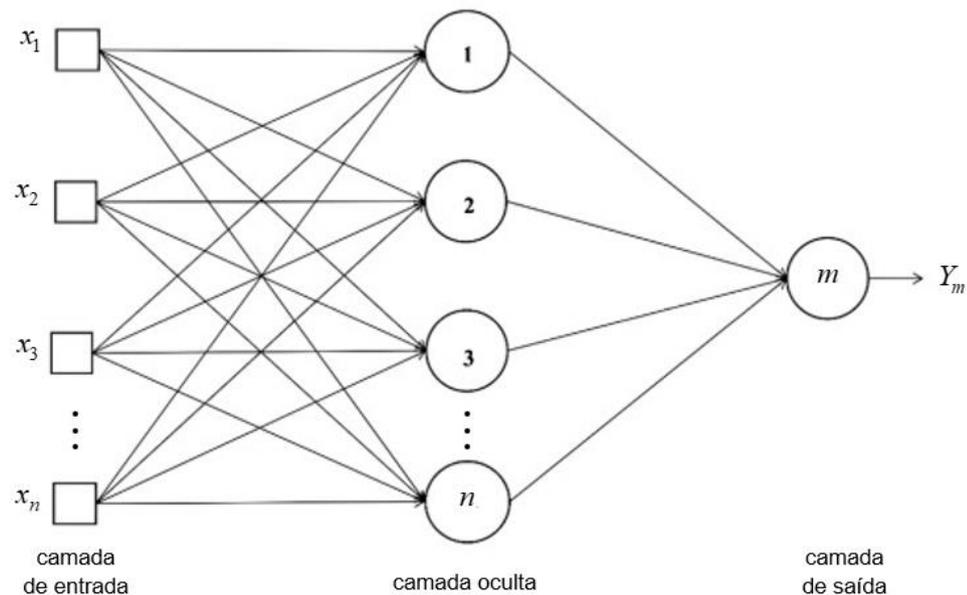


Fonte: Adaptado de Cardon e Muller (1994).

A arquitetura *Multilayer Perceptron* (MLP) é a mais usual, dentre os principais tipos de redes com arquiteturas alimentadas adiante com múltiplas camadas (do inglês, *Multilayer Feedforward Networks*) (Figura 4) (MORAIS JUNIOR; LIRA; BRITO, 2016), com destaque para aplicações, como exemplo, aproximação universal de funções, reconhecimento de padrões, identificação e controle de processos, previsão de séries temporais e otimização de sistemas (SILVA; SPATTI; FLAUZINO, 2010)

Esta arquitetura baseia-se no modo em que os neurônios estão distribuídos, sendo a mesma composta pela camada de entrada, cuja função é receber as informações oriundas do meio externo; pelas camadas ocultas em que os neurônios extraem atributos vinculados ao processo a ser inferido e também pela camada de saída, que produz os resultados finais provenientes dos processamentos realizados pelos neurônios das camadas anteriores (SILVA; SPATTI; FLAUZINO, 2010).

Figura 4 – Exemplificação da arquitetura *feedforward* de camadas múltiplas.



Fonte: Reis et al. (2018).

É importante destacar a relevância da habilidade de aprendizado da RNA, a qual se baseia na introdução de amostras (padrões ou exemplos) que expressam o comportamento do problema e após ter compreendido a relação entre as entradas e saídas, permitem generalizar soluções. O treinamento da rede é realizado por meio

da adoção de passos ordenados visando o ajuste de pesos dos neurônios (SILVA; SPATTI; FLAUZINO, 2010).

O treinamento para arquitetura MLP é denominado de treinamento supervisionado (SILVA; SPATTI; FLAUZINO, 2010), em que o supervisor externo disponibiliza para a rede, a entrada e saída almejadas (BRAGA; CARVALHO; LUDERMIR, 2007). Estes consistem nos dados de treinamento, que é fornecido à MLP que calcula a saída da rede, a qual é confrontada com a saída almejada. Logo, o erro resultante é realimentado por meio das camadas, sendo os pesos modificados conforme o algoritmo de aprendizagem ou treinamento, de modo iterativo até reduzir o erro para os dados de treinamento a níveis predefinidos pelo usuário (GOLDSCHMIDT; PASSOS, 2005).

Tem-se obtido resultados satisfatórios para resolução de problemas difíceis por meio da aplicação de MLP treinada com o algoritmo popularmente conhecido como algoritmo de retropropagação do erro (*error backpropagation*) (HAYKIN, 2001). De acordo com Silva, Spatti e Flauzino (2010), o aperfeiçoamento do algoritmo *error backpropagation* para melhoria de sua convergência, resultou nas seguintes versões deste algoritmo:

- *Levenberg-Marquardt*

É uma variação do *error backpropagation* que utiliza aproximação pelo método do Newton. Este pode ser aplicado para problemas de aproximação de função em redes com menos de 1.000 pesos cuja precisão deve ser alta e por sua rapidez em aprendizagem. Também é usual quando se realiza o treinamento para redes de tamanho pequeno ou médio em que não há limitação de memória e tempo computacional. Portanto, diante da indisponibilidade de capacidade computacional suficiente, pode-se encontrar bastante opções de algoritmos disponíveis (BEALE; HAGAN; DEMUTH, 2016).

- *Resilient-propagation*

Esta variação resulta no dinamismo da taxa de aprendizagem, pois quando apresentados sinais de gradiente iguais em duas iterações sucessivas é permitido aumentar a taxa de aprendizagem visto que está distante de um ponto de mínimo da função erro. Entretanto, quando se trata de sinais diferentes, a taxa deve ser reduzida pois, denota-se que o ponto mínimo da função foi ultrapassado (SILVA; SPATTI; FLAUZINO, 2010).

- *Error backpropagation com Momentum*

O termo momentum consiste em uma variação simples do *error backpropagation* por meio da inserção de apenas um parâmetro para ponderação do quanto as matrizes sinápticas passaram por alterações entre duas iterações anteriores e sucessivas. Logo, confere maior eficiência na convergência da rede, visto que é considerado o afastamento da solução atual em relação à solução final (SILVA; SPATTI; FLAUZINO, 2010).

Para o aprendizado das redes, divide-se o conjunto total de amostras disponíveis em subconjunto de treinamento composto por 60 a 90% das amostras e subconjunto de teste com aproximadamente 10 a 40% do conjunto total de amostras. Em relação ao subconjunto de treinamento, cada abordagem do mesmo tendo em vista ajustar os pesos dos neurônios é referida como época ou ciclo de treinamento (SILVA; SPATTI; FLAUZINO, 2010).

As RNA possuem capacidade de generalização, diante disto vale evidenciar que uma boa generalização consiste no mapeamento entrada-saída correto ou aproximadamente correto das saídas calculadas pela rede para dados não utilizados no treinamento da rede (subconjunto de teste). Em contraposição, o aprendizado de um número excessivo de exemplos entrada-saída pode levar a rede ao excesso de treinamento, implicando na sua perda de capacidade em generalizar padrões de entrada-saída semelhantes (HAYKIN, 2001).

3.4 Seleção de variáveis

3.4.1 Força bruta com aplicação de Redes Neurais Artificiais

É notória a relevância da busca por variáveis e interações que resultam em predições mais precisas (HSU, 2015). Para tanto, uma abordagem ampla de modo a obter a solução ideal consiste em explorar todas as possíveis combinações de variáveis 2^p , em que p é a dimensão do conjunto de dados, denominada de força bruta ou busca exaustiva. No entanto, para valores de p muito elevados, esta tarefa

torna-se inexecuível manualmente (ALMEIDA; PARK; CARDOSO, 2004; CATENI; COLLA, 2016).

A aplicação das RNA possibilita a identificação da influência das variáveis de entrada na predição da saída (variável resposta), pois, de acordo com Morfidis e Kostinakis (2017), o desempenho das RNA está fortemente atrelado a escolha destas variáveis.

3.4.2 *Random Forests*

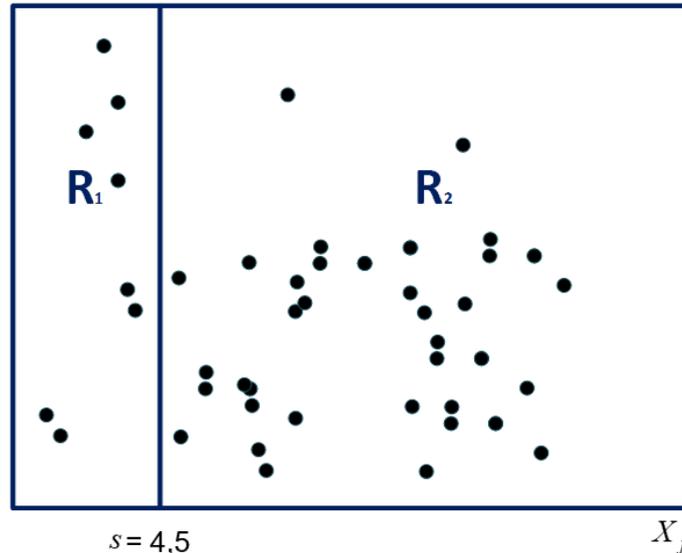
As florestas aleatórias (do inglês, *Random forests* – RF) consistem em uma coleção de árvores de decisão. Cada árvore é construída a partir de valores de um vetor aleatório, que é amostrado de forma independente e com distribuição uniforme para todas as árvores na floresta (BREIMAN, 2001).

A árvore de decisão é cultivada por meio de divisões no conjunto de dados. Cada divisão é gerada pela escolha de uma variável de divisão (X_j) e pela determinação do seu ponto de corte (S), que são estabelecidos por um algoritmo guloso. O algoritmo busca o melhor par (X_j, S) para realizar a divisão que permite a maior redução possível da soma dos quadrados do resíduo – SQR (Equação 1) Este par origina duas regiões (R_1 e R_2), em que $R_{1(j,s)} = \{X | X_j \leq s\}$ e $R_{2(j,s)} = \{X | X_j > s\}$, ou seja, um nó pai produz dois nós filhos (Figura 5). Este processo é realizado de modo iterativo para cada região, o qual define a forma da árvore (HASTIE; TIBSHIRANI; FRIEDMAN, 2008; JAMES et al., 2013).

$$SQR = \sum_{i: x_i \in R_1(j,s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i: x_i \in R_2(j,s)} (y_i - \hat{y}_{R_2})^2 \quad (1)$$

em que: \hat{y}_{R_1} = resposta média para as observações de treino em $R_{1(j,s)}$ e \hat{y}_{R_2} = resposta média para as observações de treino em $R_{2(j,s)}$.

Figura 5 – Divisão da árvore de regressão.



Fonte: Adaptado de James et al. (2013).

A árvore de decisão apresenta desvantagens como: alta variância, pouca robustez e baixa capacidade preditiva, o que as torna inferiores quando comparadas a outros métodos de regressão. Entretanto, o uso de muitas árvores de decisão, que é o fundamento dos métodos, *bagging* (no português, ensacamento), RF e *boosting* (no português, impulso), proporcionam melhorias no desempenho preditivo dos modelos (JAMES et al., 2013).

O *bagging* (*Bootstrap aggregation*) consiste em múltiplas versões de um preditor para obtenção de um preditor agregado. As versões são produzidas por meio do *bootstrap*, que forma diferentes conjuntos de dados a partir da amostragem com substituição das observações de um conjunto de dados original. Para a predição da variável resposta, quando quantitativa, toma-se a média das diferentes versões de preditores, o que resulta na redução da variância, e conseqüentemente no aperfeiçoamento da predição (BREIMAN, 1996; JAMES et al., 2013).

Este é um caso especial de RF, em que o número de variáveis preditoras consideradas em cada divisão (m) é igual ao número total de variáveis preditoras disponíveis (p). Então, na presença de uma variável preditora muito forte no conjunto de dados, juntamente com outras moderadamente fortes, a maioria ou todas as árvores usarão a forte variável preditora na divisão superior. Em decorrência disto, todas estas serão bastante semelhantes entre si. Portanto, as predições das mesmas estarão altamente correlacionadas. A média de muitas quantidades

altamente correlacionadas não leva a uma grande redução de variância como a média de muitas quantidades não correlacionadas (JAMES et al., 2013).

As RF foram propostas por Breiman (2001) a partir da modificação no *bagging* (HASTIE; TIBSHIRANI; FRIEDMAN, 2008), que resultou em uma melhoria por meio de um pequeno ajuste aleatório que descorrelaciona as árvores. Para a construção de uma RF, em cada divisão na árvore, o algoritmo nem sempre considera a maioria das variáveis preditoras disponíveis, visto que força cada divisão a considerar apenas um subconjunto das mesmas ($m \leq p$), logo as outras variáveis preditoras terão mais chances de serem escolhidas, e conseqüentemente, a média das árvores resultantes será menos variável e mais confiável (JAMES et al., 2013).

A escolha de m , denotada por *mtry*, consiste no primeiro parâmetro do ajuste do modelo. Quanto menor o seu valor, menor a correlação entre as árvores e maior a redução da variância. Além deste, outros parâmetros também podem ser ajustados, como o número de árvores (*nree*) da floresta e o tamanho da árvore pelo número de divisões (ou nós) em cada árvore (*nsplit*) e pela quantidade máxima de nós terminais (ou folhas) das árvores que a floresta pode ter (*max nodes*). As mudanças nas configurações dos parâmetros de ajuste, geralmente, fornecem resultados semelhantes, embora a contribuição de cada parâmetro seja diferente para o viés e a variância das predições e impactam na qualidade da solução final (GOLDSTEIN et al., 2011).

