



**LAÍS ALMEIDA ARAÚJO**

**ALGORITMOS HÍBRIDOS APLICADOS À BIOMETRIA  
FLORESTAL**

**LAVRAS – MG**

**2020**

**LAÍS ALMEIDA ARAÚJO**

**ALGORITMOS HÍBRIDOS APLICADOS À BIOMETRIA FLORESTAL**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, curso de Mestrado, área de concentração em Manejo Florestal, para a obtenção do título de Mestre.

Prof. Dr. Lucas Rezende Gomide  
Orientador

**LAVRAS – MG**  
**2020**

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

Araújo, Laís Almeida.

Algoritmos híbridos aplicados a biometria florestal / Laís  
Almeida Araújo. - 2020.

106 p.

Orientador(a): Lucas Rezende Gomide.

Dissertação (mestrado acadêmico) - Universidade Federal de  
Lavras, 2020.

Bibliografia.

1. Relações Alométricas. 2. Manejo Florestal. 3. Inteligência  
computacional. I. Gomide, Lucas Rezende. II. Título.

**LAÍS ALMEIDA ARAÚJO**

**ALGORITMOS HÍBRIDOS APLICADOS À BIOMETRIA FLORESTAL**

**HYBRID ALGORITHMS APPLIED TO FOREST BIOMETRY**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, curso de Mestrado, área de concentração em Manejo Florestal, para a obtenção do título de Mestre.

APROVADA EM 28 de fevereiro de 2020

Prof. Dr. Lucas Rezende Gomide UFLA

Prof. Dr. Sérgio Henrique Godinho Silva UFLA

Prof. Dr. Alexandre Behling UFPR

Prof. Dr. Lucas Rezende Gomide

Orientador

**LAVRAS – MG**

**2020**

*À minha mãe,  
Cleide Almeida Santos de Araújo*

*À minha irmã,  
Lizandra Almeida Araújo*

*Ao meu pai,  
Jaír Batista de Araújo*

*Dedico.*

## AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus e a Nossa Senhora, pelas graças alcançadas e pela oportunidade de chegar até aqui e concluir um mestrado.

Ao meu pai, Jair, e a minha mãe, Cleide (*In memoriam*), pela educação, amor e por não medir esforços para proporcionar a mim e a minha irmã, todas as oportunidades de estudo e sempre estar apoiando nas nossas decisões.

A minha irmã, por estar sempre ao meu lado e por todo carinho.

Ao meu orientador, Lucas Resende Gomide, pela orientação, paciência, confiança, ensinamentos e pelo apoio para que eu concluísse com êxito este trabalho.

A Universidade Federal de Lavras (UFLA), ao Programa de Pós-Graduação em Ciências Florestais e ao Laboratório de Estudos e Projetos em Manejo Florestal (LEMAF), pela oportunidade conferida para realização do mestrado.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão de bolsas de estudos.

Ao professor Mário Dobner Jr. por ceder os dados necessários para realização de parte do estudo desenvolvido neste trabalho.

A CEMIG pelo apoio financeiro na coleta dos dados florestais na bacia do Rio Grande, base para realização de parte do estudo desenvolvido neste trabalho.

A todos do Laboratório de Planejamento e Otimização Florestal por todo apoio, paciência, pela boa convivência, amizade, união e todos os ensinamentos compartilhados.

Aos professores do Departamento de Ciências Florestais, por todo o aprendizado e trabalhos desenvolvidos.

Aos amigos da pós-graduação, pela companhia, amizade e apoio durante essa trajetória.

As amigas de república, pela companhia, paciência e amizade.

A toda família e amigos, pelo incentivo e torcida.

Aos professores da banca examinadora, por aceitarem o convite para contribuir pela avaliação e sugestões de melhoria.

**MUITO OBRIGADA!**

## RESUMO

A compreensão e quantificação do crescimento das árvores são de suma importância para o fornecimento de informações confiáveis, a fim de possibilitar uma gestão adequada dos recursos florestais. Assim, a modelagem vem se constituindo em uma importante ferramenta, pois tem a capacidade de sintetizar conhecimentos, identificar lacunas, projetar e prever o comportamento de acordo com as diversas condições. Nesse contexto, muitas técnicas são utilizadas, destacando a inteligência computacional, com grande crescimento no setor florestal e alcançado bons resultados. Como exemplo têm-se o *Random Forest*, *Simulated Annealing* e Algoritmo Genético. Nessa dissertação esses algoritmos foram aplicados em diferentes situações para florestas equiâneas e inequiâneas. O primeiro artigo teve como objetivo principal avaliar a aplicabilidade das meta-heurísticas, algoritmo genético e *Simulated Annealing* na predição de parâmetros da função Weibull de 2 parâmetros. A partir dos resultados obtidos foi observado a adequação e o potencial das meta-heurísticas e concluiu-se que os métodos testados são consistentes e estáveis, superando os métodos clássicos independente dos regimes silviculturais envolvendo desbastes avaliados. O segundo artigo aborda a seleção de variáveis e modelagem do estoque de carbono acima do solo, em nível árvore, da vegetação nativa, por meio da construção de um método híbrido envolvendo o *Simulated Annealing* e *Random forest*. Os resultados obtidos indicaram que o método híbrido foi eficiente, visto que conseguiu diminuir o número de variáveis (menos de 16 variáveis) e identificar aquelas que mais contribuem para explicar o estoque de carbono, além de reduzir o erro das estimativas. De forma geral, foi observada a capacidade de aplicação da técnica de aprendizado de máquina nos casos avaliados.

**Palavras-chave:** Relações Alométricas. Manejo Florestal. Inteligência computacional.

## ABSTRACT

Understanding and quantifying the growth of trees are very important for providing available information, in order to enable an adequate management of forest resources. Thus, modeling has becoming an important tool, because it has the ability to synthesize knowledge, identify gaps, design and predict behavior according to different conditions. In this context, many techniques are used, highlighting computational intelligence with great growth in the forest sector and good results. As an example, we have the *Random Forest*, Simulated Annealing and Genetic Algorithm. In this dissertation, these algorithms were applied in different situations for native and planted forests. The first article had as main objective to evaluate the efficiency of the genetic algorithm and Simulated Annealing in the prediction of parameters of the Weibull function of 2 parameters. From the results it was concluded that the tested methods are consistent and stable, surpassing the classic methods regardless of silvicultural regimes involving thinning evaluated. The second article approach the selection of variables and modeling of the above-ground carbon stock, at tree level, of native vegetation, through the construction of a hybrid method involving the Simulated Annealing and Random forest. The results obtained indicated that the hybrid method was efficient, since it managed to reduce the number of variables (less than 16 variable) and identify those that contribute most to explain the carbon stock, in addition to reducing the error of the estimates. In general, the ability to apply the machine learning technique was observed in the cases evaluated.

**Keywords:** Allometric Relations. Forest Management. Computational intelligence.

## LISTA DE FIGURAS

### PRIMEIRA PARTE

- Figura 1 - Esquema simples de um neurônio artificial tipo *Perceptron*.....21
- Figura 2 - Fluxograma do Algoritmo Genético.....23
- Figura 3 - Fluxograma do algoritmo *Simulated Annealing*. ....26

### SEGUNDA PARTE – ARTIGOS

#### ARTIGO 1

- Figura 1 - Fluxograma simplificado das etapas metodológicas. ....55
- Figura 2- Gráfico de *violin plot* da variável diâmetro à altura do peito (cm) entre os diferentes tipos de desbaste. ....61
- Figura 3 - Distribuição dos parâmetros obtidos por todos os métodos avaliados em diferentes idades. ....63
- Figura 4 – Representação do diagrama de Taylor (1) envolvendo todos os ajustes e o comportamento do espaço amostral do parâmetros b e c para uma parcela submetida a 2 desbastes gerados pelo gráfico de superfície de resposta (2). ....64
- Figura 5 - Os melhores percentis 1 e 2 selecionados pelo método híbrido SA<sub>2</sub> - Percentis para as idades de 5, 12, 18, 22 e 30 anos.....67
- Figura 6 - Histogramas de frequências observadas versus a distribuição Weibull ajustada pelo melhor método por grupo, em uma parcela submetida a desbaste moderado em diferentes idades. ....68
- Figura 7 - Erros das classes acumulados (%) de cada repetição para os métodos com bom desempenho, para os diferentes tipos de tratamentos nas idades de 5, 10, 15, 22 e 30 anos.....69

#### ARTIGO 2

- Figura 1 – Mapa de localização da Bacia Hidrográfica do Rio Grande no estado de Minas Gerais, sendo (A) Classificação da vegetação, (B) Classificação dos tipos de solos, (C) Classificação altimétricas, e (D) índices de pluviosidade. ....81
- Figura 2 - Fluxograma do modelo híbrido *Simulated Annealing* e *Random Forest*.....87
- Figura 3 – Distribuição residual dos melhores modelos gerados em cada grupo para estimar o estoque de carbono (kg/árvore). ....95

## LISTA DE TABELAS

### SEGUNDA PARTE – ARTIGOS

#### ARTIGO 1

- Tabela 1 - Tratamentos de desbaste com o número de intervenções e as medições realizadas em todas as parcelas para cada tratamento. .... 54
- Tabela 2 - Estatísticas descritivas dos valores da função de avaliação obtidas entre as repetições e o tempo de processamento médio para cada técnica de ajuste ..... 62
- Tabela 3 - Síntese dos resultados obtidos pela avaliação estatística dos métodos de ajuste da distribuição Weibull nos diferentes tipos de regime de desbaste. .... 65