As RF são difundidas em diferentes áreas de pesquisa como ciência social, clínica, econometria (HAPFELMEIER; ULM, 2013), agrárias (VIEIRA; OLIVEIRA; PAIVA, 2015; CARVALHO JUNIOR et al., 2016), bioinformática (BOULESTEIX et al., 2012), genética e biologia molecular (GOLDSTEIN et al., 2011).

Em geral, os trabalhos desenvolvidos com aplicação de RF buscam compreender as interações entre as variáveis, identificar as variáveis explicativas e melhorar a capacidade de predição, visto que em algumas situações podem ser superiores aos métodos consolidados, como exemplo, regressão, análise discriminante linear, máquinas de vetor de suporte e redes neurais (HAPFELMEIER; ULM, 2013). Também são usuais na quantificação da importância das variáveis preditoras, identificação e seleção das mais relevantes para problemas de regressão e classificação (LOUPPE et al., 2013; STROBL et al., 2007; BASTOS; NASCIMENTO; LAURETTO, 2014; JANITZA; TUTZ; BOULESTEIX, 2016;

GREGORUTTI; MICHEL; SAINT-PIERRE, 2017, DÍAZ-URIARTE; DE ANDRÉS, 2006).

RF é uma técnica popular e eficiente, que pode ser aplicada em problemas de regressão e classificação (GENUER; POGGI; TULEAU-MALOT, 2010).

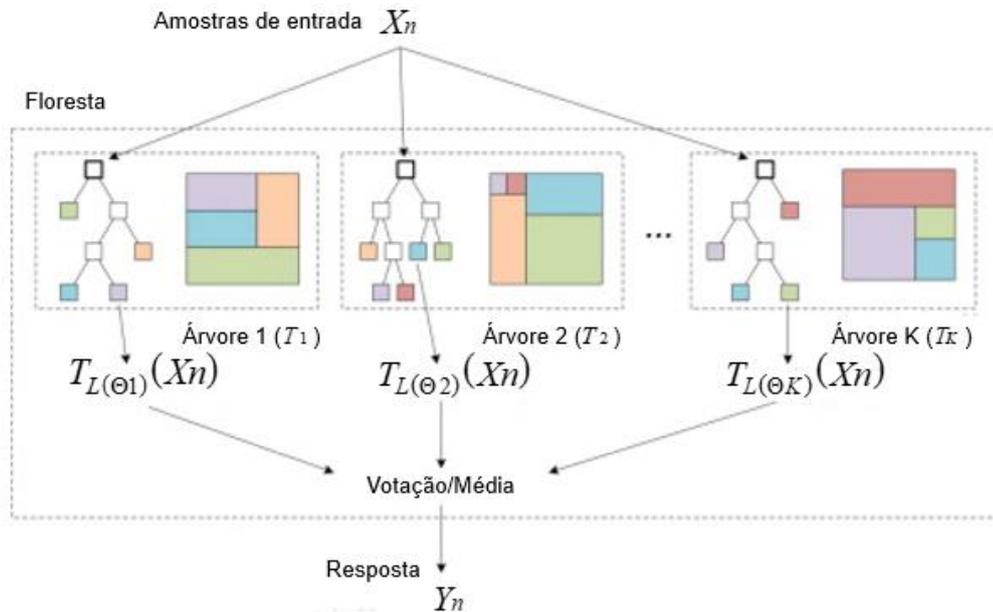
O funcionamento desta (Figura 6) baseia-se na retirada de uma diferente amostra *bootstrapp* $L(\theta)$ de tamanho n a partir de um conjunto treinamento de tamanho N , que é modificado para cada nova árvore. Logo, cada preditor de árvore $T_{L(\theta)}$ depende do vetor aleatório θ , que refere-se às amostras ensacadas (do inglês, *bagged*) do conjunto de aprendizagem original L . Para a classificação (Equação 2), uma floresta aleatória apresenta um voto de classe de cada árvore e, em seguida, classifica usando o voto majoritário. Já para a regressão (Equação 3), as previsões de cada árvore em um ponto alvo x_n são simplesmente calculadas como média (BREIMAN, 1996).

$$y'_\eta = f(x_\eta) = \text{voto majoritário} \left\{ T_{L(\theta_k)}(x_\eta) \right\}_1^K \quad (2)$$

$$y'_\eta = f(x_\eta) = \frac{1}{K} \sum_{k=1}^K T_{L(\theta_k)}(x_\eta) \quad (3)$$

em que: y'_η = resposta predita para a amostra x_η e K = número de árvores do conjunto.

Figura 6 – Funcionamento da *Random Forest* para classificação e regressão.



Fonte: Auret e Aldrich (2012).

A RF destacou-se como uma ferramenta excepcional em diferentes áreas científicas, o que pode ser atribuído a sua aplicabilidade em uma diversidade de problemas (LOUPPE et al., 2013), não demandar que o usuário tenha uma exaustiva preocupação com o ajuste de parâmetros ou outras questões relacionadas à especificação do modelo (AURET; ALDRICH, 2012), sua capacidade de construir modelos precisos e também em fornecer medidas de importância da variável (do inglês, *variable importance measures* – VIMs) (LOUPPE et al., 2013).

É notória a relevância desta última tarefa, pois em inúmeras situações objetiva-se, além da predição de uma resposta, identificar quais as variáveis preditoras tem maior influência nesta predição (LOUPPE et al., 2013). As medidas de importância variável (VIMs), podem ser obtidas automaticamente para cada variável preditora dentro do algoritmo de RF (BOULESTEIX et al., 2012).

A RF também merece destaque pela possibilidade de lidar com alta dimensão dos dados, variáveis preditoras altamente correlacionadas, captura de padrões de associação não-linear entre variáveis preditoras e respostas (BOULESTEIX et al., 2012).

A RF padrão tem duas medidas distintas para obtenção da importância da variável (VIMs): a importância de Gini (do inglês, *gini importance* – gVI) e a

importância da variável baseada na permutação (do inglês, *variable importance measure based on permutation* – pVI) (BOULESTEIX et al., 2012).

Para a obtenção da medida de importância de Gini, o classificador RF usa uma função de divisão denominada de “Índice de Gini” de modo a estabelecer qual variável dividir durante a fase de aprendizagem da árvore. Este índice mede o nível de impureza ou desigualdade das amostras atribuídas a um nó filho após a divisão do nó pai. Por exemplo, no caso da classificação binária, em que há duas classes, p representa a fração de exemplos positivos atribuídos a um determinado nó k e $1 - p$, a fração de exemplos negativos. Logo, o índice de Gini em m é definido conforme a Equação 4 (QI, 2012).

$$G_k = 2p(1-p) \quad (4)$$

Sempre que a divisão de um nó é feita usando uma determinada variável, o valor de Gini para os dois nós descendentes é menor que o nó pai. O valor de importância de Gini de uma variável em uma única árvore é então definido como a soma da redução do índice Gini (do nó pai para o nó filho) sobre todos os nós nos quais a variável específica é usada para dividir. A importância na floresta é definida como a soma ou a média de seu valor de importância entre todas as árvores na floresta (QI, 2012).

A importância da variável baseada na permutação é obtida em cada árvore (Equação 7) pela diferença entre o erro quadrático médio (do inglês, *mean squared error* – *mse*) das amostras dentro da bolsa (*bagging*) e fora da bolsa (do inglês, *out-of-bag* – OOB), conforme as Equações 5 e 6, respectivamente.

$$mse = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (5)$$

$$mse^{OOB} = \frac{1}{n_{OOB}} \sum_{i=1}^{n_{OOB}} (Y_i - \hat{Y}_i^{OOB})^2 \quad (6)$$

$$pVI = \omega_j(T_L) = mse(T_{L(\theta)}) - mse(T_{L^j_{OOB}(\theta)}) \quad (7)$$

em que: pVI = importância da variável por árvores; n = número de observações; Y_i = valores observados; \hat{Y}_i = valores estimados; n_{OOB} = número de observações para a amostra OOB; \hat{Y}^{OOB} = valores estimados para as observações da amostra OOB e $L^j_{OOB}(\theta)$ = amostra de aprendizagem OOB com a variável j permutada.

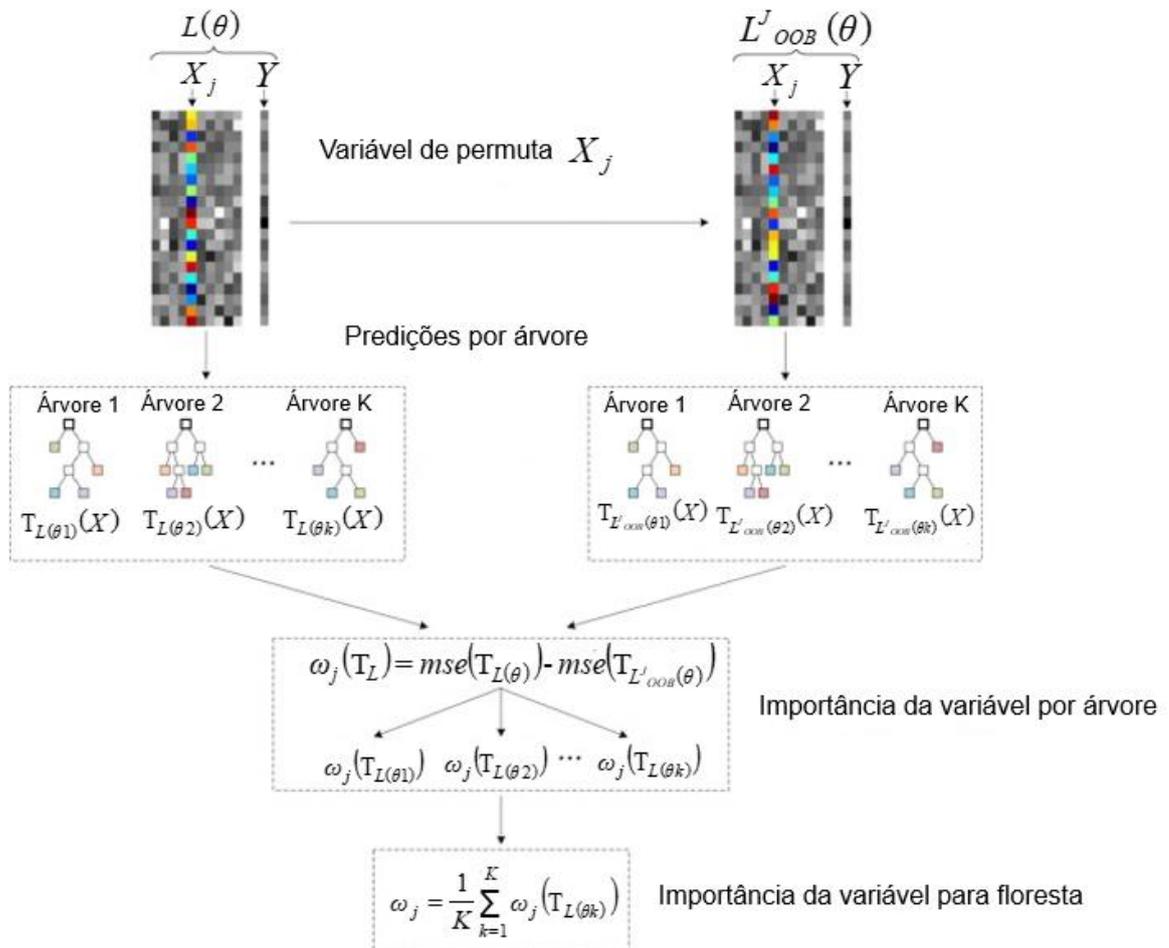
A média das medidas de importância das árvores individuais resultam na importância da variável para a floresta, que pode ser obtida conforme a Equação 8:

$$\omega_j = \frac{1}{K} \sum_{k=1}^K \omega_j(T_{L(\theta_k)}) \quad (8)$$

A obtenção da importância da variável X_j para modelos de regressão com aplicação de florestas aleatórias, pode ser demonstrada conforme a Figura 7. Para isto, outras variáveis de entrada (excluindo X_j) são representadas pelas matrizes em escala de cinza, enquanto a coluna de escala cinza consiste no vetor de resposta Y . Já, o vetor colorido é o intervalo de valores de X_j . Logo, um modelo de floresta aleatória foi construído a partir do conjunto de dados de treinamento original, denotado por $L(\theta)$ (AURET; ALDRICH, 2012; NGUYEN; HUANG; NGUYEN, 2015).

Outro conjunto de aprendizagem $L^j_{OOB}(\theta)$ é construído permutando aleatoriamente os valores variáveis de X_j . As previsões da resposta da árvore são então feitas para os conjuntos de treinamento originais e permutados, para todas as árvores K do conjunto e para todas as amostras N . A importância da variável X_j por árvore é calculada com a diferença do erro quadrático médio (mse) das N amostras dos dados originais e permutados. A importância média de X_j para floresta é obtida com a média das K árvores. (AURET; ALDRICH, 2012). Logo, a permutação de um preditor "importante" deverá aumentar o erro OOB, de modo a gerar um pVI elevado (BOULESTEIX et al., 2012).

Figura 7 – Demonstração da obtenção da importância da variável baseada na permutação.



Fonte: Auret e Aldrich (2012).

No gVI, um preditor "importante" é frequentemente selecionado dentro de um subconjunto de preditores, obtendo um alto gVI, ao contrário do pVI, que fundamenta-se na precisão da predição e não no critério de divisão. Assim, este é obtido pela diferença entre o erro OOB (do inglês, *out-of-bag error*) resultante de um conjunto de dados obtido por permutação aleatória do preditor de interesse e o erro OOB resultante do conjunto de dados original.

Embora o pVI seja mais frequentemente usado na prática, a questão da escolha do tipo VIM e as propriedades desses VIMs ainda são assuntos de pesquisas recentes (BOULESTEIX et al., 2012), valendo ressaltar que o uso de gVI é recomendável apenas para classificação (GOLDSTEIN et al., 2011), além de apresentar tendenciosidade na preferência por preditores qualitativos, quando são

constituídos por muitas classes, o que é menos evidente quando usa-se o pVI (STROBL et al., 2007).

3.4.3 Algoritmo de *Garson*

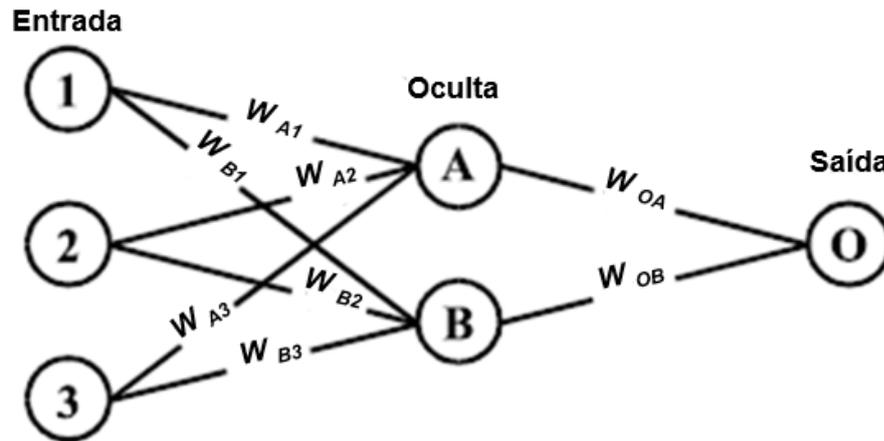
Há diferentes medidas para quantificar a contribuição relativa das variáveis de entrada de uma Rede Neural Artificial do tipo *Multilayer Perceptron (MLP)* (FISCHER, 2015). Estas medidas são passíveis de aplicação em vários campos científicos (PENTÓS, 2016). Alguns exemplos são: o algoritmo de *Garson*, os pesos globais de conexão e as derivadas parciais. (OLIVEIRA; PEDROLLO; CASTRO, 2015).

Diante desta ampla variedade de medidas de contribuição da variável, ainda não há um consenso sobre a mais confiável. No entanto, Fisher (2015) recomenda que estudos empíricos que visam quantificar a importância relativa das variáveis de entrada devem preferir o algoritmo de *Garson* em relação aos pesos globais de conexão proposto por Olden et al. (2004).

O algoritmo desenvolvido por *Garson* (1991) consiste em um método de divisão dos pesos de conexão da rede neural para determinar a importância relativa de cada variável de entrada em relação à resposta predita. Este, depois foi modificado por Goh (1995) por meio do uso dos valores absolutos dos pesos de conexão, de modo a assegurar que não houvesse influência contrária em função de valores positivos e negativos (ZHOU et al., 2015). Logo, não fornece a direção da relação entre as variáveis de entrada e de saída (OLDEN; JACKSON, 2002).

O funcionamento do algoritmo de *Garson* por meio da aplicação da RNA, cuja estrutura pode ser exemplificada na Figura 8, fundamenta-se no uso dos pesos (W) entre os neurônios da camada de entrada (1, 2 e 3) e oculta (A e B) e entre a camada oculta e de saída (O) (OLDEN; JACKSON, 2002).

Figura 8 – Demonstração dos pesos resultantes de ajustes com aplicação de RNA.



Fonte: Olden e Jackson (2002).

Para o cálculo das importâncias relativas de cada variável, devem ser atendidas às seguintes etapas descritas por Olden e Jackson (2002):

- Obtenção da contribuição de cada neurônio de entrada para o neurônio de saída, via neurônios da camada oculta, calculados a partir da multiplicação dos pesos entre as camadas entrada-oculta e oculta-saída, conforme a Equação 9 exemplificando a conexão entre a entrada 1 e a saída via neurônio A da camada oculta.

$$C_{A1} = W_{A1} W_{OA} \quad (9)$$

- Determinação da contribuição de cada neurônio de entrada para a saída é calculada em relação a cada neurônio oculto (r) e posteriormente, realiza-se o somatório das contribuições de cada neurônio de entrada (S). Estes valores são obtidos, respectivamente, de acordo com as equações 10 e 11.

$$r_{A1} = \frac{|C_{A1}|}{|C_{A1}| + |C_{A2}| + |C_{A3}|} \quad (10)$$

$$S_1 = r_{A1} + r_{B1} \quad (11)$$

- Cálculo da importância relativa para cada variável entrada conforme a Equação 12.

$$RI_1(\%) = \frac{S_1}{(S_1 + S_2 + S_3)} 100 \quad (12)$$

4 MATERIAL E MÉTODOS

4.1 Descrição dos dados

O estudo foi realizado com dados, cedidos por empresa florestal, oriundos de testes clonais em povoamentos de eucalipto, distribuídos na região leste do estado de Minas Gerais.

As informações cadastrais (variáveis qualitativas) foram organizadas em: identificação das 7 sub-regiões dos plantios, 2 classes de relevo (encosta e baixada) e 52 classes de material genético, em que foram utilizados 386 registros em nível de árvore (Tabela 1).

As variáveis quantitativas do inventário florestal contínuo consideradas neste estudo foram idade (anos), *DAP* (cm), altura total (m) e volume ($\text{m}^3 \cdot \text{ha}^{-1}$).

As variáveis climáticas destes povoamentos foram temperatura média ($^{\circ}\text{C}$), precipitação total média (mm), velocidade do vento ($\text{m} \cdot \text{s}^{-1}$), déficit hídrico (mm) e déficit pressão de vapor (hPA) advindas de sete estações climáticas dispostas na região de estudo. Desta forma, para os clones pertencentes a cada teste clonal, foram utilizados os dados referentes às estações mais próximas, em termos de médias mensais com base no período entre a data do plantio e a data de amostragem da densidade da madeira (período de crescimento do povoamento). Além destas, a altitude também foi considerada.

A densidade básica da madeira foi obtida pela amostragem de um disco de madeira a 1,30 metros de altura (*DAP*) de pelo menos três indivíduos (repetições) dos clones pertencentes a cada teste clonal, sendo considerada a média para cada clone. A determinação da densidade em $\text{kg} \cdot \text{m}^{-3}$ foi realizada por meio do método de balança hidrostática, fundamentado na NBR 11941 (ABNT, 2003). A análise descritiva das variáveis quantitativas foi descrita na Tabela 1. Já na Figura 9, a variável idade apresentou alta correlação com a densidade básica da madeira, enquanto esta foi moderada para altura e volume.

Tabela 1 – Número de observações da densidade básica da madeira e estatísticas descritivas das variáveis quantitativas do Inventário florestal contínuo, das variáveis

DB x DAP				DB x HT				DB x VTCC			
Densidade observada (kg/m ³)	OBSERVAÇÕES			OBSERVAÇÕES			OBSERVAÇÕES			Densidade observada (kg/m ³)	
	DB	SUB	TOTAL	SUB	SUB	SUB	SUB	SUB	TOTAL		
600											
550											
500											
450											
400											
350											
300											
250											
200											
150											
100											
50											
0											
estatística	DAP	HT	VTCC	TM	PPT	V	DEF	DPV	ALT	DB	
Mínimo	2,5	7,4	0,142	17,4	72,9	1,0	7,2	3,8	240,0	354,1	
Média	4,5	14,5	0,1731	20,2	106,5	2,7	16,5	6,3	800,3	467,5	
Máximo	7,1	20,0	0,4042	23,3		4,4	33,2	9,3	1273,0	550,4	
Desvio	1,5	1,7	0,0946	2,1		1,2	6,2	1,6	403,9	37,4	
	DAP (cm)	HT (m)	VTCC (m ³ /ha)	HT (m)	VTCC (m ³ /ha)						

climáticas, da altitude e da densidade básica da madeira

em que: I= idade (anos); DAP= diâmetro a altura do peito (cm); HT= altura total (m); VTCC= volume total com casca (m³.ha⁻¹); TM= temperatura média (°C); PPT= precipitação total média (mm); V= velocidade do vento (m.s⁻¹); DEF= déficit hídrico (mm); DPV = déficit pressão de vapor (hPA); ALT= altitude (m), SUB= sub-regiões e DB= densidade básica da madeira (kg.m⁻³).

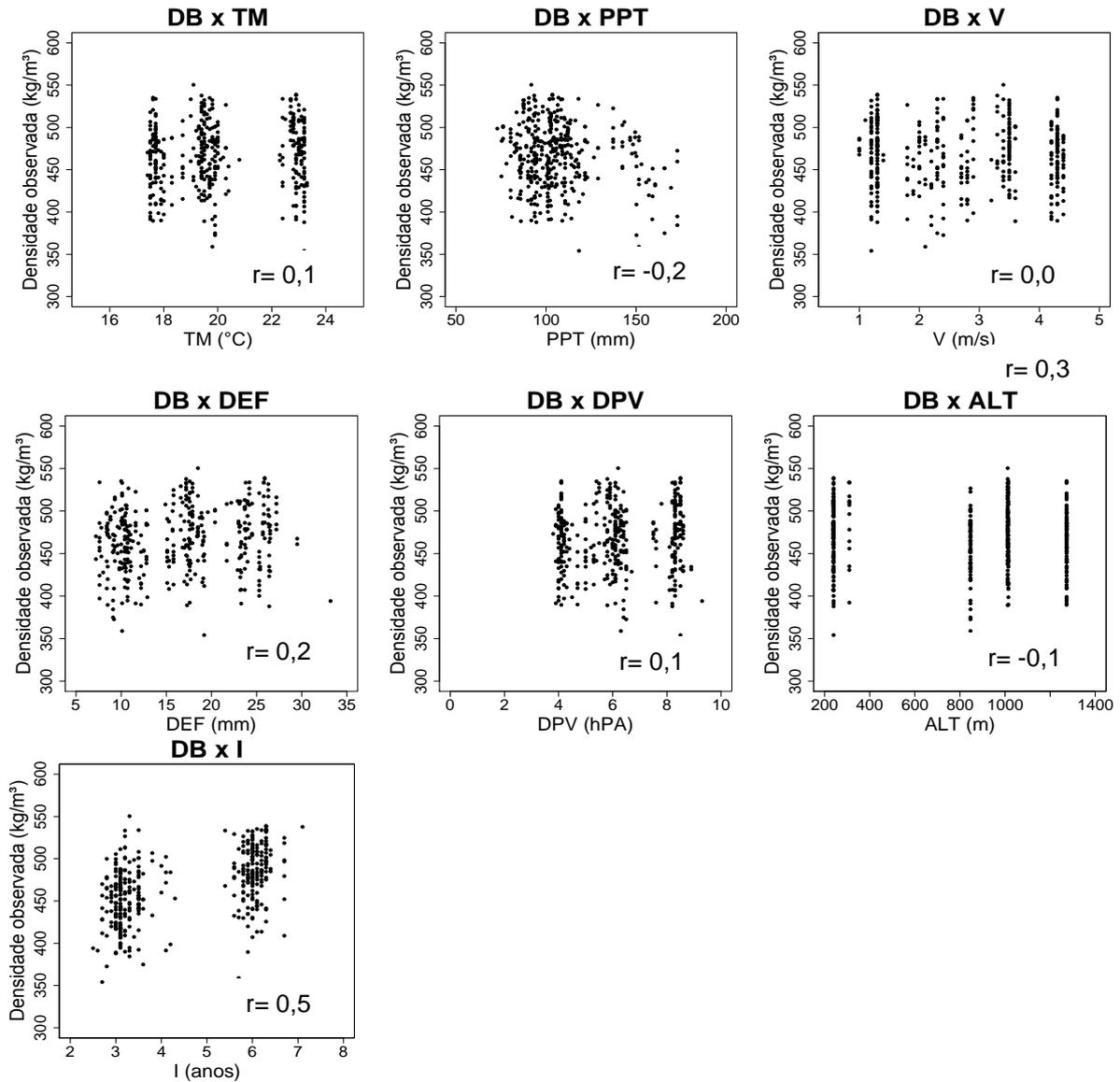


Figura 9 – Correlação entre a densidade básica da madeira e as variáveis quantitativas do estudo.

em que: I= idade (anos); DAP= diâmetro a altura do peito (cm); HT= altura total (m); VTCC= volume total com casca ($\text{m}^3 \cdot \text{ha}^{-1}$); TM= temperatura média ($^{\circ}\text{C}$); PPT= precipitação total média (mm); V= velocidade do vento ($\text{m} \cdot \text{s}^{-1}$); DEF= déficit hídrico (mm); DPV = déficit pressão de vapor (hPA); ALT= altitude (m) e DB= densidade básica da madeira ($\text{kg} \cdot \text{m}^{-3}$).

4.2 Seleção de variáveis

4.2.1 Preparação do banco de dados

A base de dados foi dividida de forma aleatória e estratificada em dois conjuntos, um para o treinamento com 248 observações e outro para a generalização com 138 observações, o que resultou em uma proporção de aproximadamente 65% para treinamento e 35% para generalização.

As variáveis qualitativas (clone, sub-região e relevo) foram utilizadas para estratificação da base, de modo a assegurar a ocorrência das mesmas no treinamento e na generalização, e também, quando necessário, a maior representatividade no treinamento. Quanto às variáveis quantitativas, garantiu-se que o treinamento abrangesse a amplitude das mesmas.