#### ARTIGO 2

- Tabela 1 - Distribuição diamétrica dos indivíduos selecionados na amostragem para o procedimento de cubagem rigorosa e coleta das demais variáveis morfométricas. 82
- Tabela 2 - Análise exploratória do teor de Carbono (kg) nos diferentes compartimentos da árvore. .... 89
- Tabela 3 - Equações ajustadas via método *Stepwise* e do modelo de Spurr linearizado para a estimativa do estoque de Carbono total, no fuste, nos galhos e nas folhas. .... **Erro! Indicador não definido.**
- Tabela 4 - Estatísticas de ajuste dos modelos ajustados para gerar equações para estimar o estoque de Carbono (kg) para os compartimentos total, fuste, galho e folha. .... 90
- Tabela 5 - Coeficientes de correlação linear de Pearson entre o estoque de Carbono dos compartimentos total, fuste, galho e folha e as variáveis selecionadas entre os melhores métodos de seleção utilizados. .... 96

## SUMÁRIO

<b>PRIMEIRA PARTE</b> .....	<b>12</b>
<b>1</b> <b>INTRODUÇÃO GERAL</b> .....	<b>13</b>
<b>2</b> <b>REVISÃO DE LITERATURA</b> .....	<b>14</b>
2.1 <b>Estoque de carbono nas florestas</b> .....	<b>14</b>
2.2 <b>Métodos de predição</b> .....	<b>16</b>
2.2.1 <b>Análise de regressão</b> .....	<b>16</b>
2.2.2 <b>Inteligência Artificial</b> .....	<b>17</b>
2.2.2.1 <b>Random Forest</b> .....	<b>18</b>
2.2.2.2 <b>Redes Neurais Artificiais</b> .....	<b>20</b>
2.2.2.3 <b>Support Vector Machine (SVM)</b> .....	<b>22</b>
2.2.2.4 <b>Algoritmo Genético</b> .....	<b>23</b>
2.2.2.5 <b>Simulated Annealing</b> .....	<b>25</b>
2.2.2.6 <b>Métodos híbridos</b> .....	<b>27</b>
2.3 <b>Distribuição Diamétrica</b> .....	<b>28</b>
2.4 <b>Distribuição Weibull</b> .....	<b>30</b>
2.4.1 <b>Método dos Momentos</b> .....	<b>32</b>
2.4.2 <b>Método dos Percentis</b> .....	<b>33</b>
2.4.3 <b>Método Máxima Verossimilhança</b> .....	<b>34</b>
2.5 <b>Métricas de avaliação da aderência das distribuições</b> .....	<b>35</b>
<b>3</b> <b>CONSIDERAÇÕES GERAIS</b> .....	<b>37</b>
<b>REFERÊNCIAS</b> .....	<b>37</b>
<b>SEGUNDA PARTE – ARTIGOS</b> .....	<b>49</b>
<b>ARTIGO 1 – ESTIMAÇÃO DE PARÂMETROS DA FUNÇÃO WEIBULL UTILIZANDO META-HEURÍSTICAS</b> .....	<b>50</b>
<b>ARTIGO 2 – ESTOQUE DE CARBONO ACIMA DO SOLO: UM ESTUDO SOBRE A OTIMIZAÇÃO NA SELEÇÃO DE VARIÁVEIS EXPLICATIVAS E MODELAGEM</b> .....	<b>75</b>
<b>APÊNDICE A – Pseudocódigo do Algoritmo Genético 1</b> .....	<b>106</b>
<b>APÊNDICE B – Pseudocódigo do Algoritmo Genético 2</b> .....	<b>107</b>
<b>APÊNDICE C – Pseudocódigo do Algoritmo Genético 1 - Percentis</b> .....	<b>108</b>
<b>APÊNDICE D – Pseudocódigo do Algoritmo Genético 2 - Percentis</b> .....	<b>109</b>
<b>APÊNDICE E – Pseudocódigo do <i>Simulated Annealing</i> 1</b> .....	<b>110</b>
<b>APÊNDICE F – Pseudocódigo do <i>Simulated Annealing</i> 2</b> .....	<b>111</b>
<b>APÊNDICE G – Pseudocódigo do <i>Simulated Annealing</i> 1 - Percentis</b> .....	<b>112</b>
<b>APÊNDICE H – Pseudocódigo do <i>Simulated Annealing</i> 2 - Percentis</b> .....	<b>113</b>
<b>APÊNDICE I – Pseudocódigo do <i>Simulated Annealing</i> e <i>Random Forest</i></b> .....	<b>114</b>

## **PRIMEIRA PARTE**

## 1 INTRODUÇÃO GERAL

Esforços na compreensão e quantificação do crescimento das árvores são cruciais para o fornecimento de informações confiáveis, permitindo assim a gestão adequada dos recursos florestais. Em vista disso, salienta-se o interesse no crescimento e na produção das florestas para promover um planejamento criterioso da produção, por meio de prescrições de regimes de manejo adequados, que asseguram a qualidade do produto final (LIMA FILHO et al., 2012). Nesse sentido, a criação de modelos preditivos vem se constituindo em uma importante técnica, pois permite sintetizar conhecimentos, identificar lacunas, projetar e prever o comportamento frente a diversas condições. Dessa forma, tem sido fundamental o suporte dos modelos matemáticos para tomadas de decisões inerentes ao planejamento florestal (MAESTRI; SANQUETTA; ARCE, 2003).

Segundo Campos e Leite (2009) os modelos de crescimento e produção podem ser expressos de diferentes formas de acordo com as variáveis envolvidas. Os modelos em nível de povoamento são mais empregados por ser mais fácil a amostragem, porém não explica de forma direta a variação do tamanho das árvores dentro do povoamento. Os modelos de distribuição diamétrica estimam a quantidade de árvores e a produção de madeira por hectare por classe de diâmetro, e a produção total pode ser obtida por meio do somatório da variável em estudo em todas as classes. Nessa categoria são utilizadas as funções de densidade probabilidade que expressam a frequência por classe diamétrica pré-definida, sendo a função Weibull a mais empregada (MIGUEL et al., 2010). Esta função dispõe de diversos métodos de estimativa de seus parâmetros, em que cada método apresenta vantagens e desvantagens. Dessa maneira, vários estudos são elaborados com o objetivo de fornecer subsídios técnicos na obtenção dos parâmetros, e assim proporcionar mais confiabilidade na escolha do método de estimativa dos parâmetros da função.

Já a categoria de árvores individuais é mais complexa que as demais, no entanto, apresentam vantagens de produzir informações detalhadas sobre a dinâmica da estrutura e estoque do povoamento, visto que cada árvore é estudada considerando suas características individualmente (CAMPOS; LEITE, 2009). Além disso, esse nível de detalhe propicia um viés ecológico elevado quando aplicado, principalmente, em florestas nativas. Desta forma é possível entender melhor a dinâmica, sucessão e até processos ecofisiológicos de uma árvore ou de um conjunto (ORELLANA et al., 2016). Contudo, a quantificação das alterações da biomassa acima do solo, além de medições sucessivas, necessita também de equações

alométricas robustas capazes de estimar os incrementos de biomassa ou carbono das árvores (SILLET et al., 2019). Devido às florestas tropicais serem consideradas como uma fonte líquida de carbono, torna-se necessário o conhecimento dos estoques de carbono na biomassa florestal (RATUCHNE et al., 2015), o que pode ser realizado por meio da regressão e inteligência artificial, por exemplo.

Nos últimos anos, a aplicação de técnicas de aprendizado de máquina tem crescido no setor florestal e alcançado bons resultados. Essa técnica apresenta vantagens sob os métodos clássicos, como a independência de conhecimento e suposições sobre as relações entre as variáveis utilizadas para o ajuste. Além disso, permite o uso de variáveis categóricas e numéricas em um mesmo modelo e auxilia em estudos que envolvem um número elevado de variáveis, possibilitando, dessa forma, a seleção de variáveis e melhoria do desempenho dos modelos (CORDEIRO et al., 2015; HONG et al., 2018). Nesse contexto, a utilização dos métodos híbridos tem avançado para resolução de problemas com maior nível de complexidade, pois combinam uma série de estratégias. Os métodos híbridos têm apresentado grande eficiência e eficácia na resolução de problemas difíceis e em larga escala de diversas áreas da ciência, como projetos de engenharia, programação de mineração de dados e problemas de produção (MAFARJA; MIRJALILI, 2018). A exemplo disso têm-se os estudos de Mafarja e Mirjalili (2017) e Wang et al. (2017) que encontraram bons resultados na utilização de modelos híbridos para seleção de recursos.

Diante disso, este trabalho tem como objetivo aplicar algoritmos de aprendizado de máquina em problemas envolvendo estimativas de parâmetros da função Weibull e modelagem de atributos florestais. O primeiro artigo trata-se de avaliar as meta-heurísticas, algoritmo genético e *Simulated Annealing* na predição de parâmetros e identificação dos percentis ótimos para a função Weibull. Já o segundo artigo aborda a seleção de variáveis e modelagem do estoque de carbono acima do solo, em nível de árvore, para espécies da bacia do Rio Grande, Minas Gerais, por meio da construção de um método híbrido envolvendo o *Simulated Annealing* e *Random Forest*.