4.2.2 Processamento das Redes Neurais Artificiais

A modelagem da densidade foi feita por meio de RNA de arquitetura MLP (*Multilayer Perceptron*) com uma camada de entrada e duas camadas de processamento, sendo estas, uma camada oculta e uma de saída. O número de neurônios na camada de entrada variou conforme o número de variáveis utilizadas em cada combinação, sendo necessário um neurônio para cada variável quantitativa e um neurônio para cada classe de variável qualitativa. Dessa maneira, ocorreu a transformação dummy das variáveis qualitativas clone (52), relevo (2) e sub-região (7), ou seja, as quantidades de neurônios na camada de entrada das RNA para estas variáveis resultaram em, respectivamente, cinquenta e dois, dois e sete neurônios. A camada de saída contemplou um único neurônio, que forneceu os valores estimados da densidade básica da madeira (kg.m^{-3}) para cada árvore do conjunto de treinamento e do conjunto de generalização.

O número de neurônios da camada oculta foi definido conforme a Equação 13 estabelecida por Li, Chow e Yu (1995), com resultados arredondados para o valor mais alto.

$$N_o = \frac{\sqrt{1 + 8N_e} - 1}{2} \quad (13)$$

em que: N_o = número de neurônios na camada oculta; N_e = número de neurônios na camada de entrada.

O processamento das RNA foi realizado por meio do software Matlab versão R2017a. Foi utilizado o algoritmo de treinamento *Levenberg-Marquardt* e a função de ativação tangente hiperbólica na camada oculta e linear na camada de saída.

O critério de parada adotado foi o número de ciclos igual a 1.000 e a *early stopping*. Este último permite a verificação constante do processo de treinamento a partir do uso de subconjuntos de validação, em que o mesmo é finalizado quando ocorre o aumento do erro quadrático entre os ciclos sucessivos (SILVA; SPATTI; FLAUZINO, 2010). Logo, o treinamento finalizou-se quando um desses critérios foi alcançado.

4.2.3 Aplicação do método força bruta com base em Redes Neurais Artificiais

O método da força bruta consistiu na abordagem de todas as combinações possíveis conforme o número de variáveis de entrada (Equação 14) obtidas no inventário, climáticas e cadastrais (clone, relevo e sub-região) para estimação da densidade básica da madeira, ou seja, desde a inclusão de todas as variáveis até o uso de apenas uma (Tabela 2).

$$C_{p,m} = \frac{p!}{(p-m)!m!} \quad (14)$$

em que: p = número total de variáveis preditoras disponíveis e m = número de variáveis preditoras utilizadas como entradas da RNA ($m < p$).

Tabela 2 – Quantidade total de combinações de acordo com o número de variáveis utilizadas para os ajustes

Nº DA COMBINAÇÃO	Nº DE VARIÁVEIS	Nº TOTAL DE COMBINAÇÕES AVALIADAS
1	13	1
2	12	13
3	11	78
4	10	286

5	9	715
6	8	1.287
7	7	1.716
8	6	1.716
9	5	1.287
10	4	715
11	3	286
12	2	78
13	1	13
TOTAL		8.191

Foram treinadas 5 RNA para cada número de combinação (totalizando 40.955 redes) e selecionada a que obteve melhor desempenho e menor tempo de treinamento em cada número de combinação e dentre todas as redes avaliadas.

4.2.4 Aplicação do algoritmo de *Garson*

Foram treinadas 10 RNA para estimar a densidade básica da madeira com as configurações descritas anteriormente, em função de todas as variáveis preditoras (cadastrais, de IFC e climáticas). Finalizado o processamento das RNA e de posse dos pesos provenientes destes ajustes, iniciou-se o cálculo da importância relativa (IR%) para cada uma das variáveis na estimação da densidade, conforme a Equação 15, obtida a partir das Equações 9, 10 e 11.

$$IR(\%) = \frac{S_1}{(S_1 + S_2 + \dots + S_n)} 100 \quad (15)$$

em que: (S)= somatório das contribuições de cada neurônio de entrada (1,2,..., n).

Para a seleção das variáveis, estas foram ordenadas de maneira decrescente com base na média aritmética dos valores de importância relativa das 10 redes treinadas. A quantidade de variáveis selecionadas foi determinada conforme o número ótimo de variáveis de entrada das Redes Neurais Artificiais no método força bruta.

4.2.5 Aplicação do método *Random Forest*

Foi realizada por meio do *Random Forest* – RF, a modelagem da densidade básica da madeira (DB) em função de todas as variáveis preditoras (cadastrais, IFC e climáticas). Para o treinamento, utilizou-se dez configurações, formadas a partir da determinação dos parâmetros: número de variáveis preditoras consideradas em cada divisão (*mtry*) igual a $p/3$ (JAMES et al., 2013) e número de árvores (*ntree*), que variou entre 50 a 500, com intervalos de 50. O processamento foi realizado por meio do *software* Matlab versão R2017a.

Foram treinadas 10 configurações de RF para obtenção das estimativas das importâncias das variáveis preditoras para a floresta, conforme a Equação 16.

$$\omega_j = \frac{1}{K} \sum_{k=1}^K \omega_j(\tau_{L(\theta_k)}) \quad (16)$$

em que: $\omega_j(\tau_L)$ = importância da variável j ; K = número de árvores do conjunto.

Com base nos valores de importância, as variáveis preditoras foram ordenadas de maneira decrescente para a RF treinada.

4.3 Estimação da Densidade Básica da Madeira

A densidade básica da madeira foi estimada através de RNA pelos seguintes métodos:

- todas as combinações possíveis na estratégia da força bruta ou busca exaustiva;
- com as variáveis mais importantes selecionadas pelo algoritmo de *Garson*;
- com as variáveis mais importantes selecionadas pelo *Random Forest* e
- com as variáveis do Inventário Florestal contínuo.

4.4 Avaliação das estimativas geradas pelos métodos aplicados

A avaliação das estimativas resultantes da aplicação de Redes Neurais Artificiais nas fases de treinamento e generalização foi efetuada por meio de estatísticas e análises gráficas de resíduos. Foram consideradas as seguintes estatísticas: correlação (r) entre as densidades estimadas e observadas, a raiz do erro quadrado médio percentual (do inglês, *root mean squared error* – RMSE%) e o viés (V), que estão apresentadas na Tabela 3.

Tabela 3 – Estatísticas de avaliação das estimativas da densidade básica da madeira durante o treinamento e da generalização das Redes Neurais Artificiais

ESTATÍSTICAS	FÓRMULAS
Correlação ($r_{y\hat{y}}$)	$r_{Y\hat{Y}} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(\hat{Y}_i - \hat{Y}_m)}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2 \sum_{i=1}^n (\hat{Y}_i - \hat{Y}_m)^2}}$
Viés (%)	$V(\%) = \frac{100}{\bar{Y}} \cdot \frac{\sum_{i=1}^n Y_i - \sum_{i=1}^n \hat{Y}_i}{n}$
Raiz do erro quadrado médio (%)	$RMSE(\%) = \frac{100}{\bar{Y}} \cdot \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}}$

em que: Y_i = densidade observada; \hat{Y}_i = densidade estimada; \bar{Y} = média da densidade observada; \hat{Y}_m = média da densidade estimada e n = número de observações.

O RMSE quantifica a magnitude dos erros (WILLMOTT; MATSUURA, 2005) e os menores valores resultantes do mesmo, indicam maior exatidão (DORESWAMY; VASTRAD, 2013).

O viés aponta a diferença entre o valor esperado do estimador e o observado (BATISTA; COUTO; SILVA FILHO, 2014), sendo assim, valores de viés igual a zero e aliados a alta precisão resultam em estimativas com maior exatidão (VAN LAAR; AKÇA, 2007).

Em relação às análises gráficas, observaram-se os gráficos de dispersão dos erros percentuais e os histogramas de frequência percentual dos erros percentuais, sendo o erro percentual de cada observação ser obtido pela Equação 17.

$$Erro_i(\%) = \frac{Y_i - \hat{Y}_i}{Y_i} \cdot 100 \quad (17)$$

em que: $Erro_i(\%)$ = erro de cada observação.

5 RESULTADOS E DISCUSSÃO

5.1 Seleção das variáveis preditoras e estimativa da densidade com o método da força bruta

A avaliação das 40.955 Redes Neurais Artificiais para predição da densidade básica da madeira, das quais 5 redes foram treinadas para cada combinação (1.891) das 13 variáveis preditoras, resultou em 13 redes selecionadas (Figura 4) em cada número de variáveis de entrada possível (1 a 13). Para a seleção das redes foram demandados diferentes tempos de processamento (minutos) em um computador com processador Intel (R) Xeon (R) CPU E5 2630 V3 2,40 GHz com 32 GB de memória (RAM).

O método da força bruta não identificou a importância individual das variáveis, mas permitiu obter a combinação destas que conferiu melhor desempenho das RNA na predição da densidade básica da madeira (menor RMSE%, V% e maior $r_{y\hat{y}}$). Logo, com base nas estatísticas para os dados de validação (Tabela 4), a combinação de cinco variáveis (clone, idade, volume total, déficit hídrico e temperatura média) resultou na RNA de melhor desempenho, seguida da RNA com a combinação de nove variáveis, sendo as mesmas da anterior, acrescidas da altura total, precipitação total média, velocidade do vento e relevo.

A RNA com todas as variáveis disponíveis foi inferior as demais, com exceção da RNA com apenas uma variável, de acordo com as estatísticas observadas para os dados de validação. Portanto, é fundamental a seleção das variáveis para predição da densidade básica da madeira.

Neste contexto, Oliveira, Pedrollo e Castro (2015), avaliaram a simplificação de modelos hidrológicos baseados em RNA e verificaram que a eliminação de variáveis irrelevantes ou até mesmo redundantes, proporcionaram a redução do RMSE e o aumento do potencial preditivo. A eliminação de variáveis também propicia a minimização do custo, redução do tempo de coleta de dados, simplifica a manipulação e interpretação dos fenômenos (ALMEIDA; PARK; CARDOSO, 2004).

Tabela 4 – Combinações ótimas das variáveis predictoras conforme o número de entradas, tempo de processamento (T), arquitetura e estatísticas utilizadas para avaliar o desempenho das Redes Neurais Artificiais

Nº	T (min)	VARIÁVEIS	ARQUITETURA	TREINO			VALIDAÇÃO		
				RMSE (%)	V (%)	$r_{y\hat{y}}$	RMSE (%)	V (%)	$r_{y\hat{y}}$
13	0,6	C, I, VTCC, DEF, HT, TM, DPV, PPT, DAP, V, ALT, SUB, R	71-12-1	2,76	-1,46	0,96	5,52	-1,58	0,76
12	3	C,I,VTCC,DEF,HT,TM,PPT,DAP,V,ALT,SUB,R	70-12-1	3,06	1,07	0,94	5,10	0,48	0,76
11	15	C,I,VTCC,DEF,HT,TM,DPV,PPT,DAP,V,ALT	62-11-1	3,47	1,23	0,92	4,64	0,83	0,81
10	30	C,I,VTCC,DEF,HT,TM,DPV,DAP,V,SUB	67-12-1	3,17	-0,48	0,92	4,61	-0,81	0,81
9	57	C,I,VTCC,DEF,HT,TM,PPT,V,R	61-11-1	2,50	0,43	0,95	4,47	-0,09	0,81
8	78	C,I,VTCC,DEF,PPT,DAP,V,SUB	65-11-1	2,74	0,25	0,94	4,61	-0,04	0,80
7	82	C,I,VTCC,DEF,HT,DAP,ALT	58-11-1	3,38	0,30	0,91	4,55	-0,11	0,81
6	83	C,I,VTCC,DEF,TM,DPV	57-11-1	3,65	-0,31	0,90	4,47	-0,73	0,82
5	50	C, I, VTCC, DEF, TM	56-11-1	2,85	-0,01	0,94	4,43	-0,17	0,82
4	28	C, I, DEF, DPV	55-10-1	3,77	0,21	0,89	4,62	-0,28	0,80
3	11	C, I, TM	54-10-1	3,87	0,40	0,88	4,84	-0,40	0,78
2	1	C, I	53-10-1	4,44	-0,09	0,84	4,98	-0,47	0,77
1	0,5	C	52-10-1	5,85	-0,83	0,71	6,29	-0,59	0,58

em que: C= clone; I= idade; TM= temperatura média; DEF= déficit hídrico; DPV= déficit pressão de vapor; VTCC= volume total com casca; PPT= precipitação total média; DAP= diâmetro a altura do peito; HT= altura total; ALT= altitude, V= vento; SUB= sub-região; R= relevo; $r_{y\hat{y}}$ = coeficiente de correlação entre as densidades observadas e estimadas; V (%)= viés; RMSE (%) = raiz do erro quadrado médio; T= tempo de processamento e Arquitetura = n° de neurônios na camada entrada-n° de neurônios na camada oculta-n° de neurônios na camada de saída.

A variável clone foi selecionada em todas as combinações com melhor desempenho em termos de qualidade de estimativa no treinamento e na validação (Tabela 4). A interação do clone com as variáveis do IFC e climáticas, avaliadas neste estudo, contribuiu expressivamente na obtenção de melhores estimativas de densidade da madeira. Para Silva et al. (2009), a inclusão do clone às variáveis de entrada da RNA, DAP e HT, resultou na melhoria das estimativas de volume. Binoti et al. (2015), avaliaram a eficiência das RNA na prognose da produção de povoamentos equiâneos de eucalipto e verificaram a predominância do clone para as 12 redes pré-selecionadas.

O clone (material genético) está relacionado ao volume e a densidade da madeira. Esta última é herdada de forma aditiva (passível de forte controle genético) para espécies de *Eucalyptus* (MUNERI; RAYMOND, 2000; TAN et al., 2018). Este comportamento resulta em diferentes padrões de variação da densidade (ALZATE; TOMAZELLO FILHO; PIEDADE, 2005). Portanto, o efeito do clone na densidade básica da madeira é relevante na seleção e no melhoramento genético das espécies e clones de *Eucalyptus* plantados no Brasil (BOTREL et al., 2010; PROTÁSIO et al., 2014).