## **2 REVISÃO DE LITERATURA**

### **2.1 Estoque de carbono nas florestas**

Atualmente, sabe-se que as florestas desempenham um papel importante em relação ao sequestro biológico de C e na retirada do CO<sub>2</sub> da atmosfera, por meio de processos fisiológicos nos diferentes compartimentos da planta e também no solo (CHENG; WANG; JIANG, 2007; DINIZ et al., 2015; SANQUETTA et al., 2014). Além disso, segundo a FAO (2011), as florestas tropicais cobrem extensas áreas em todo planeta, cerca de 17 milhões de km<sup>2</sup>, o que corresponde a 20% da superfície terrestre. Esses ecossistemas florestais exercem um papel essencial no ciclo global de carbono (C), o qual contribui com aproximadamente, 36% das trocas líquidas de carbono entre a atmosfera e a vegetação terrestre. O ciclo global de carbono está relacionado com os ecossistemas tropicais que, ao mesmo tempo, emitem e absorvem o dióxido de carbono (CO<sub>2</sub>) atmosférico, realizando um balanço natural entre as taxas de respiração e produção. Sendo assim, as alterações resultantes de distúrbios naturais ou antrópicos provocam mudanças no fluxo de C (DELPIERRE et al., 2012; PAIVA; REZENDE; PEREIRA, 2011; VIEIRA et al., 2011). No Brasil, as mudanças do uso da terra, junto com o desmatamento, são as principais práticas que colaboram para as alterações desse fluxo (MATTHEWS et al., 2014).

A capacidade das florestas em sequestrar carbono da atmosfera faz com que amenize de forma eficiente as mudanças climáticas, por isso projetos como o REDD + (Redução de Emissões por Desmatamento e Degradação Florestal) foram desenvolvidos para remunerar os países e os projetos que inserissem conservação florestal em seus territórios, fazendo com que aumentem as reduções das emissões de CO<sub>2</sub>. Projetos como esse proporcionariam o desenvolvimento de países subdesenvolvidos, uma vez que a compra de créditos de carbono de países desenvolvidos compensariam suas próprias reduções de emissões (ANGELSEN, 2016; BENJAMINSEN; KAARHUS, 2018). Uma estratégia de sucesso já praticada diz respeito ao desenvolvimento ambientalmente sustentável, sendo um projeto envolvendo o pagamento por serviços ambientais (PSA). Essa modalidade tem importância no Brasil e no mundo e por isso tem sido explorado desde 2006. Os projetos REDD + são práticas voluntárias que visam a conservação e manejo adequado dos ecossistemas, sendo que o provedor desses serviços é retribuído por meio de remuneração financeira proveniente de compradores de serviços. Para tal, é necessário que haja fornecedores de serviços ambientais, assim como compradores. Esse projeto resulta na quantidade e qualidade no abastecimento de água a jusante, no sequestro de carbono e conservação da biodiversidade (PAGIOLA; GLEHN; TAFFARELLO, 2013).

As florestas contribuem no equilíbrio do ciclo de carbono, uma vez que esse é fixado na biomassa tanto acima, quanto abaixo do solo, além das árvores acumularem em sua estrutura orgânica elementos nocivos à qualidade de vida. Logo, ter conhecimento a respeito do estoque de C na biomassa florestal pode oferecer subsídios importantes para a tomada de decisões relacionadas ao manejo, à proteção e à valoração das florestas (RATUCHNE et al., 2015).

Nesse contexto, Scolforo et al. (2015) estimaram o estoque de C acima do solo, com base nos dados do Inventário Florestal de Minas Gerais, presentes nos biomas Mata Atlântica, Savana e Floresta Semiárida, por meio de técnicas de interpolação. Os valores médios encontrados foram de 35,05, 45,73 e 12,40 Megagramas (Mg, 1 Mg equivale a 1.000 Kg)·ha<sup>-1</sup> para os biomas mencionados, respectivamente. Já Moraes et al. (2013), ao avaliarem o estoque total de C em um fragmento de cerradão, também no estado de Minas Gerais, encontraram um valor médio de 36,78 MgC·ha<sup>-1</sup> o que corresponde a uma contribuição de 26,3% de C da parte aérea. Por outro lado, Ribeiro et al. (2011), ao utilizarem métodos de estimativa direta (método destrutivo) em uma área de cerrado no Norte de Minas Gerais, constataram quantidades de biomassa aérea entre 12,89 e 107,36 MgC·ha<sup>-1</sup>. Essa amplitude dos valores demonstra uma grande variação dessa fitofisionomia. Por isso, a quantificação mais precisa dos estoques de C e das metodologias aplicadas tornam-se importantes na determinação da extensão do sequestro de C pelas florestas (AMARO et al., 2013; CHINEMBIRI et al., 2013).

## 2.2 Métodos de predição

### 2.2.1 Análise de regressão

A análise de regressão é um método simples utilizado para estudar relações funcionais entre variáveis. Essa relação é apresentada na forma de um modelo em que associa uma variável dependente com uma ou mais variáveis explicativas ou preditoras (Equação 1), e pode ser representada por expressões lineares ou não lineares. A técnica tem o propósito de prever e/ou estimar o valor médio de um conjunto de dados (CHATTERJEE; HADI, 2006), sendo  $Y$  a variável dependente,  $X_n$  as variáveis explicativas em que  $n$  é o número de variáveis utilizadas, e  $\varepsilon$  é o erro.

$$Y = f(X_1, X_2, \dots, X_n) + \varepsilon \quad (1)$$

A regressão tem sido utilizada na estimativa de diversas variáveis, para a predição e prognose da produção florestal ao longo do tempo. Como exemplo, tem-se em Lee, Seo e Choi (2017) o uso de oito modelos clássicos em biometria florestal para *Pinus densiflora*, *Pinus koraiensis* e *Larix kaempferi*, e encontraram que o modelo de Spurr foi o mais preciso nas estimativas de volume. Zhang et al. (2016) construíram equações de biomassa acima do solo a partir dos modelos de volume para árvores, em que constataram bons ajustes dos modelos com valores de coeficiente de determinação acima de 0,90. As estimativas de volume em florestas nativas a partir de modelos de regressão, foram analisadas por Vibrans et al. (2015). Os autores testaram nove modelos de volume e identificaram coeficientes de correlação altamente significativos, bom comportamento residual para normalidade, aleatoriedade e homocedasticidade para todos os modelos. O modelo de Schumacher e Hall apresentou o melhor ajuste com menor valor de AIC (Critério de Informação de Akaike) tanto para modelos genéricos quanto os específicos por espécie. Esses resultados obtidos mostram a capacidade dos métodos de regressão em construir modelos e estimar seus parâmetros.

Contudo, quando se dispõe de dados com um grande número de variáveis e observações Gong, Zhang e Liu (2018) destacam que as técnicas tradicionais como os mínimos quadrados não podem ser aplicadas de forma direta a esse conjunto de dados, sendo necessário identificar as variáveis com maior importância. Além disso, as abordagens de regressão podem ter dificuldade em descrever a complexidade de uma floresta sendo uma alternativa a utilização de modelos de aprendizado de máquina (AHMED et al., 2015).

### **2.2.2 Inteligência Artificial**

Os métodos de inteligência de artificial são amplamente aplicados em diversas áreas da ciência, como no setor aeroespacial, petróleo, gás, energia e outras (SAMAGULINA; SAMIGULINA, 2018). No setor florestal, também tem sido utilizado uma gama de técnicas computacionais e de modelagem matemática para diferentes situações (LEITE et al., 2016), como exemplo a *artificial neural network* (DIAMANTOPOULOU et al., 2015), o *Random Forest* (AHMED et al., 2015), o *simulated annealing* (DONG et al., 2015) e *support vector machine* (CORDEIRO et al., 2015). Os métodos de aprendizado de máquinas têm sido utilizados para estimar o estoque de C, como Were et al., (2015), que avaliaram e compararam o desempenho do *Random Forest* (RF), *Support Vector Machines for Regression* (SVR) e Redes Neurais Artificiais (RNA) para previsão e mapeamento da variabilidade de estoques de

C do solo na Reserva Florestal Oriental de Mau, no Quênia. Os autores encontraram valores médios de 105,39 MgC·ha<sup>-1</sup> (RF); 103,75 MgC·ha<sup>-1</sup> (SVR); e 105,48 MgC·ha<sup>-1</sup> (RNA), e constataram que o SVR apresentou a menor tendência e o menor erro médio quadrático seguido do modelo de RNA. Outros pesquisadores também têm estudado os métodos de aprendizado de máquina como exemplo, Vahedi (2016) que utilizou a RNA e equações alométricas específicas do norte do Irã para verificar e examinar a precisão desses métodos na predição da biomassa acima do solo em florestas inequiâneas.

### 2.2.2.1 *Random Forest*

Um método bastante difundido na literatura é o *Random Forest* (RF), sendo um algoritmo popular e eficaz tanto para questões de classificação como regressão, fundamentado nos princípios de agregação de modelos. O RF combina várias árvores de decisão binária utilizando amostras de *bootstrap* resultantes da amostra de aprendizagem  $L$ , e seleciona aleatoriamente em cada nó da árvore um subconjunto de variáveis explicativas  $X$  (diferentemente da estratégia do modelo CART (*Classification and Regression Trees*) o *Random Forest*, em cada nó, um dado número de variáveis de entrada (denominado *mtry*) é selecionado aleatoriamente e a melhor divisão é calculada somente dentro deste subconjunto). Para que as árvores da floresta sejam árvores máximas nenhuma poda é realizada (BREIMAN, 2001; GENUER, POGGI, TULEAU-MALOT, 2010). Assim, cada árvore de decisão é produzida de forma independente, sem nenhuma remoção, e cada nó é separado ao usar uma quantidade de recursos (*mtry*). O número de árvores (*Ntree*) é estipulado pelo usuário, e o algoritmo gera árvores com alta variância e baixa polarização (BREIMAN, 2001; BELGIU, DRAGUT, 2016).