Ao considerar apenas duas variáveis para predição da densidade básica da madeira, a combinação clone e idade resultou em melhores estimativas. Isto pode ser atribuído ao efeito significativo desta interação na densidade, ou seja, há uma dependência entre estas variáveis para espécies de *Eucalyptus*, o que permite a seleção do material genético superior em uma idade específica para atender determinada finalidade industrial (PROTÁSIO et al., 2014).

Logo, para o uso apropriado torna-se fundamental a caracterização da madeira, principalmente, quando oriunda de árvores jovens devido à tendência de colheita destas em idade precoce. A maior proporção de lenho juvenil e menor de lenho adulto nestas árvores, pode ocasionar problemas na qualidade do produto final, pois, o lenho juvenil possui densidade mais baixa que o lenho adulto. A maior densidade básica do lenho adulto é atribuída ao processo de desenvolvimento das árvores com o aumento da idade, que resulta no aumento da espessura da parede das fibras e na diminuição da frequência e da dimensão de vasos (LATORRACA; ALBUQUERQUE, 2000; SETE JR. et al., 2012; MENESES et al., 2015; HSING; PAULA; PAULA, 2016).

Posterior ao clone, a idade foi a variável que obteve maior frequência nas melhores combinações, uma vez que possui interação com o ambiente, o qual estabelece as condições de crescimento para as espécies. O aumento da idade implica em maiores valores de densidade básica da madeira (MENESES et al., 2015). Em seguida, o volume total com casca e o déficit hídrico também foram frequentes nas melhores combinações (Tabela 4).

5.2 Seleção das variáveis preditoras e estimativa da densidade com o algoritmo de *Garson*

Nas 10 RNA treinadas para avaliar a importância das variáveis preditoras da densidade básica da madeira pelo algoritmo de *Garson* (Tabela 5), o RMSE (%), V (%) e o $r_{y\hat{y}}$ apresentaram valores similares no treinamento e na validação. Notou-se uma pequena queda na qualidade das estimativas durante a validação, que pode ser explicada pela baixa representatividade de algumas classes de variáveis qualitativas na base de dados. Resultados semelhantes foram constatados por Nunes e Görgens (2016) a partir do uso de RNA para predição do diâmetro em diferentes fitofisionomias (Cerrado, Floresta semi-decídua e Floresta Tropical).

Tabela 5 – Estatísticas das Redes Neurais Artificiais utilizadas para cálculo da importância relativa das variáveis preditoras da densidade básica da madeira pelo algoritmo de *Garson*

RNA	TREINO			VALIDAÇÃO		
	RMSE (%)	V (%)	$r_{y\hat{y}}$	RMSE (%)	V (%)	$r_{y\hat{y}}$
1	4,28	0,11	0,87	5,15	0,42	0,74
2	3,85	-0,48	0,88	5,59	-0,68	0,70
3	3,81	0,04	0,88	5,51	-0,16	0,72
4	3,67	-0,17	0,89	5,43	-0,47	0,73
5	4,57	-1,62	0,86	5,35	-1,29	0,74
6	4,00	-0,46	0,88	5,70	-1,55	0,72
7	4,49	1,86	0,86	6,03	1,80	0,69
8	4,27	-0,08	0,86	6,81	0,22	0,60
9	3,72	-0,29	0,89	6,00	-1,36	0,70
10	3,94	0,99	0,88	5,73	0,82	0,69

em que: $r_{y\hat{y}}$ = coeficiente de correlação entre as densidades observadas e estimadas; V (%)= viés e RMSE (%) = raiz do erro quadrado médio.

Ocorreu uma grande variação dos valores de importância relativa entre as variáveis, principalmente, diante da superioridade representada pelo clone. Além desta variação, nota-se na Tabela 6 a diferença entre os valores de máxima e mínima importância relativa para cada variável. A inconstância dos valores de importância relativa das variáveis de entrada é inevitável em RNA (PENTÓS, 2016), em que um dos motivos é atribuído aos diferentes valores iniciais aleatórios utilizados para um mesmo método (ONÃ; GARRIDO, 2014). A variabilidade da importância relativa de cada variável também depende do método utilizado (PENTÓS, 2016).

A partir das médias das importâncias relativas obtidas com a aplicação do algoritmo de *Garson* para as 10 RNA, as variáveis foram ranqueadas, de modo decrescente, conforme a influência na predição da densidade básica da madeira (Tabela 6).

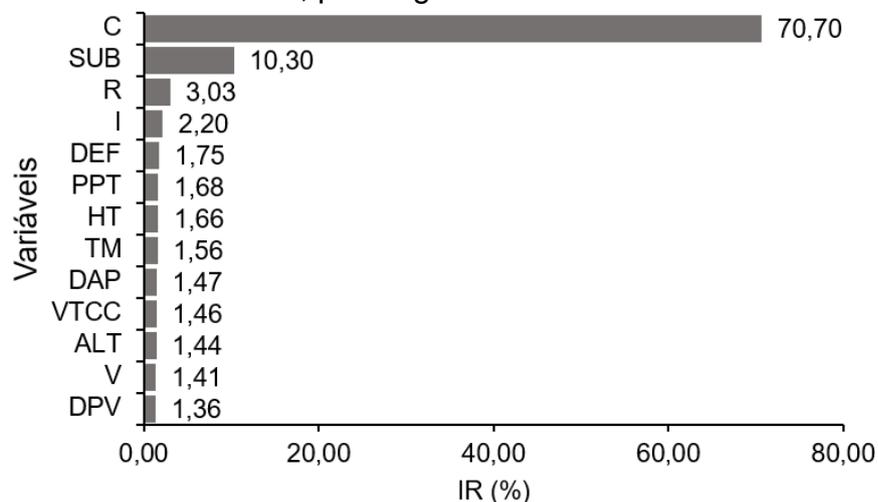
Tabela 6 – Importância relativa das variáveis predictoras (IR%) na predição da densidade básica da madeira, pelo algoritmo de *Garson*

VARIÁVEL	MÍNIMA (%)	MÉDIA (%)	MÁXIMA (%)	RANKING
C	66,40	70,70	73,96	1
SUB	9,26	10,30	11,63	2
R	2,28	3,03	4,90	3
I	1,52	2,20	3,23	4
DEF	1,03	1,75	2,55	5
PPT	1,03	1,68	2,17	6
HT	1,08	1,66	2,44	7
TM	1,04	1,56	1,97	8
DAP	0,88	1,47	2,17	9
VTCC	0,93	1,46	2,09	10
ALT	0,97	1,44	1,75	11
V	1,00	1,41	1,72	12
DPV	1,00	1,36	1,80	13

em que: C= clone; I= idade; TM= temperatura média; DEF= déficit hídrico; DPV= déficit pressão de vapor; VTCC= volume total com casca; PPT= precipitação total média; DAP= diâmetro a altura do peito; HT= altura total; ALT= altitude, V= vento; SUB= sub-região e R= relevo.

A variável clone destacou-se das demais variáveis (superior em até 50 vezes) por apresentar a maior importância relativa, seguida da sub-região. Em um nível abaixo, estão o relevo, a idade e o déficit hídrico. Já as demais variáveis, apresentaram valores menores que 1,7%. A variável com menor importância relativa foi o DPV (Figura 10).

Figura 10 – Média da importância relativa das variáveis predictoras (IR%) na predição da densidade básica da madeira, pelo algoritmo de *Garson*.



em que: C= clone; I= idade; TM= temperatura média; DEF= déficit hídrico; DPV= déficit pressão de vapor; VTCC= volume total com casca; PPT= precipitação total média; DAP= diâmetro a altura do peito; HT= altura total; ALT= altitude, V= vento; SUB= sub-região e R= relevo.

A densidade da madeira foi influenciada pela sub-região (Figura 10), uma vez que o ambiente estabelece as condições de crescimento (variáveis climáticas e de altitude) para as espécies, fato que corrobora com os estudos desenvolvidos por Kube, Raymond e Banham (2001), Latorraca e Albuquerque (2000) e Mauri et al. (2015).

Outro exemplo de condição de crescimento, o relevo, apesar de apresentar menor influência quando comparado ao clone e sub-região, contribuiu na densidade da madeira. Para Mauri et al. (2015) ao estudar clones de *Eucalyptus urophylla* x *Eucalyptus grandis*, com seis anos de idade, em diferentes condições de crescimento (relevo e altitude), os maiores valores de densidade básica foram verificados em locais com topografia (relevo) plana. Também foi concluído que a altitude não influenciou na variação desta propriedade, fato que corrobora com o resultado deste estudo, em que a altitude ocupou uma das últimas posições do ranking das variáveis.

A influência do déficit hídrico na predição da densidade da madeira é explicada pela água ser o mais limitante dentre os fatores ambientais que propiciam estresse a uma planta, visto que determina as relações hídricas e promove mudanças em seu metabolismo (SCHWIDER et al., 2013). O déficit hídrico tende a ocasionar a formação de tecido vascular mais denso, por espessamentos de parede de fibra e vaso, e conseqüentemente o aumento da densidade (BOURIAUD et al., 2005; MARTÍNEZ- CABRERA et al., 2009; CHRISTENSEN-DALSGAARD; ENNOS, 2011).

Para as variáveis qualitativas, cada classe consistiu em uma entrada das RNA, que associada aos pesos, utilizados para o cálculo da importância, resultaram em um maior valor de IR (%). Portanto, a seleção das mesmas pode ter ocorrido em função da quantidade de classes e não da relevância destas para a predição da resposta (STROBL et al., 2007).

5.3 Seleção das variáveis predictoras e estimativa da densidade com a *Random Forest*

A importância das variáveis de entrada na *Random Forest* foi obtida com base nas 10 RF, visto que apresentaram estatísticas semelhantes (Tabela 7). Para Auret e Aldrich (2012), a *Random Forest*, mesmo com pouca capacidade preditiva, gera medidas confiáveis de importância das variáveis.

Tabela 7 – Estatísticas utilizadas para avaliar o desempenho do *Random forest* na estimação da densidade básica da madeira, para os dados de treino e validação

RF	TREINO			VALIDAÇÃO		
	RMSE (%)	V (%)	$r_{y\hat{y}}$	RMSE (%)	V (%)	$r_{y\hat{y}}$
1	3,19	-0,29	0,94	5,08	-0,21	0,75
2	3,19	-0,27	0,94	5,09	-0,45	0,75
3	3,21	-0,27	0,94	5,10	-0,50	0,75
4	3,21	-0,27	0,94	5,07	-0,49	0,75
5	3,20	-0,28	0,94	5,07	-0,48	0,75
6	3,20	-0,31	0,94	5,07	-0,52	0,75
7	3,20	-0,31	0,94	5,08	-0,52	0,75
8	3,20	-0,30	0,94	5,08	-0,50	0,75
9	3,19	-0,30	0,94	5,09	-0,50	0,75
10	3,19	-0,30	0,94	5,08	-0,51	0,75

em que: $r_{y\hat{y}}$ = coeficiente de correlação entre as densidades observadas e estimadas; V (%) = viés e RMSE (%) = raiz do erro quadrado médio.

A partir da média das importâncias das variáveis baseada na permutação (pVI), obtidas para as 10 RF, estas foram ranqueadas de modo decrescente de acordo com a influência na predição da densidade básica da madeira (Tabela 8).

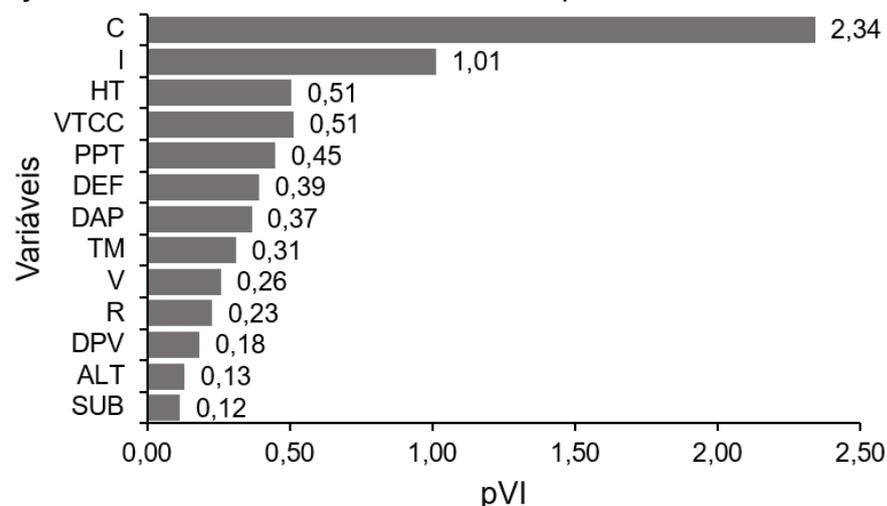
O clone sobressaiu diante das demais variáveis, ao apresentar a maior média de valor de importância (Figura 11). Na sequência, estão a idade, altura total, volume total com casca e precipitação total média. As outras variáveis resultaram em valores inferiores a 0,40.