O bom desempenho do *Random Forest* é explicado por Breiman (2001), por meio da boa qualidade de cada árvore (baixo viés), mais a pequena correlação entre as árvores da floresta, que é descrita como a correlação de predições em amostras *out-of-bag* (OOB). Essa amostra é o grupo de observações que não são utilizadas para a construção da árvore atual, sendo usada para estimar o erro da previsão e avaliar a importância da variável (VI). A quantificação da importância da variável é uma questão relevante, para interpretação, em vários problemas que complementam a seleção de variáveis. O cálculo de VI mais empregado para uma dada variável é o acréscimo na média do erro de uma árvore (erro quadrático médio para regressão e taxa de classificação inexata para a classificação) na floresta. Os valores

analisados desta variável são randomicamente permutados nas amostras OOB. Um grande número de procedimentos de seleção de variáveis é fundamentado na colaboração da importância da variável para classificação e estimativa de modelos a fim de gerar, avaliar e comparar um conjunto de modelos (GENUER; POGGI; TULEAU-MALOT, 2010).

O *Random Forest* torna-se cada vez mais popular e mostra ser poderoso em diversas aplicações, como para análise dos dados de expressão gênica (DÍAZ-URIARTE; ALVAREZ DE ANDRÉS, 2006), previsão da ocorrência do tipo de vegetação com base nas condições ambientais (PETERS et al., 2008), classificação das práticas agrícolas baseado nas imagens de satélite *Landsat* (WATTS et al., 2009), no reconhecimento das fontes principais de contaminação por nitrato em águas subterrâneas (RODRIGUEZ-GALIANO et al., 2017), na modelagem da riqueza de espécies de plantas vasculares em uma floresta com o uso de dados de LiDAR (LOPATIN et al., 2016), entre outras.

A maioria dos trabalhos aplicando o RF busca entender as interações entre variáveis, identificar as variáveis preditoras, e a capacidade de predição que algumas vezes são superiores aos métodos já estabelecidos, como a regressão, análise discriminante linear, máquinas de vetor de suporte e redes neurais (HAPFELMEIER; ULM, 2013). Na área florestal, Görgens, Montagui, Rodriguez (2015) compararam o *random forest* (RF), as redes neurais artificiais (RNA), a regressão vetorial de suporte (SVM), e um modelo de regressão para a predição do volume de plantios de híbridos *Eucalyptus grandis* e *Eucalyptus urophylla*, com dados obtidos por meio de um *Airbone Laser Scanning* (ALS). Os quatro modelos testados obtiveram valores do coeficiente de determinação ( $R^2$ ) próximos, porém o RF e o modelo de regressão apresentaram os menores erros,  $31,83 \text{ m}^3.\text{ha}^{-1}$  e  $30,56 \text{ m}^3.\text{ha}^{-1}$ , respectivamente.

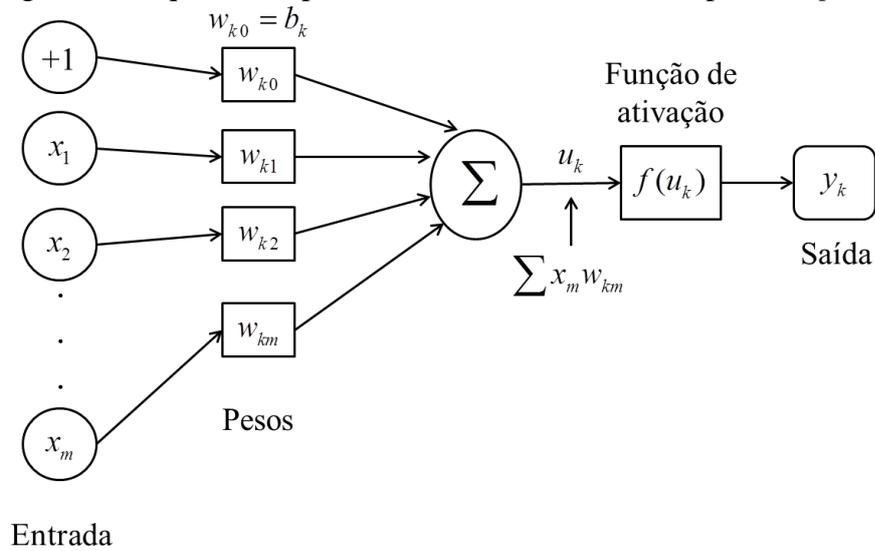
Zhang et al. (2017) aplicaram a regressão linear múltipla (MLR) e o *random forest* (RF) para prever a distribuição de carbono orgânico no solo com o uso de variáveis contínuas e categóricas, em uma área de recuperação no leste da China. Nos resultados encontrados, os dois modelos demonstraram capacidade de predição satisfatória, visto que estes apresentaram baixos índices de erro e altos valores de  $R^2$ , entretanto, o RF teve um desempenho superior. Este desempenho é justificado pelos autores, devido a habilidade do RF em lidar com relações não lineares e hierárquicas entre as variáveis. Resultado semelhante sobre a superioridade do RF em comparação com a regressão linear também foi constatado por Guo et al. (2015).

### 2.2.2.2 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNA), outro método bastante estudado, são sistemas computacionais paralelos constituídos de unidades de processamento simples, denominada de neurônios artificiais, que calculam algumas funções matemáticas geralmente não-lineares. O procedimento mais usual para a solução de problemas por meio de RNA inicia-se por uma fase de aprendizagem, onde um conjunto de exemplos é apresentado a rede, que extrai de forma automática as características necessárias para representar a informação e produzir a resposta do problema. Finalizada a fase de aprendizagem, a rede está pronta para posteriores aplicações, se mostrar a capacidade de generalizar, ou seja, fornecer respostas coerentes para dados desconhecidos (BRAGA; CARVALHO; LUDEMIR, 2007).

As camadas que compõem a rede são organizadas e interligadas por muitas conexões, que podem estar associados a pesos. Estes pesos armazenam o conhecimento expresso no modelo e ponderam a entrada recebida por cada neurônio da rede. O funcionamento das redes é baseado na estrutura física do cérebro humano (BRAGA; CARVALHO; LUDEMIR, 2007). Nesse caso, o uso de neurônio artificial (Figura 1) se resume em sinais de entrada, representados por  $x_m$ , em que cada uma é multiplicada por pesos ( $w_k$ ) positivos ou negativos em função das sinapses correspondentes. O *bias* ( $b$ ) é considerado como uma entrada com valor constante igual a 1 e é multiplicado por um peso igual a  $b_k$ . Posteriormente, ocorre a soma ponderada entre cada entrada e seus respectivos pesos, resultando em um valor  $u_k$ . Este valor é modificado por uma função de ativação que pode ser limiar, sigmoidal, hiperbólica ou semilinear. Ao final é produzido uma saída única  $y_k$  do neurônio artificial (BISHOP, 1995).

Figura 1 - Esquema simples de um neurônio artificial tipo *Perceptron*.



Fonte: Adaptado de Bishop (1995)

A arquitetura *Multilayer Perceptron* (MLP) é a mais usual dentre os tipos de rede, sendo composta por camada de entrada, oculta e saída, cada uma com um conjunto de nós conectados entre si trabalhando de forma paralela na transformação dos dados de entrada em valores de saída. A rede é treinada de forma supervisionada, ou seja, um conjunto de dados para os valores de entrada e saída esperados são fornecidos. O método consiste em primeiro atribuir os pesos às conexões de forma aleatória e escolher os parâmetros de entrada. O valor da saída é calculado e comparado com o esperado e então o erro é determinado. Este processo continua de maneira iterativa até que a convergência entre a saída calculada e a esperada seja alcançada (CHOI et al., 2012; CONFORTI et al., 2014; WERE et al., 2015).

O algoritmo utilizado para esse tipo de aprendizagem é o *backpropagation* (retropropagação), o qual modifica os pesos automaticamente a fim de reduzir o erro da saída, dessa forma a rede treinada classifica os dados adequadamente (CONFORTI et al., 2014). De acordo com Reis et al. (2016), as redes neurais artificiais obtiveram performance satisfatória na prognose do crescimento individual de árvores em florestas tropicais após colheita. Vahedi (2016) modelou a biomassa acima do solo em uma floresta no norte do Irã, utilizando as redes neurais artificiais em comparação com as equações alométricas específicas do local. O modelo de RNA demonstrou resultado mais preciso, apresentando menor erro médio quadrático. Miguel et al. (2015) analisaram a eficácia da aplicação de modelos de regressão e RNA associadas com imagens de satélite na predição do volume de madeira e biomassa acima do solo em um fragmento de cerradão no município de Palmas – TO. Os resultados obtidos

mostraram estimativas eficazes, visto que os valores reais não diferiram estatisticamente dos valores estimados para ambos os modelos, porém as RNA apresentaram maior acurácia.

### 2.2.2.3 *Support Vector Machine (SVM)*

A máquina de vetores de suporte (do inglês – *Support Vector Machine* – SVM) é um algoritmo que se caracteriza em um conjunto de métodos de aprendizado supervisionado relacionados à família dos classificadores lineares generalizados. É considerado um dos métodos mais robustos para serem utilizados na área de inteligência artificial. Os SVM's foram propostos originalmente por Vapnik (1995) e têm superado muitos dos sistemas em uma ampla variedade de aplicações. Diferentemente das técnicas tradicionais, que procuram aprimorar o desempenho sobre o conjunto de treinamento, os SVM's buscam minimizar a probabilidade de classificar de maneira errada padrões ainda não vistos pela distribuição de probabilidades de dados (MARCO JÚNIOR; SIQUEIRA, 2009).