Tabela 8 – Ranking das variáveis preditoras da densidade básica da madeira pelo *Random Forest*

VARIÁVEL	MÍNIMA	MÉDIA	MÁXIMA	RANKING
C	2,23	2,34	2,43	1
I	0,90	1,01	1,06	2
HT	0,45	0,51	0,74	3
VTCC	0,41	0,51	0,56	4
PPT	0,34	0,45	0,68	5
DEF	0,33	0,39	0,53	6
DAP	0,28	0,37	0,43	7
TM	0,17	0,31	0,55	8
V	0,21	0,26	0,31	9
R	0,17	0,23	0,31	10
DPV	0,15	0,18	0,25	11
ALT	0,08	0,13	0,18	12
SUB	-0,03	0,12	0,24	13

em que: C= clone; I= idade; TM= temperatura média; DEF= déficit hídrico; DPV= déficit pressão de vapor; VTCC= volume total com casca; PPT= precipitação total média; DAP= diâmetro a altura do peito; HT= altura total; ALT= altitude, V= vento; SUB= sub-região e R= relevo.

Figura 11 – Média da importância das variáveis preditoras baseada na permutação (pVI) para predição da densidade básica da madeira, pelo *Random Forest*



em que: C= clone; I= idade; TM= temperatura média; DEF= déficit hídrico; DPV= déficit pressão de vapor; VTCC= volume total com casca; PPT= precipitação total média; DAP= diâmetro a altura do peito; HT= altura total; ALT= altitude, V= vento; SUB= sub-região e R= relevo.

Neste método de seleção, a altura total foi uma das variáveis com maior importância na predição da densidade da madeira, o que pode ser atribuído a sua correlação moderada (0,4) com a densidade da madeira (Figura 9).

Semelhante ao pVI da altura, o volume total com casca, foi uma importante variável para predição da densidade. Em relação a isto, Rezende, Saglietti e Chaves (1998), ao avaliarem a variação da densidade da madeira em diferentes níveis de produtividade, confirmaram que o ganho em produtividade (m^3) ocasionou o decréscimo da densidade em *Eucalyptus grandis* aos 8 anos de idade.

Em seguida, a precipitação total média também influenciou na densidade. A elevada pluviosidade estimula a atividade do meristema cambial e, conseqüentemente, a divisão celular, de modo a aumentar a proporção da madeira juvenil ou do lenho inicial no anel anual de crescimento, que propicia a formação de madeira de menor densidade (MENESES et al., 2015; SETTE JUNIOR et al. 2016).

A variável com menor média de importância foi a sub-região, o que é justificado pela RF, diferente do algoritmo de *Garson*, não ter apresentado preferência por variáveis qualitativas. Uma das vantagens do RF é saber lidar com a mistura de variáveis preditoras, qualitativas e quantitativas (DÍAZ-URIARTE; DE ANDRES, 2006).

5.4 Resultados para estimação da densidade básica da madeira de árvores de *Eucalyptus*

As variáveis foram classificadas em diferentes ordens de importância para cada método de seleção: força bruta, algoritmo de *Garson* e RF. Gevrey, Dimopoulos e Lek (2003) utilizaram diferentes métodos de seleção (PaD, perfil, pesos, perturbação e versões do *stepwise*) que também resultaram na variação da classificação das variáveis, o que pode ser atribuído às diferentes abordagens matemáticas de cada método.

A combinação com uso de cinco variáveis de entrada (método da força bruta) que conferiu melhor desempenho na predição da densidade da madeira e as cinco variáveis de maior importância obtidas pelo algoritmo de *Garson* e *Random Forest* (Tabela 9), demonstraram unanimidade na escolha das variáveis clone e idade. Em relação as demais, o volume total com casca foi obtido pelo *Random Forest* e a força bruta, e o déficit hídrico foi comum entre a força bruta e o algoritmo de *Garson*.

Tabela 9 – Variáveis de entrada para as Redes Neurais Artificiais com uso de cinco variáveis selecionadas a partir dos métodos, força bruta, algoritmo de *Garson*, *Random Forest* e do Inventário florestal contínuo.

RNA	MÉTODO DE SELEÇÃO	VARIÁVEIS
1	Força bruta (combinação)	C, I, VTCC, DEF, TM
2	Algoritmo de <i>Garson</i>	C, SUB, R, I, DEF
3	<i>Random Forest</i>	C, I, HT, VTCC, PPT
4	C + I + Variáveis do IFC	C, I, DAP, HT, VTCC

em que: C= clone; I= idade; TM= temperatura média; DEF= déficit hídrico; DPV= déficit pressão de vapor; VTCC= volume total com casca; PPT= precipitação total média; DAP= diâmetro a altura do peito; HT= altura total; ALT= altitude, V= vento; SUB= sub-região e R= relevo.

5.4.1 Redes Neurais Artificiais

Para as quatro RNA (Tabela 9), os coeficientes de correlação foram maiores para os dados de treinamento quando comparados aos dados de validação (Tabela 10). Zanuncio et al. (2017) ao avaliarem a RNA na predição de propriedades físicas (teor de umidade de equilíbrio e inchaço volumétrico total) e mecânicas (módulo de elasticidade e de ruptura) em função da temperatura utilizada no tratamento da madeira de *E. grandis*, encontraram coeficientes de correlação acima de 95% para os dados de treinamento e acima de 94% para os dados de validação. Para a predição das propriedades colorimétricas observou-se maiores valores para a validação (acima de 98%) quando comparado ao treinamento (acima de 96%).

Fernández et al. (2008), ao estudarem a predição das propriedades mecânicas da madeira (tração perpendicular às fibras de madeira, resistência à flexão e módulo de elasticidade) a partir do uso das propriedades físicas (densidade, teor de umidade, inchaço, espessura e absorção de água) como variáveis de entrada da RNA, obtiveram coeficientes de correlação para os dados de treinamento (acima de 0,98) e validação (acima de 0,86).

Tabela 10 – Estatísticas utilizadas para avaliar o desempenho das Redes Neurais Artificiais (RNAs) na estimação da densidade básica da madeira de árvores de *Eucalyptus*

RNA	TREINO			VALIDAÇÃO		
	RMSE (%)	V (%)	$r_{y\hat{y}}$	RMSE (%)	V (%)	$r_{y\hat{y}}$
1	2,85	-0,01	0,94	4,43	-0,17	0,82
2	3,21	1,16	0,93	5,44	0,79	0,74
3	4,47	0,07	0,84	5,35	0,20	0,73
4	2,97	-0,21	0,93	6,49	-0,52	0,64

em que: $r_{y\hat{y}}$ = coeficiente de correlação entre as densidades observadas e estimadas; V (%) = viés e RMSE (%) = raiz do erro quadrado médio.

A RNA 1 apresentou as melhores estatísticas para os dados de treino e validação, com menor RMSE (%), V (%) e maior $r_{y\hat{y}}$, comparada as RNA 2, 3 e 4. A RNA 3 teve menor RMSE (%) e V (%) com relação a RNA 2 e 4 para os dados de validação. No treino, a RNA 4 compara às RNA 2 e 3 apresentou o menor RMSE (%), já na validação consistiu no maior RMSE (%) e menor $r_{y\hat{y}}$.

Nas Figuras 12 e 13, respectivamente, são mostradas as análises gráficas de correlação (a) entre as densidades observadas e estimadas pelas RNA, distribuição de erros percentuais (b) e histograma dos erros percentuais (c) para o treinamento e validação.

Para os dados de treinamento, o gráfico de correlação (Figura 12a) entre as densidades observadas e estimadas demonstrou a mesma tendência (dados próximos à linha de 45°) para a RNA 1, 2 e 4, diferente da RNA 3, em que os dados ficaram mais distantes da linha. Já, em relação aos dados de validação, a RNA 1 destacou-se com os dados mais próximos à linha.

Para os dados de treinamento, a maior dispersão dos erros (Figura 12b) é apresentada pela RNA 3.

A partir do histograma de erros percentuais (Figura 12c e 13c), verifica-se que as RNA 1, 2 e 4 apresentaram cerca de 90% dos erros nas classes de ± 5 , para os dados de treinamento. Em relação aos dados de validação, destacou-se a RNA 1 com maior representatividade dos erros nas classes de ± 5 , quando comparada as demais RNA. Fernández et al. (2008) encontraram uma maior variação dos erros (%), quando comparados aos deste estudo, para os dados de validação, na predição da tração perpendicular às fibras de madeira entre -17,65 e 40,40 (com média de

14,65), da resistência à flexão entre -42,07 e 33,52 (com média de 13,79) e do módulo de elasticidade entre -29,13 e 5,85 (com média de 12,20).

De modo geral, almeja-se uma RNA que proporcione maior exatidão, para os dados de treinamento e também para os dados de validação. Pode-se considerar a RNA 1 como a mais eficiente em extrair características importantes dos dados de treinamento e validação, visto que, demonstrou desempenho superior às demais, com base nos resultados estatísticos (Tabela 10) juntamente com as análises gráficas (Figuras 12 e 13). Pode-se verificar que a mesma conseguiu reduzir a dispersão dos erros, de modo a proporcionar estimativas de maior exatidão.

Figura 12 – Densidade básica estimada versus densidade básica observada, distribuição dos erros percentuais das estimativas de densidade básica e histograma de erros das estimativas de densidade básica, obtidas por Redes Neurais Artificiais (RNA), para os dados de treinamento.

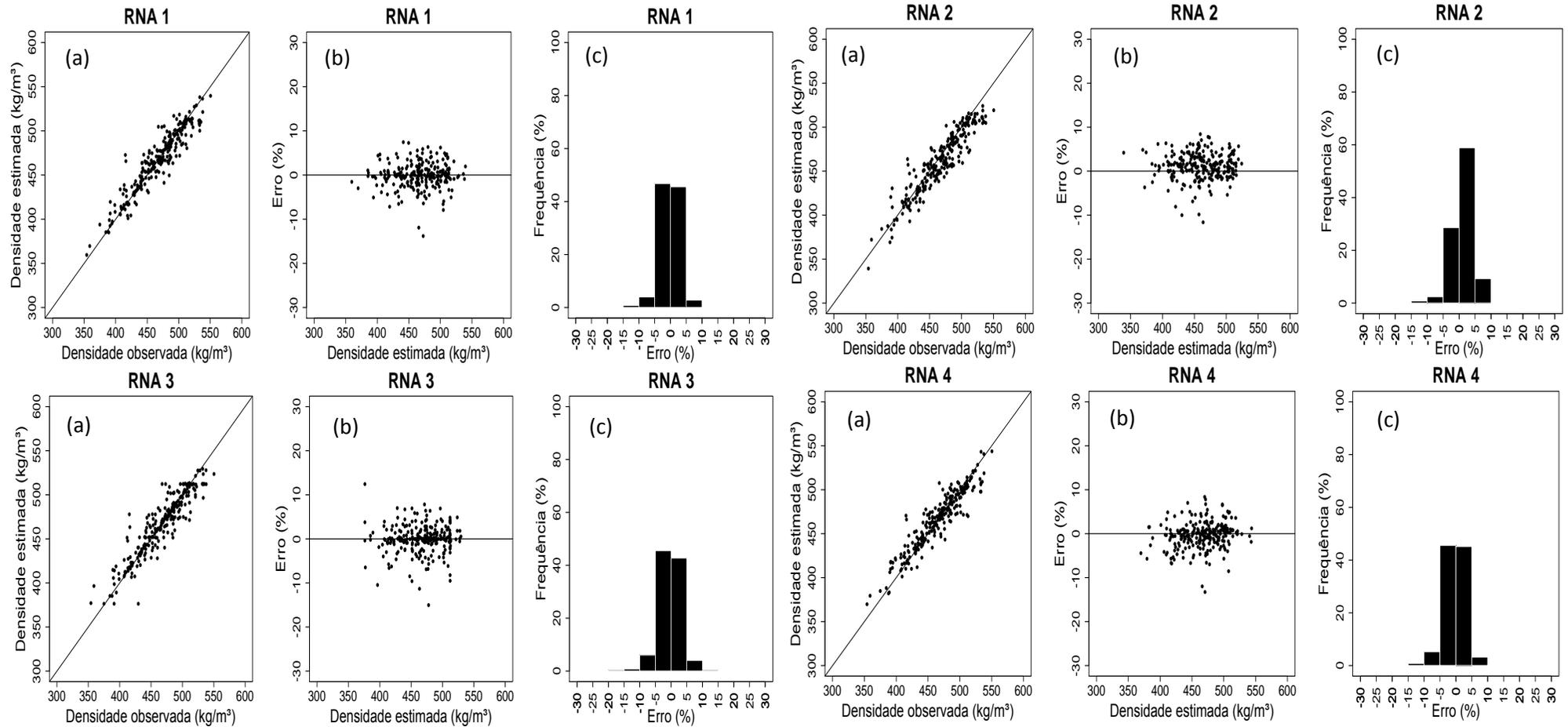
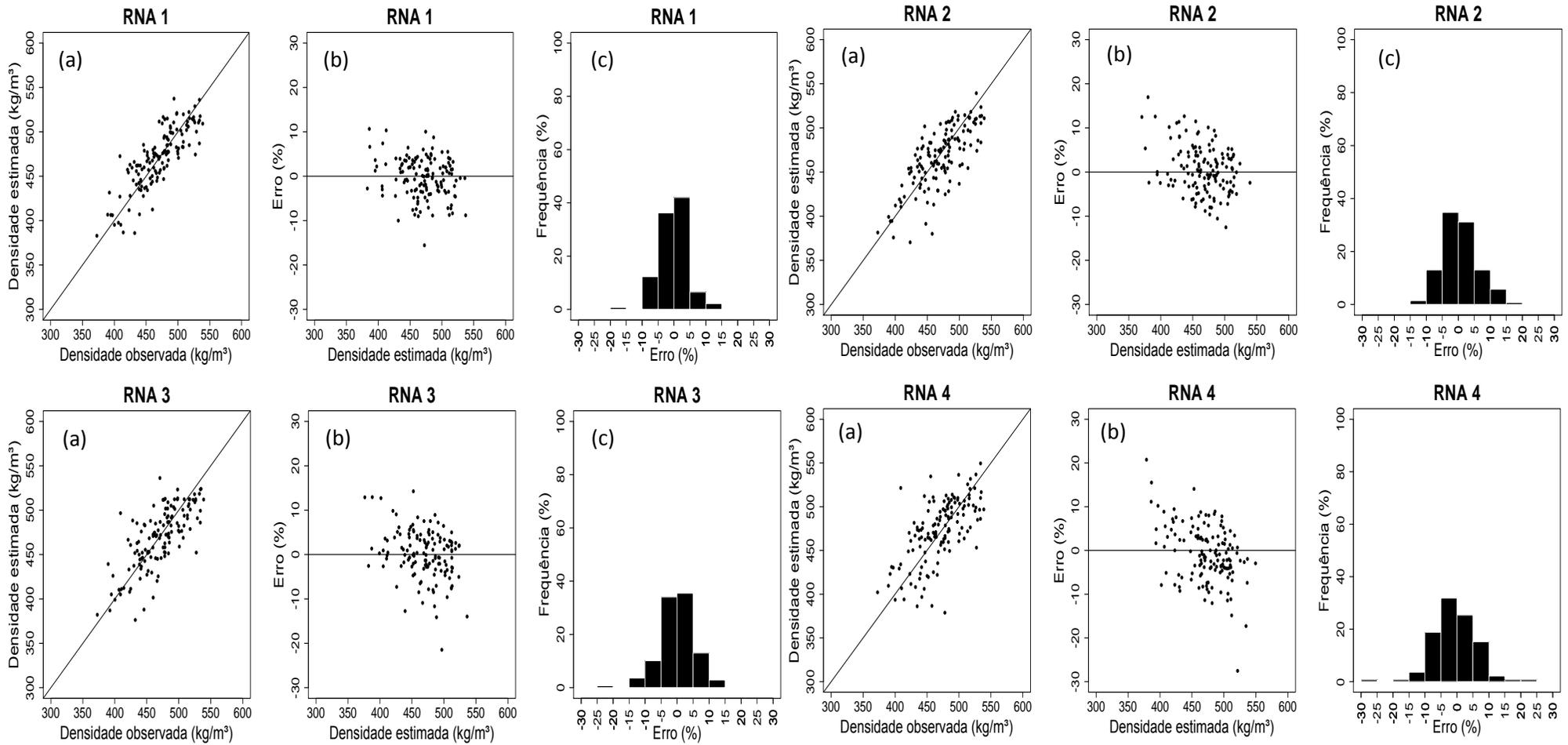


Figura 13 – Densidade básica estimada versus densidade básica observada, distribuição dos erros percentuais das estimativas de densidade básica e histograma de erros das estimativas de densidade básica, obtidas por Redes Neurais Artificiais (RNA), para os dados



de validação.

(c)

O método da força bruta foi o mais apropriado para seleção de variáveis, em termos de avaliação de todas as combinações possíveis, porém este pode ser inviável, em algumas ocasiões, pelo número de combinações geradas, tempo de processamento e avaliação dos resultados. Logo, foi demonstrado neste estudo que o algoritmo de *Garson* e o *Random Forest* foram alternativas viáveis para seleção de variáveis, visto que apresentaram resultados satisfatórios quanto a aproximação dos resultados obtidos pela força bruta e menor tempo de processamento. Neste contexto, a obtenção de estimativas de densidade básica da madeira com boa exatidão é atribuída a interação das variáveis selecionadas em cada método (Tabela 9).

6 CONCLUSÕES

Todas as variáveis, qualitativas e quantitativas utilizadas como entradas das redes foram importantes, sendo predominantes as variáveis clone e idade.

A seleção de variáveis pelo método força bruta com aplicação de Redes Neurais Artificiais, apesar de terem apresentado desempenho satisfatório, demandaram maior tempo para avaliação dos resultados e maior esforço computacional comparado ao algoritmo de *Garson* e o *Random Forest*. Logo, estes dois últimos métodos foram alternativas viáveis para seleção de variáveis.

As Redes Neurais Artificiais podem ser utilizadas para a estimação da densidade básica, em função das variáveis obtidas no inventário florestal, variáveis climáticas e altitude.

REFERÊNCIAS

ALMEIDA, G. M.; PARK, S. W.; CARDOSO, M. Variables selection for neural networks identification for Kraft recovery boilers. **IFAC Proceedings Volumes**, v. 37, n. 16, p. 91-96, 2004.

ALZATE, S. B. A.; TOMAZELLO FILHO, M.; PIEDADE, S. M de S. Variação longitudinal da densidade básica da madeira de clones de *Eucalyptus grandis* Hill ex *Maiden*, *E. saligna* Sm. e *E. grandis* x *urophylla*. **Scientia Forestalis**, v. 68, p. 87-95, 2005.

ANDRADE, M. C. N. de. **Crescimento micelial, produção e características bromatológicas do shiitake em função de linhagens e de propriedades físicas e químicas de espécies e clones de eucalipto**. 2007. 195 f. Tese (Doutorado em Agronomia) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Botucatu, 2007.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS – ABNT. **NBR 11941 - 02: Madeira – Determinação da densidade básica**. Rio de Janeiro: ABNT, 2003. 6 p.

AURET, L.; ALDRICH, C. Interpretation of nonlinear relationships between process variables by use of random forests. **Minerals Engineering**, v. 35, p. 27-42, 2012.

BASTOS, D. G. D.; NASCIMENTO, P. S.; LAURETTO, M. S. Análise Empírica de Desempenho de Quatro Métodos de Seleção de Características para *Random Forests*. **iSys - Revista Brasileira de Sistemas de Informação**, v. 7, n. 2, p. 25-47, 2014.

BATISTA, J. L. F.; COUTO, H. T. Z.; SILVA FILHO, D. F. da. **Quantificação de recursos florestais: árvores, arvoredos e florestas**. São Paulo: Oficina de Textos, 2014. 384 p.

BEALE, M. H.; HAGAN, M. T.; DEMUTH, H. B. **Matlab: neural network toolbox™ user's guide**. Natick: Mathworks, 2016. 416 p.

BERGÈS, L.; NEPVEU, G.; FRANCO, A. Effects of ecological factors on radial growth and wood density components of sessile oak (*Quercus petraea* Liebl.) in Northern France. **Forest ecology and management**, v. 255, n. 3-4, p. 567-579, 2008.

BINOTI, M. L. M. da 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.

BOSCHETTI, W. T. N. et al. Parâmetros dendrométricos e excentricidade da medula em árvores inclinadas de eucalipto. **Scientia Forestalis**. v. 43, n. 108, p. 781-789, 2015.

BOTREL, M. C. G. et al. Seleção de clones de *Eucalyptus* para biomassa florestal e qualidade da madeira. **Scientia Forestalis**, v. 38, n. 86, p. 237-245, 2010.

- BOULESTEIX, A. L. et al. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, v. 2, n. 6, p. 493-507, 2012.
- BOURIAUD, O. et al. Intra-annual variations in climate influence growth and wood density of *Norway spruce*. **Tree physiology**, v. 25, n. 6, p. 651-660, 2005.
- BRAGA, A. de P.; CARVALHO, A. P. de L. F.; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: LTC, 2007. 238 p.
- BRASIL, M. A. M.; VEIGA, R. A. de. A.; MELLO, H. do A. Densidade básica de madeira de *Eucalyptus grandis* HILL EX MAIDEN, aos 3 anos de idade. **IPEF**, n. 19, p. 63-76, 1979.
- BRAZ, R. L. et al. Parâmetros dendrométricos e resistência mecânica das árvores de clones de *Eucalyptus* em áreas sujeitas à ação dos ventos. **Ciência Florestal**, v. 24, n. 4, p. 947-958, 2014.
- BRAZ, R. L. et al. Deformação Residual Longitudinal de Clones de *Eucalyptus* em local sujeito a ventos. **Floresta e Ambiente**, v. 24, p. e00028013, 2017.
- BREIMAN, L. Bagging predictors. **Machine learning**, v. 24, n. 2, p. 123-140, 1996.
- BREIMAN, L. Random Forests. **Machine learning**, v. 45, n. 1, p. 5-32, 2001.
- BRUDER, E. M. **Métodos de determinação da densidade básica e aparente da madeira de *Eucalyptus sp.*** 2012. 99 f. Dissertação (Mestrado em Agronomia) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Botucatu, 2012.
- BRUDER, E. M.; REZENDE, M. A. de; COSTA, V. E. Densidade de *Eucalyptus sp.* próxima a umidade de equilíbrio estimado pelo método de imersão. **Revista Energia na Agricultura**, Botucatu, v. 31, n. 1, p. 38-47, 2016.
- CAIXETA, R. P. et al. Propriedades e classificação da madeira aplicadas à seleção de genótipos de *Eucalyptus*. **Revista Árvore**, v. 27, n. 1, p. 43-51, 2003.
- CAMPOS, J. C. C.; LEITE, H. G. **Mensuração Florestal: perguntas e respostas**. 3. ed. Viçosa, MG: Editora UFV, 2009. 548 p.
- CARDON, A.; MÜLLER, D. N. **Introdução Às Redes Neurais Artificiais**. 1994. 30 f. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 1994.
- CARVALHO JUNIOR, W. et al. Regressão linear múltipla e modelo *Random Forest* para estimar a densidade do solo em áreas montanhosas. **Pesquisa Agropecuária Brasileira**, v. 51, n. 9, p.1428-1437, 2016.
- CATENI, S.; COLLA, V. Variable Selection for Efficient Design of Machine Learning-Based Models: Efficient Approaches for Industrial Applications. In: International

Conference on Engineering Applications of Neural Networks. **Springer**. 2016. p. 352-366.

CHRISTENSEN-DALSGAARD, K. K.; ENNOS, A. R. Effects of drought acclimation on the mechanical properties of *Ochroma pyramidale*, *Betula pendula* and *Acacia karroo* tree seedling stems. **Forestry**, v. 85, n. 2, p. 215-223, 2011.

COUTO, A. M. et al. Modelagem da densidade básica da madeira de *Eucalyptus grandis* e *Eucalyptus urophylla* por métodos não destrutivos. **Cerne**, v. 19, n. 1, p. 27-34, 2013.

DEMUNER, W. P. **Predição do impacto da madeira em fábrica kraft de eucalipto**. 2011. 24 f. Monografia (Graduação em Engenharia Industrial Madeireira) – Universidade Federal do Espírito Santo, Jerônimo Monteiro, 2011.

DÍAZ-URIARTE, R.; DE ANDRES, S.A. Gene selection and classification of microarray data using random forest. **BMC bioinformatics**, v. 7, n. 1, p. 3, 2006.

DIAS, D. C et al. Uso da técnica de resistografia e de variáveis dendrométricas na modelagem da densidade básica de povoamentos clonais de *Eucalyptus*. **Ciência Florestal**, v. 27, n. 2, p. 609-619, 2017.

DORESWAMY, H.; VASTRAD, C. M. Performance analysis of neural network models for oxazolines and oxazoles derivatives descriptor dataset. **International Journal of Information Sciences and Techniques**, v. 3, n. 6, p. 111-123, 2013.

DOWNES, G.; DREW, D. Climate and growth influences on wood formation and utilisation. **Southern Forests**, v. 70, n. 2, p. 155-167, 2009.

FERNANDES, D. E. et al. Influência da produtividade de clones híbridos de eucalipto na densidade da madeira e na polpação Kraft. **Scientia Forestalis**, v. 39, n. 90, p. 143-150, 2011.

FERNÁNDEZ, F. G. et al. Prediction of standard particleboard mechanical properties utilizing an artificial neural network and subsequent comparison with a multivariate regression model. **Forest Systems**, v. 17, n. 2, p. 178-187, 2008.

FERREIRA, M.; KAGEYAMA, P. Y. Melhoramento genético da densidade da madeira de eucalipto. **IPEF**, v. 6, n. 20, p. a.1-a.14, 1978.

FISCHER, Andreas. How to determine the unique contributions of input-variables to the nonlinear regression function of a multilayer perceptron. **Ecological modelling**, v. 309, p. 60-63, 2015.

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

FOELKEL, C. E. B. Madeira do Eucalipto: da floresta ao digestor. **IPEF**, v. 6, n. 20, p. e.1-e.25, 1978.

GARSON, G. D. Interpreting neural network connection weights. **AI expert**, v. 6, n. 4, p. 46-51, 1991.

GENUER, R.; POGGI, J.M.; TULEAU-MALOT, C. Variable selection using random forests. **Pattern Recognition Letters**, v. 31, n. 14, p. 2225–2236, 2010.

GEVREY, M.; DIMOPOULOS, I.; LEK, S. Review and comparison of methods to study the contribution of variables in artificial neural network models. **Ecological modelling**, v. 160, n. 3, p. 249-264, 2003.

GIVNISH, T. J. et al. Determinants of maximum tree height in Eucalyptus species along a rainfall gradient in Victoria, Australia. **Ecology**, v. 95, n. 11, p. 2991-3007, 2014.

GOH, A. T. C. Back-propagation neural networks for modeling complex systems. **Artificial Intelligence in Engineering**, v. 9, n. 3, p. 143-151, 1995.

GOLDSTEIN, B. A.; POLLEY, E. C.; BRIGGS, F.B.S. Random forests for genetic association studies. **Statistical applications in genetics and molecular biology**, v. 10, n. 1, 2011.

GOLDSCHMIDT, R.; PASSOS, E. **Data mining: um guia prático, conceitos, técnicas, ferramentas, orientações e aplicações**. Rio de Janeiro: Campus. 2005. 259 p.