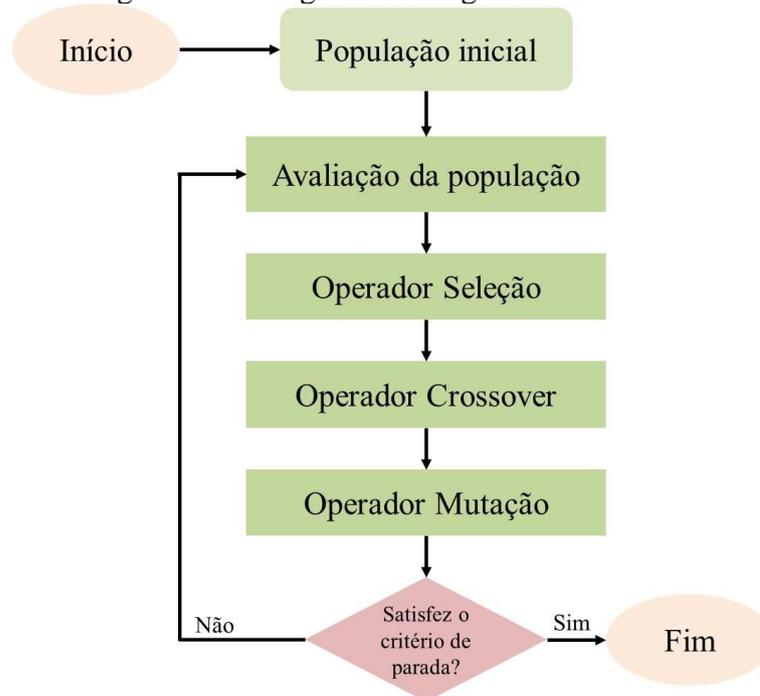
O SVM é baseado em um algoritmo que encontra o hiperplano que maximize a separação entre as classes. Os dados com distâncias mínimas desse hiperplano formam os vetores de suporte, que serão elementos críticos para o conjunto de treinamento. Se todos os outros elementos do treinamento fossem removidos, o mesmo hiperplano de separação seria encontrado (SANCHEZ-HENANDEZ; BOYD; FOODY, 2007; TIRELLI; GAMBÀ; PESSANI, 2012). Desta forma, é possível utilizar o SVM para obter alta precisão de classificação com um número reduzido de amostras de treinamento (ZHENG, et al. 2015).

O método SVM exhibe várias vantagens sobre outras técnicas de aprendizado de máquina: raramente alcança um sobreajuste (*overfitting*) dos dados de treinamento; fornece resultados mais competitivos que os melhores métodos aplicados para classificação; apresenta excelente desempenho de generalização na resolução de problemas não lineares e séries temporais; trabalha com um pequeno conjunto de dados (TIRELLI; GAMBÀ; PESSANI, 2012). Löw et al. (2013) utilizaram o SVM para classificação de culturas em paisagens irrigadas, juntamente com o RF, para a seleção de um subconjunto de características com a finalidade de obter maior acurácia. Os resultados encontrados evidenciaram um aumento na precisão e uma redução da classificação. Omer et al. (2017) compararam o desempenho da RNA e SVM na estimativa das concentrações de nitrogênio e carbono nas folhas em uma floresta nativa no norte da África do Sul. De acordo com os resultados obtidos, o método SVM superou a RNA com estimativas relativamente mais exatas.

### 2.2.2.4 Algoritmo Genético

O Algoritmo Genético (AG), proposto por Holland (1975), é um dos algoritmos de computação evolucionária utilizado com frequência, sendo amplamente empregado para produzir soluções úteis em problemas de otimização e busca (WANG et al., 2017). Assim, o AG é baseado nos princípios da evolução por meio da seleção natural e na “sobrevivência do mais apto”, em que utiliza uma população de indivíduos onde são submetidos a operadores genéticos, como seleção, crossover (cruzamento) e mutação, que gera alterações nos indivíduos e permite, dessa forma, com que a população alcance uma melhor área do espaço de busca (YU; XU, 2014; RAHMAN; SALEH, 2018). O funcionamento do AG pode ser demonstrado de acordo com o fluxograma apresentado na Figura 2.

Figura 2 - Fluxograma do Algoritmo Genético.



Fonte: Adaptado de Cheong e Koh (2019).

Este algoritmo é composto, principalmente, pelos componentes: codificação, função de aptidão (*fitness*), população, operadores genéticos e parâmetros de controle. Inicialmente, é realizada a codificação da solução candidata de forma adequada ao problema, por meio de cromossomos, ou seja, indivíduos. Esses indivíduos são representados por um vetor de comprimento definido, podendo ser codificado em variáveis binárias, forma mais usual, e

dessa maneira cada gene adquire o valor 0 ou 1 que caracteriza os atributos de cada indivíduo (ABDULHAMED; TAWFEEK; KESHK, 2018; ZHI; LIU, 2019). Após estabelecer a codificação, são gerados aleatoriamente os indivíduos que formam a população inicial. Em seguida, cada indivíduo é avaliado por meio de uma função de aptidão (*fitness*), que analisa quão próximo o indivíduo está da solução desejada. É de suma importância que a função de aptidão seja bastante representativa e diferencie de forma adequada as soluções boas e ruins, visto que, se houver baixa precisão na avaliação, as soluções ótimas podem ser rejeitadas durante o processamento do algoritmo e ainda explorar soluções não promissoras (KARIMI; KANI, 2019; REBOUÇAS FILHO et al., 2019).

Logo após, é aplicado o operador genético de seleção que tem por finalidade comparar os indivíduos da população inicial para seleção dos mais aptos que podem ser utilizados na iteração subsequente (GHAMISI; BENEDIKTSSON, 2015). Este operador pode ser realizado por meio de diversas técnicas, como o torneio e a roleta. A seleção por torneio é mais usual devido sua simplicidade e menor complexidade computacional, e funciona basicamente com a competição dos indivíduos escolhidos aleatoriamente. Depois de selecionados os indivíduos, é utilizado o operador genético crossover que atua de forma a misturar os atributos dos pais para produzir filhos (cromossomos) potenciais. Este operador é identificado como o esquema de busca mais importante e eficaz, pois permite a geração de novos cromossomos candidatos (CHUANG; CHEN; HWANG, 2015). E finalmente, o operador mutação que altera aleatoriamente um ou alguns genes dos cromossomos, de acordo com a probabilidade definida, e tem por objetivo impedir a convergência prematura da população, o que possibilita a oportunidade de escapar de ficar preso em um ótimo local (YU; XU, 2014).

A obtenção de bons resultados provenientes do AG é decorrente da definição correta dos parâmetros de probabilidade de crossover e mutação, pois essas indicam a frequência com que essas operações ocorram. Normalmente, o operador crossover recebe uma probabilidade próxima a 1, diferente do operador mutação que recebe baixo valor de probabilidade (BALIEIRO et al., 2014; METAWA; HASSAN; ELHOSENY, 2017). O processo é repetido ao longo de várias gerações, desde a etapa de avaliação até a mutação, para a obtenção de soluções cada vez melhores até que um critério de parada seja satisfeito. Alguns critérios de finalização podem ser adotados, como um número máximo de iterações ou quando a população convergir para uma solução que possui melhor valor da função de aptidão (CHUANG; CHEN; HWANG, 2015; MOHAMMADI et al., 2019).

Yalcinkaya, Senoglu e Yolcu (2018) utilizaram o AG para obter os estimadores de máxima verossimilhança dos parâmetros da distribuição Skew Normal (SN) e compararam a eficiência deste algoritmo com as técnicas iterativas tradicionais. Os autores concluíram que o AG é mais eficiente que os outros métodos iterativos avaliados, além disso, apontam que o AG também pode ser aplicado para outros tipos de distribuições.

### **2.2.2.5 Simulated Annealing**

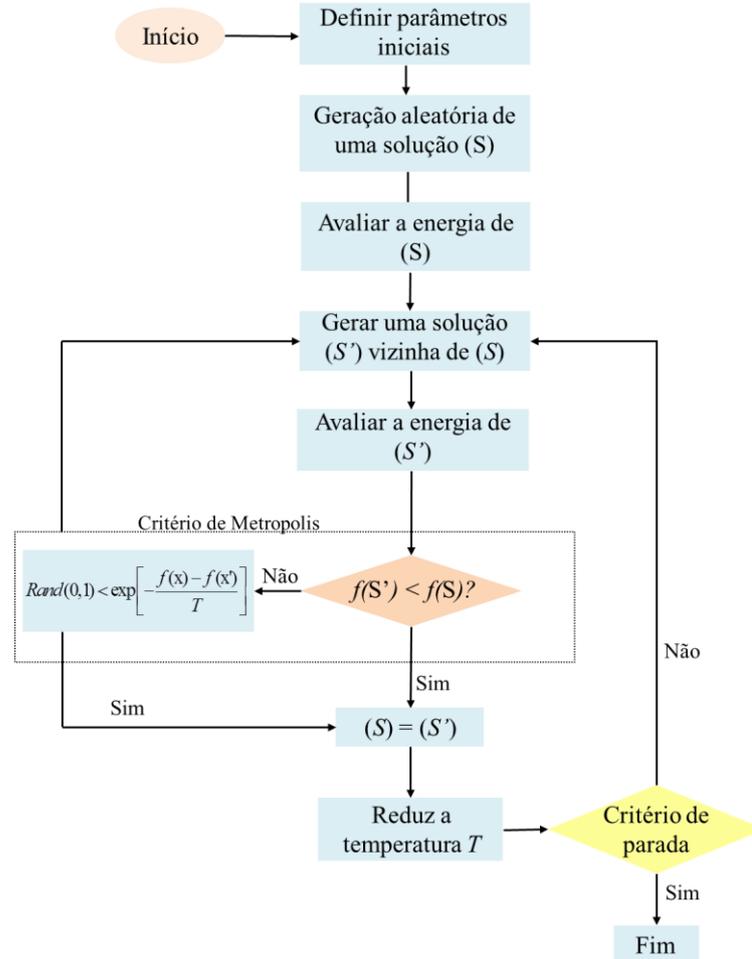
O *Simulated Annealing* (SA), ou recozimento simulado, é um algoritmo de busca local estocástico que, por meio de uma solução inicial pesquisa iterativamente a vizinhança da solução atual. O algoritmo é inspirado no estudo de Metropolis et al. (1953), que foi proposto a fim de descobrir a configuração de equilíbrio para um conjunto de átomos a uma temperatura definida. Posteriormente, Kirkpatrick, Gelatt e Vecchi (1983) e Černý (1985) aplicaram o SA para resolver problemas de otimização combinatória (ABBASI et al, 2011; FRANZIN; STÜTZLE, 2019). O funcionamento do SA pode ser representado na Figura 3.