GONÇALVES, Débora Romano Camilo. **Avaliação do efeito da face de exposição solar sobre o crescimento de plantios comerciais de *Eucalyptus sp* na região sudoeste do estado de São Paulo**. 2015. 90 f. Dissertação (Mestrado em Recursos Florestais) – Universidade de São Paulo, Piracicaba, SP, 2015.

GOUVÊA, A. F. G. et al. Determinação da densidade básica da madeira de *Eucalyptus* por diferentes métodos não destrutivos. **Revista Árvore**, v. 35, n. 2, p. 349-358, 2011.

GREGORUTTI, B.; MICHEL, B.; SAINT-PIERRE, P. Correlation and variable importance in random forests. **Statistics and Computing**, v. 27, n. 3, p. 659-678, 2017.

HAYKIN, S. **Redes neurais: princípios e práticas**. 2. ed. Porto Alegre: Bookman, 2001. 900 p.

HAYKIN, S. **Neural networks and learning machines**. 3. ed. USA: Pearson, 2009. 906 p.

HAPFELMEIER, A.; ULM, K. A new variable selection approach using random forests. **Computational Statistics & Data Analysis**, v. 60, p. 50-69, 2013.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2. ed. California: Springer, 2008. 745 p.

HSING, T. Y.; PAULA, N. F de; PAULA, R. C. de. Características dendrométricas, químicas e densidade básica da madeira de híbridos de *Eucalyptus grandis* X *Eucalyptus urophylla*. **Ciência Florestal**, v. 26, n. 1, 2016.

HSU, D. Identifying key variables and interactions in statistical models of building energy consumption using regularization. **Energy**, v. 83, p. 144-155, 2015.

INDÚSTRIA BRASILEIRA DE ÁRVORES – IBÁ. **Relatório anual 2017**. Brasília, DF: IBÁ, 2017. 80 p.

JAIN, A. K.; MAO, J.; MOHIUDDIN, K. Artificial Neural Networks: A Tutorial. **Computer**, v. 29, n. 3, p.31-44, 1996.

JAMES, G. et al. **An introduction to statistical learning-with applications in R**. New York: Springer. 2013. 440 p.

JANITZA, S.; TUTZ, G.; BOULESTEIX, A. L. Random forest for ordinal responses: prediction and variable selection. **Computational Statistics & Data Analysis**, v. 96, p.57-73, 2016.

KUBE, P.; RAYMOND, C.; BANHAM, P. W. Genetic parameters for diameter, basic density, cellulose content and fibre properties for *Eucalyptus nitens*. **International Journal of Forest Genetics**, v. 8, n. 4, p. 285-294, 2001.

LATORRACA, J.V de F.; ALBUQUERQUE, C. E. C. Efeito do rápido crescimento sobre as propriedades da madeira. **Floresta e Ambiente**, v. 7, n. 1, p. 279-291, 2000.

LEITE, H. G. et al. Redes Neurais Artificiais para a estimação da densidade básica da madeira. **Scientia Forestalis**, v. 44, n. 109, p. 149-154, 2016.

LI, J.Y.; CHOW, T.W.S.; YU, Y.L. The estimation theory and optimization algorithm for the number of hidden units in the higher-order feedforward neural network. In: **Neural Networks**, 1995. Proceedings., IEEE International Conference on. IEEE, 1995. p. 1229-1233.

LOPES, G. de A.; GARCIA, J. N. Densidade básica e umidade natural da madeira de *Eucalyptus saligna* Smith, de Itatinga, associadas aos padrões de casca apresentados pela população. **Scientia Forestalis**, n. 62, p. 13-23, 2002.

LOUPPE, G. et al. Understanding variable importances in forests of randomized trees. **Advances in neural information processing systems**, p 431-439, 2013.

MARTÍNEZ-CABRERA, H. I. et al. Wood anatomy and wood density in shrubs: responses to varying aridity along transcontinental transects. **American Journal of Botany**, v. 96, n. 8, p. 1388-1398, 2009.

MAURI, R. et al. Wood density of clones of *Eucalyptus urophylla* x *Eucalyptus grandis* in different conditions of growth. **Floresta**, v. 45, n. 1, p. 193 - 202, 2015.

MEHROTRA, K.; MOHAN, C. K.; RANKA, S. **Elements of Artificial Neural Networks**. Cambridge: Mit, 1997. 344 p.

MENESES, V. A. et al. Efeito da idade e do sítio na densidade básica e produção de massa seca de madeira em um clone do *Eucalyptus urophylla*. **Scientia Forestalis**, v. 43, n. 105, p. 101-116, 2015.

MIFSUD, B. M. Victoria's tallest trees. **Australian Forestry**, v. 66, n. 3, p. 197-205, 2003.

MILAGRES, F. R. **Avaliação da madeira de híbridos de *Eucalyptus globulus* com *E. grandis* e *E. urophylla*, para produção de celulose, utilizando espectroscopia NIR**. 2009. 125 f. Dissertação (Mestrado em Ciência Florestal) – Universidade Federal de Viçosa, Viçosa, MG, 2009.

MONTEIRO, T. C. et al. Influência do lenho de tração nas propriedades físicas da madeira de *Eucalyptus sp.* **Journal of Biotechnology and Biodiversity**, v. 1, n. 1, p. 6-11, 2010.

MORFIDIS, K.; KOSTINAKIS, K. Seismic parameters' combinations for the optimum prediction of the damage state of R/C buildings using neural networks. **Advances in Engineering Software**, v. 106, p. 1-16, 2017.

MORAIS JÚNIOR, A. A.; LIRA, J. O. B.; BRITO, R. P. Análise transiente e estimativa de variáveis críticas em uma coluna industrial de alta pureza utilizando redes neurais artificiais. **Revista interdisciplinar de pesquisa em engenharia**, v.2, n.9, p. 77-97, 2016.

MUNERI, A.; RAYMOND, C. A. Genetic parameters and genotype-by-environment interactions for basic density, pilodyn penetration and stem diameter in *Eucalyptus globulus*. **Forest Genetics**, v. 7, n. 4, p. 317-328, 2000.

NGUYEN, T.T.; HUANG, J.Z.; NGUYEN, T.T. Unbiased feature selection in learning random forests for high-dimensional data. **The Scientific World Journal**, v. 2015, p.1-18, 2015.

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.

OLDEN, J. D.; JACKSON, D. A. Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. **Ecological modelling**, v. 154, n. 1-2, p. 135-150, 2002.

OLDEN, J. D.; JOY, M. K.; DEATH, R. G. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. **Ecological Modelling**, v. 178, n. 3, p. 389-397, 2004.

OLIVEIRA, G. G.; PEDROLLO, O. C.; CASTRO, N. M. Simplifying artificial neural network models of river basin behaviour by an automated procedure for input

variable selection. **Engineering Applications of Artificial Intelligence**, v. 40, p. 47-61, 2015.

OLIVEIRA, G. M. V. et al. Efeito do ambiente sobre a densidade da madeira em diferentes fitofisionomias do estado de Minas Gerais. **Cerne**, v. 18, n. 2, p. 345-352, 2012.

OÑA, J. de; GARRIDO, C. Extracting the contribution of independent variables in neural network models: a new approach to handle instability. **Neural Computing and Applications**, v. 25, p. 859-869, 2014.

PENTOSÓ, K. The methods of extracting the contribution of variables in artificial neural network models—Comparison of inherent instability. **Computers and Electronics in Agriculture**, v. 127, p. 141-146, 2016.

PFAUTSCH, S. et al. Climate determines vascular traits in the ecologically diverse genus *Eucalyptus*. **Ecology Letters**, v. 19, n. 3, p. 240-248, 2016.

POORTER, L. et al. The importance of wood traits and hydraulic conductance for the performance and life history strategies of 42 rainforest tree species. **New phytologist**, v. 185, n. 2, p. 481-492, 2010.

PROTÁSIO, T. de P. et al. Qualidade da madeira e do carvão vegetal oriundos de floresta plantada em Minas Gerais. **Pesquisa Florestal Brasileira**, v. 34, n. 78, p. 111-123, 2014.

QI, Y. Random forest for bioinformatics. In: **Ensemble Machine Learning**. USA: Springer, 2012. p. 307-323.

REIS, L. P. et al. Estimation of mortality and survival of individual trees after harvesting wood using artificial neural networks in the amazon rain forest. **Ecological Engineering**, v. 112, p. 140-147, 2018.

RIBEIRO, F. de A.; ZANI FILHO, J. Variação da densidade básica da madeira em espécies/procedências de *Eucalyptus spp.* **IPEF**, n. 46, p. 76-85, 1993.

REZENDE, M. A. de; SAGLIETTI, J. R. C.; CHAVES, R. Variação da massa específica da madeira de *Eucalyptus grandis* aos 8 anos de idade em função de diferentes níveis de produtividade. **Scientia Forestalis**, n. 53, p. 71-78, 1998.

ROSADO, A. M. et al. Seleção simultânea de clones de eucalipto de acordo com produtividade, estabilidade e adaptabilidade. **Pesquisa Agropecuária Brasileira**, v. 47, n. 7, p. 966-973, 2012.

SCHWIDER, Y. S. et al. Efeito do déficit hídrico sobre o crescimento de eucalipto em diferentes condições microclimáticas. **Enciclopédia Biosfera**, v. 9, n. 16, p. 888-900, 2013.

SCOLFORO, J. R. S. **Biometria Florestal**: parte I: modelos de regressão linear e não-linear, parte II: modelos para relação hipsométrica, volume, afilamento e peso de matéria seca. Lavras: UFLA/FAEPE, 2005. 352 p.

SCOLFORO, J. R. S.; THIERSCH, C. R. **Biometria Florestal**: medição, volumetria e gravimetria. Lavras: UFLA/FAEPE, 2004. 285 p.

SETTE JUNIOR, C. R. et al. Crescimento em diâmetro do tronco das árvores de *Eucalyptus grandis* W. Hill. ex. Maiden e relação com as variáveis climáticas e fertilização mineral. **Revista árvore**, v. 34, n. 6, p.979-990, 2010.

SETTE JUNIOR, C. R. et al. Efeito da idade e posição de amostragem na densidade e características anatômicas da madeira de *Eucalyptus grandis*. **Revista Árvore**, v. 36, n. 6, 2012.

SHIMOYAMA, V. R. de S.; BARRICHELO, L. E. G. Densidade básica da madeira, melhoramento e manejo florestal. **IPEF**, v. 6, n. 20, p. 1-22, 1989.

SILVA, I. N. da; SPATTI, D. H.; FLAUZINO, R. A. **Redes neurais artificiais**: para engenharia e ciências aplicadas. São Paulo: Artliber, 2010. 399 p.

SILVA, S.R.; SCHIMIDT, F. Redução de variáveis de entrada de redes neurais artificiais a partir de dados de análise de componentes principais na modelagem de oxigênio dissolvido. **Quim. Nova**, v. 39, n. 3, p. 273-278, 2016.

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, n. 6, p. 1133-1139, 2009.

STEIN, F. da R. **Modelagem de produção industrial de celulose kraft com modelos aditivos generalizados e redes neurais**. 2010. 101 f. Dissertação (Mestrado Profissional em Tecnologia de Celulose e Papel) – Universidade Federal de Viçosa, Viçosa, MG, 2010.

STROBL, C. et al. Bias in random forest variable importance measures: Illustrations, sources and a solution. **BMC bioinformatics**, v. 8, n. 1, p.25, 2007.

TAN, B. et al. Genomic relationships reveal significant dominance effects for growth in hybrid *Eucalyptus*. **Plant Science**, v. 267, p. 84-93, 2018.

THOMAS, D. S.; MONTAGU, K. D.; CONROY, J. P. Changes in wood density of *Eucalyptus camaldulensis* due to temperature—the physiological link between water viscosity and wood anatomy. **Forest Ecology and Management**, v. 193, n. 1-2, p. 157-165, 2004.

THOMAS, D. S.; MONTAGU, K. D.; CONROY, J. P. Temperature effects on wood anatomy, wood density, photosynthesis and biomass partitioning of *Eucalyptus grandis* seedlings. **Tree Physiology**, v. 27, n. 2, p. 251-260, 2007.

TRUGILHO, P. F. Densidade básica e estimativa de massa seca e de lignina na madeira em espécies de *Eucalyptus*. **Ciência e Agrotecnologia**, v. 33, n. 5, p. 1228-1239, 2009.

TRUGILHO, P. F. et al. Estimate of fixed carbon in the wood of a hybrid clone of *Eucalyptus urophylla* and *Eucalyptus grandis*. **Cerne**, v. 16, p. 33–40, 2010.

VAN LAAR, A.; AKÇA, A. **Forest Mensuration**. 2. ed. Dordrecht: Springer, 2007. 383 p.

VIEIRA, F. D.; OLIVEIRA, S.R. de M.; PAIVA, S. R. Metodologia baseada em técnicas de mineração de dados para suporte à certificação de raças de ovinos. **Embrapa Informática Agropecuária-Artigo em periódico indexado (ALICE)**, v. 35, n. 6, p.1172-1186, 2015.

VITAL, B. R. Métodos de determinação da densidade da madeira. Viçosa, MG: **SIF**, 1984. 21p. (Boletim técnico, 1).

WILLMOTT, C. J.; MATSUURA, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. **Climate research**, v. 30, p. 79-82, 2005.

ZANUNCIO, A. J. V. et al. Prediction of the physical, mechanical and colorimetric properties of *Eucalyptus grandis* heat-treated wood using artificial neural networks. **Scientia Forestalis**, v. 45, n. 113, p. 109-118, 2017.

ZHOU, B. et al. Relative importance analysis of a refined multi-parameter phosphorus index employed in a strongly agriculturally influenced watershed. **Water, Air, & Soil Pollution**, v. 226, n. 3, p. 25, 2015.