A implementação do SA necessita da seleção de alguns parâmetros de controle, sendo essas a representação da solução, a temperatura inicial, a definição da função de avaliação, o cronograma de resfriamento e o critério de parada (ABBASI et al, 2011). O algoritmo começa a partir de uma solução inicial ( $S$ ) gerada aleatoriamente, sendo representada por um vetor dimensional com comprimento definido. A dimensão do mesmo está relacionada com o problema em estudo. Além disso, o parâmetro temperatura ( $T$ ) é determinado para um valor inicial ( $T_0$ ). Em seguida, é calculada a função de avaliação para a solução inicial  $S$ . A partir desse ponto, em cada iteração, uma solução vizinha ( $S'$ ) da solução atual  $S$  é produzida, em que também é calculada a função de avaliação. A melhor solução é sempre aceita e a pior também conforme um critério determinado pela probabilidade de Boltzmann e pela temperatura atual. A temperatura inicialmente permanece alta a fim de proporcionar que o algoritmo aceite soluções piores, e gradativamente a temperatura é reduzida por meio de uma taxa para assim diminuir a probabilidade de aceitar soluções piores. A redução da temperatura segue de acordo com o cronograma de resfriamento definido (MAFARJA; MIRJALILI, 2017; ASSAD; DEEP, 2018).

Nesse sentido, a determinação da taxa de resfriamento influencia na procura pelas soluções, pois quanto mais lenta for a taxa de redução da temperatura, maiores são as oportunidades de encontrar soluções ótimas, no entanto, será necessário um maior número de

iterações do SA e isso aumentaria o tempo computacional (COURCHELLE et al., 2019). Abbasi et al. (2006) declaram que, normalmente, é utilizada uma taxa que varia de 0,75 a 0,95.

Figura 3 - Fluxograma do algoritmo *Simulated Annealing*.



Fonte: Adaptado de Gomide, Arce e Silva (2013).

O processo é repetido inúmeras vezes até que um critério de parada seja atendido. Esse critério pode ser quando o sistema atingir o nível de energia pretendido, ou seja, uma temperatura de resfriamento. Além disso, também pode ser adotado como critério o número total de soluções geradas ou uma taxa de aceitação, em que leva em consideração o número de soluções aceitas em relação à quantidade de soluções geradas (ABBASI et al, 2011). Abbasi et al. (2006) aplicaram o SA para maximizar a função de verossimilhança com o objetivo de estimar os parâmetros da função Weibull, e assim propor uma metodologia para modelar a distribuição de falhas em estudos de confiabilidade. A partir dos resultados obtidos, os autores

concluíram que a metodologia envolvendo o SA produz bons resultados para esse problema produzindo estimativas precisas.

#### 2.2.2.6 Métodos híbridos

Nos últimos anos, devido à complexidade dos problemas aplicados, bem como o surgimento de ideias envolvendo a integração de algoritmos, tem-se o avanço na construção de métodos híbridos que combinam uma série de estratégias. Estes algoritmos tornaram-se extremamente úteis no reconhecimento de padrões e predição do comportamento de variáveis. Dentro desse ambiente fértil, as metaheurísticas apresentaram sua eficiência e eficácia na resolução de problemas difíceis e em larga escala no projeto de engenharia, aprendizado de máquina, programação de *data mining* e problemas de produção (MAFARJA e MIRJALILI, 2018). Nas metaheurísticas existem dois critérios que precisam ser considerados: o estudo de regiões ainda não exploradas dentro do espaço de busca, conhecido como diversificação, e a exploração concentrada das regiões com as melhores soluções encontradas, intensificação. No entanto, esses critérios são divergentes e o equilíbrio entre eles ocasionará em um aumento no desempenho do algoritmo de busca. Ao combinar pelo menos duas técnicas, obtém-se um aumento no desempenho de cada técnica, sendo, portanto, uma alternativa para atingir esse equilíbrio (MAFARJA; MIRJALILI, 2017). Frequentemente, percebe-se que o método híbrido tem capacidade de identificar uma boa solução, enquanto que utilizar somente um método normalmente prende-se a uma solução inadequada (KABIR; SHAHJAHAN; MURASE, 2011).

Nesse contexto, Hong et al. (2018) aplicaram o algoritmo genético (AG) para selecionar a melhor combinação de variáveis referente a incêndios florestais, e utilizaram os métodos *random forest* (RF) e a máquina de vetores de suporte (SVM) para elaboração de um mapa de susceptibilidade a incêndios florestais. Os modelos RF e SVM otimizados apresentaram resultados mais precisos em comparação com os modelos originais, sendo o RF otimizado superior a todos os outros. Bui et al. (2017) também apresentaram uma abordagem híbrida para modelagem espacial da suscetibilidade a incêndios florestais utilizando um sistema de inferência Neural Fuzzy (NF) para determinar o modelo de incêndio florestal, enquanto que a *Particle Swarm Optimization* (PSO) foi aplicado para investigar os melhores parâmetros do modelo. O desempenho foi avaliado pela área sob a curva (*Area under the curve ROC - AUC*) e várias medidas estatísticas, além de ser comparado com os métodos RF

e SVM. Os resultados encontrados mostraram maiores valores para as métricas de avaliação e maior desempenho do modelo proposto em comparação aos outros, sendo assim uma ferramenta alternativa promissora para modelagem da suscetibilidade de incêndios.

Na área de sensoriamento remoto, a classificação de imagens é um procedimento corriqueiro, contudo pode ser melhorado quando modelos híbridos são utilizados, sendo observado uma robustez e alto desempenho (Wang et al., 2017). Os autores anteriormente citados utilizaram uma técnica de sensoriamento remoto com base no SVM, no qual seus parâmetros e a seleção de atributos foram definidos por meio do algoritmo de otimização de colônia de formigas (*Ant Colony Optimization – ACO*) associado com o AG. Reddy, Gebreslasie e Ismail (2017) aplicaram uma abordagem híbrida com a regressão por mínimos quadrados parciais e RF para previsão de atributos de uma floresta de eucalipto a partir de dados de sensoriamento remoto, e constataram uma performance superior deste método na predição do volume e altura das árvores. Samadianfard, Ghorbani e Mohammadi (2018) observaram comportamento satisfatório do modelo de redes neurais artificiais em conjunto com o algoritmo de otimização de vagalumes (*Firefly Algorithm - FA*) para modelagem da temperatura do solo em diferentes profundidades.

Diversos estudos envolvendo modelos híbridos demonstraram melhor desempenho, quando comparados com os algoritmos de forma separada, assim como Mafarja e Abdullah (2013) que apresentaram uma hibridização com o *simulated annealing* (SA) para aprimorar a capacidade de busca local do AG. Manimala, Selvi e Ahila (2011) também utilizaram o AG e SA juntos para produzir um algoritmo de seleção de características e otimizar os parâmetros do SVM na classificação de distúrbios de energia. Recentemente, os algoritmos colônia de abelhas artificiais (*Artificial Bee Colony - ABC*) e evolução diferencial (*Differential Evolution - ED*) foram combinados com o objetivo de seleção de recursos para classificação (ZORARPACI; ÖZEL, 2016).

### **2.3 Distribuição Diamétrica**

A distribuição diamétrica de uma floresta, segundo Robinson (2004), compreende um histograma de frequência do diâmetro à altura do peito (DAP) das árvores, que é obtida a partir do agrupamento dos indivíduos em classes de diâmetro. As florestas podem apresentar diferentes tipos de distribuição de acordo com a estrutura, como exponencial negativa, unimodal e multimodal. Em relação a esses tipos de distribuição do diâmetro, a unimodal é

representada por um único valor para a moda sendo característica de florestas equiâneas. Já a distribuição decrescente, ou exponencial negativo, é determinada pela redução da frequência de indivíduos conforme o diâmetro aumenta, caracterizando assim uma floresta inequiânea em que ocorre regeneração contínua e apresenta uma diversidade de espécies em diferentes idades. Por último, a distribuição multimodal que indica mais de um ponto com maior frequência, mas é dificilmente encontrada na natureza (LOETSCH et al., 1973; TEO et al., 2012).

Uma das formas de descrever a estrutura do diâmetro de uma floresta ou de uma espécie é por meio do uso de funções de densidade probabilidade (fdp's), destacando-se as funções Weibull, Beta,  $S_B$ , Gamma, log-normal e Normal. Estas proporcionam a estimativa da probabilidade de árvores ocorrerem em cada classe de diâmetro predefinidas, sendo que as funções aplicadas para representar a distribuição diamétrica, deve ser adequadamente testada e selecionada com o objetivo de identificar qual delas apresenta maior consistência para caracterizar o comportamento do diâmetro (MACHADO et al., 2010). A escolha da função a ser utilizada leva em consideração as particularidades da floresta ou da espécie, o grau de dificuldade na estimativa dos parâmetros, a flexibilidade na descrição da amplitude de distribuição e o nível de complexidade para o cálculo de sua integral (BURKHART; TOMÉ, 2012).

De acordo com Arce (2004), a distribuição diamétrica é a ferramenta mais simples e poderosa para definir a estrutura da floresta. Nesse sentido, o conhecimento da estrutura diamétrica de uma floresta permite caracterizar o estoque em crescimento, identificar diferentes tipologias, definir a amplitude de variação do diâmetro, além de verificar a intensidade de regeneração natural (MACHADO et al., 2009; TEO et al., 2012). Sob o ponto de vista da produção, é possível indicar o estoque de madeira no período anterior à exploração, determinar o rendimento da floresta para diversos produtos e aplicação na modelagem de crescimento e produção.

Portanto, a distribuição diamétrica é uma importante técnica para subsidiar a tomada de decisão e o planejamento do manejo a ser empregado em uma determinada área (MACHADO et al., 2009; ORELLANA et al., 2017). Esse recurso tem sido utilizado em diversos trabalhos para vários tipos de floresta e região, alcançando diferentes objetivos (LIMA e LEÃO, 2013). A exemplo, Zhang e Lei (2010) identificaram que a distribuição Weibull é a que melhor explica a distribuição diamétrica da espécie *Pinus tabulaeformis*, e

Miguel et al. (2010) que verificaram bons resultados ao aplicar a função Weibull na prognose da produção em classes de diâmetro para plantios de *Eucalyptus urophylla*.

## 2.4 Distribuição Weibull

A distribuição Weibull foi proposta inicialmente por Fisher e Tippett em 1928, para estudos de valores extremos. Mais tarde, Waloddi Weibull desenvolveu de forma independente, em 1939, estudos sobre resistência dos materiais. Esses estudos ganharam destaque no período pós-guerra, que resultou na associação do nome de Waloddi Weibull (BAILEY; DELL, 1973). Em 1973, essa distribuição foi introduzida no meio florestal e desde então é utilizada em diversos trabalhos (PALAHÍ; PUKKALLA; BLASCO, 2007; GORGOSO-VARELA; ROJO-ALBORECA, 2014b; SANQUETTA et al., 2015; PACH; PODLASKI, 2015).

Na ciência florestal, a distribuição Weibull é amplamente empregada devido a simplicidade em relação à estimativa de seus parâmetros e a sua flexibilidade para se adaptar tanto para floresta nativa como plantada (OGANA; GORGOSO-VARELA, 2015). Esta pode ser apresentada com 3 parâmetros, sendo eles, os parâmetros locação ( $a$ ), escala ( $b$ ) e forma ( $c$ ). O parâmetro locação é relacionado ao diâmetro mínimo do povoamento e quando igual a zero denota-se distribuição Weibull 2 parâmetros, com início na origem (BURKHART; TOMÉ, 2012), e os parâmetros escala e forma são sempre valores positivos.

A função de densidade probabilidade Weibull de 2 parâmetros e sua função acumulativa é apresentada pelas Equações 2 e 3, em que  $x$ : variável aleatória, no caso diâmetro central da classe ( $x \geq 0$ );  $b$  e  $c$ : parâmetros escala e forma da distribuição ( $b, c > 0$ ), respectivamente.

$$f(x) = \left(\frac{c}{b}\right) \left(\frac{x}{b}\right)^{c-1} \exp\left[-\left(\frac{x}{b}\right)^c\right] \quad (2)$$

$$F(x) = 1 - \exp\left[-\left(\frac{x}{b}\right)^c\right] \quad (3)$$

A função de densidade probabilidade Weibull de 3 parâmetros e sua função acumulativa é apresentada pelas Equações 4 e 5, em que  $x$ : variável aleatória, no caso

diâmetro central da classe ( $x \geq 0$ );  $a$ ,  $b$  e  $c$ : parâmetros locação ( $a \geq 0$ ), escala e forma da distribuição ( $b, c > 0$ ), respectivamente.

$$f(x) = \left(\frac{c}{b}\right) \left(\frac{x-a}{b}\right)^{c-1} \exp\left[-\left(\frac{x-a}{b}\right)^c\right] \quad (4)$$

$$F(x) = 1 - \exp\left[-\left(\frac{x-a}{b}\right)^c\right] \quad (5)$$

O parâmetro escala ( $b$ ) está associado à amplitude de variação mínima e máxima da função, sendo que quanto menor o valor de  $b$  mais leptocúrtica será a curva de distribuição, e quanto maior for o valor de  $b$  mais platicúrtica será a distribuição. Já o parâmetro forma ( $c$ ) determina as várias formas que a distribuição pode apresentar. Assim, quando  $c$  assume valor próximo de 3,6, tem-se uma distribuição normal e simétrica, e quando menor que 3,6 a distribuição é assimétrica à direita e a esquerda caso seja maior que 3,6. A forma exponencial negativa é um caso particular quando  $c$  é igual a 1 (BAILEY; DELL, 1973; BURKHART; TOMÉ, 2012).

Existem diversas maneiras de ajustar a função Weibull, dos quais Acitas, Aladag e Senoglu (2019) mencionam a máxima verossimilhança (ML), método dos momentos (MM), máxima verossimilhança modificada (MML), mínimos quadrados (LS), percentis e outros. No entanto, Orellana et al. (2017) apontam que alguns métodos de ajuste podem apresentar melhores estimativas em comparação a outros, sendo assim, os estudos sobre distribuição diamétrica geralmente avaliam várias funções e diversos métodos, com o objetivo de encontrar a função e o método que melhor represente o conjunto de dados. O método de máxima verossimilhança é frequentemente considerado o mais adequado, pois é assintoticamente mais eficiente, sendo bastante empregado na estimativa dos parâmetros da função Weibull (LEI, 2008).

Gorgoso-Varela e Rojo-Alboreca (2014a) compararam a distribuição Weibull ajustada pelos métodos de máxima verossimilhança, momentos e percentis, e  $S_B$  de Johnson com ajuste pelos métodos de máxima verossimilhança condicional, momentos, moda e Knoebel-Burkhart para as espécies de *Quercus robur* e *Betula pubescens*. Os autores encontraram que Weibull ajustada por meio do método dos percentis apresentou melhor ajuste para *Quercus robur*, já a espécie *Betula pubescens* foi mais bem ajustada pela função  $S_B$  de Johnson pelos métodos dos percentis e máxima verossimilhança. Fonton et al. (2013) ao avaliarem os métodos de máxima verossimilhança, momentos e percentis para estimativa dos parâmetros

da função Weibull em um povoamento de *Tectona grandis* Lf. submetidos a dois regimes de manejo, concluíram que os três métodos proporcionam boas estimativas das distribuições, no entanto, o método dos momentos mostrou superioridade ao apresentar menores índices de erros.

Mayrinck et al. (2018) avaliaram a estrutura horizontal de povoamentos de *Khaya ivorensis* (mogno africano) por meio das funções Beta, Gamma,  $S_B$  de Johnson e Weibull ajustadas por diferentes métodos (máxima verossimilhança, método dos momentos, percentis, moda, regressão linear e Knoebel-Burkhart). Os autores constataram que os melhores métodos de ajuste foram os momentos e a máxima verossimilhança para a função  $S_B$  de Johnson e o método dos percentis para Weibull. Carretero e Alvarez (2013) também avaliaram diferentes métodos de ajuste (regressão não linear, máxima verossimilhança, momentos e percentis) da função Weibull para distribuição diamétrica de *Quercus suber* e encontraram que o método que obteve melhor desempenho foi o de regressão não linear.

Outros métodos estão sendo desenvolvidos e avaliados com o propósito de melhorar o processo de estimativa dos parâmetros da distribuição Weibull como exemplo, o estudo de Diamantopoulou et al. (2015) que aplicaram as Redes Neurais Artificiais na modelagem da distribuição do diâmetro e conseguiram estimativas precisas para os parâmetros escala e forma. Carvalho et al. (2016) também encontraram boa acurácia ao utilizar a técnica das Redes Neurais Artificiais na modelagem da estrutura diamétrica de florestas tropicais. Além de Acitas, Aladag e Senoglu (2019), que aplicaram a otimização de enxames de partículas para estimativa dos parâmetros da distribuição Weibull e alcançaram bons resultados.

#### **2.4.1 Método dos Momentos**

O método dos momentos é uma técnica utilizada para obter estimadores dos parâmetros da distribuição que é baseada na relação dos momentos da amostra com os momentos da distribuição correspondente, sendo a resolução obtida a partir de um sistema de  $k$  equações com  $k$  incógnitas. Esse método apresenta algumas desvantagens, como as estimativas fornecidas que podem estar fora do espaço dos parâmetros, principalmente para as amostras pequenas e, além disso, a obtenção das estimativas por meio desse método não consideram totalmente as informações compreendidas na amostra (ZASADA, 2013).

No meio florestal, o método dos momentos é bastante difundido devido à correlação dos parâmetros escala e forma com os momentos amostrais centrados na origem. No caso de

distribuições diamétricas, o primeiro momento amostral refere-se ao diâmetro médio do povoamento e o segundo ao diâmetro quadrático (LEI, 2008).

Os parâmetros escala e forma podem ser determinados diretamente pela média e variância da amostra, sendo a variância estimada a partir do diâmetro médio e quadrático. Uma vez que a média e a variância da distribuição do diâmetro são conhecidas em qualquer momento específico, o parâmetro forma é obtido por processo iterativo, e posteriormente o parâmetro escala é encontrado. A relação entre a média e o desvio padrão da amostra revela imediatamente a forma aproximada da distribuição (SHIFLEY; LENTZ, 1985; DIAMANTOPOULOU et al., 2015). O parâmetro locação é obtido de forma independente a partir de uma percentagem do diâmetro mínimo.

Arias-Rodil et al. (2018) avaliaram o método dos momentos da distribuição Weibull para plantios de *Pinus radiata* utilizando informações provenientes de LiDAR e constataram que apenas 28 a 40% das distribuições de diâmetro foram calculados com precisão, no entanto, os autores ressaltam que esse resultado pode ter sido influenciado pelas características das informações utilizadas. Sghaier et al. (2016) encontraram resultados semelhantes nas estimativas dos parâmetros da distribuição Weibull ao aplicar o método dos momentos e máxima verossimilhança.

#### **2.4.2 Método dos Percentis**

O método dos percentis é mais simples, claro e independente para estimar os parâmetros da distribuição Weibull quando comparado com outros métodos, como a da máxima verossimilhança e o dos momentos. Assim, para a estimativa dos parâmetros por meio desse método são necessárias apenas funções auxiliares com a finalidade de estimar os dois diâmetros percentis com segurança semelhante às funções para estimar os limites diamétricos, diâmetro mínimo e máximo, que também devem ser empregados nessa ou em quaisquer outras hipóteses de método (WENDLING; EMERENCIANO, HOSOKAWA, 2011). Quando não se conhece os valores dos percentis que descrevem a distribuição Weibull em relação ao conjunto de dados utilizado, deve-se testar várias combinações de percentis para o conjunto de dados que compõem a amostra.

Embora o método dos percentis apresente maior simplicidade para a estimativa dos parâmetros, nos últimos anos, encontram-se poucos trabalhos com a aplicação desse método. Dentre eles, Adeyemi e Adesoye (2016) que avaliaram alguns métodos de ajuste dos

parâmetros de Weibull para caracterizar o diâmetro de uma floresta no sul da Nigéria. Esses autores encontraram que o método baseado em percentis proporcionou melhores estimativas dos parâmetros, que envolveram os percentis 25 e 95.

Ogana e Gorgoso-Varela (2015) analisaram os métodos da máxima verossimilhança, o método dos momentos e o método dos percentis, para a distribuição Weibull em povoamentos de floresta nativa na Nigéria. Os autores concluíram que todos os métodos apresentaram bons ajustes para os parâmetros, no entanto, ao ponderar a simplicidade e a facilidade para obtenção das estimativas, o método dos percentis pode ser considerado como uma ferramenta viável a ser empregada.

### **2.4.3 Método Máxima Verossimilhança**

O estudo de Fisher, em 1922, atuou como pioneiro no desenvolvimento do método da máxima verossimilhança, sendo considerado um dos métodos mais versáteis e confiáveis. O autor propôs esse método como alternativa ao método dos momentos e dos mínimos quadrados, uma vez que o primeiro era arbitrário na escolha das equações dos momentos já o segundo por não ser invariável sob as mudanças de escala das variáveis. Os estimadores de máxima verossimilhança consistem na busca pelos valores de parâmetros que maximizam a função de máxima verossimilhança, a partir dos dados disponíveis (RINNIE, 2008).

O método de máxima verossimilhança é um procedimento bastante difundido para distribuição Weibull na silvicultura, visto que apresenta certas propriedades desejáveis como estimadores consistentes, assintoticamente normais e eficientes. (LEI, 2008; GUERA et al., 2018). De acordo com Cao e McCarty (2006), as estimativas produzidas por esse método são estatisticamente consistentes o que confere melhor qualidade de ajuste, em comparação aos outros métodos, no entanto, demanda maiores recursos computacionais. Razali, Salih e Mahdi (2009) compararam os métodos dos momentos, máxima verossimilhança e mínimos quadrados para estimar os parâmetros da função Weibull de 2 e 3 parâmetros. Os resultados obtidos mostraram que o método dos momentos proporcionou melhores estimativas devido apresentar menores valores de erro quadrático médio. Já Castro et al. (2017) estudaram a distribuição diamétrica de *Eucalyptus* spp. considerando diferentes métodos de ajuste (máxima verossimilhança e aproximação linear) da função Weibull. Os autores encontraram que a função Weibull com 3 parâmetros foi melhor ajustada pelo método de máxima verossimilhança, com 96% de aderência ao teste de Kolmogorov – Smirnov. Diversos

trabalhos na área florestal utilizaram o método de máxima verossimilhança como Ribeiro et al. (2014), Cheng et al. (2016) e Souza et al. (2016).

## 2.5 Métricas de avaliação da aderência das distribuições

Na avaliação da qualidade do ajuste das distribuições pode-se utilizar os testes de aderência, em que são fundamentais para analisar se a distribuição de probabilidade da variável de interesse pode ser representada por meio de uma função de distribuição de probabilidade conhecida (CARGNELUTTI FILHO; MATZENAUER; TRINDADE, 2004). Os testes de aderência apresentam estatísticas e critérios de decisão diferentes, no entanto as hipóteses testadas são iguais. A hipótese de nulidade é de que a variável aleatória adere à distribuição específica, já a hipótese alternativa é oposta a de nulidade (TORMAN; COSTER; RIBOLDI, 2012).

Dentre os vários testes utilizados para avaliar a aderência das distribuições, são encontrados os testes: Kolmogorov – Smirnov, Índice de Reynolds, Qui-quadrado entre outros. De acordo com Frey (2016), o teste de Kolmogorov – Smirnov (KS) é mais frequentemente utilizado. Esse teste consiste na comparação da distribuição acumulativa observada com a distribuição acumulativa estimada, obtida pela função de densidade probabilidade. O ponto que apresenta a maior divergência, em módulo, entre as duas distribuições é a estatística do teste. Se o valor de KS encontrado for menor que o valor tabelado, que depende da quantidade dos dados e do nível de significância escolhida, então a função de densidade probabilidade possui aderência aos dados (DIAMANTOPOULOU et al., 2015). Esse teste não paramétrico é mais sensível nas regiões próximas da mediana da distribuição do que nas caudas e, portanto deve ser utilizado apenas para as distribuições contínuas. No entanto, diversos trabalhos na área florestal tem aplicado o teste KS para determinar a distribuição diamétrica, em que encontraram resultados satisfatórios na utilização desse método (CAO, 2004; BINOTI et al., 2012; GORGOSO-VARELA; ROJO-ALBORECA, 2014a; RIBEIRO et al., 2014).

O Índice de Reynolds proposto por Reynolds et al. (1988) é definido como a soma ponderada das diferenças absolutas entre as distribuições observada e estimada em cada classe. A ponderação pode ser feita baseada no volume ou valor (ZHANG; PACKARD; LIU, 2003), e área basal (TORRES-ROJO; TORRES-MARGAÑA; MIRELES, 2000). Assim, um baixo valor para o Índice de Reynolds indica pequena diferença entre as distribuições

observada e estimada (QIN; CAO; BLOUIN, 2007). Além disso, Orellana et al. (2017) afirmam que essa ferramenta é adequada e de fácil aplicação em análises de distribuições diamétricas. Outros estudos também utilizaram esse índice para avaliação das funções de densidade probabilidade, como Sghaier et al. (2016), Maltamo et al. (2017) e Pogoda, Ochal e Orzel (2019).

O teste de qui-quadrado ( $\chi^2$ ) foi desenvolvido por Karl Pearson e é baseado na soma do quadrado dos desvios das frequências observada e estimada. A estatística de  $\chi^2$  é dada pela Equação 6, em que:  $k$  é o número de classes,  $Fo_i$  é a frequência observada da  $i$ -ésima classe,  $Fe_i$  é a frequência estimada da  $i$ -ésima classe. O valor de  $\chi^2$  obtido é comparado com um valor tabelado, conforme o nível de significância preestabelecido e os graus de liberdade da distribuição. No entanto, esse teste apresenta limitações quando dispõe de dados com poucas classes, sendo dessa forma mais específico para dados agrupados, uma vez que as classes com valores menores que três ou cinco são agrupadas em outras classes (CATALUNHA et al., 2002). Algumas aplicações do teste de qui-quadrado para avaliar distribuições do diâmetro podem ser verificadas nos trabalhos de Soares et al. (2010), Zheng e Zhou (2010) e Mohammadi, Fallahchai e Hashemi (2016).

$$\chi^2 = \sum_{i=1}^k \left( \frac{(Fo_i - Fe_i)^2}{Fe_i} \right) \quad (6)$$

A partir do Diagrama de Taylor (TAYLOR, 2001) é possível comparar e avaliar o desempenho de diferentes modelos ou métodos. Esse diagrama permite resumir graficamente o grau de correspondência entre um conjunto de dados estimados em relação ao conjunto de dados observados. A similaridade entre os dois conjuntos de dados é quantificada a partir de três estatísticas: coeficiente de correlação, desvio padrão e a raiz do erro quadrático médio. O ângulo de azimute do diagrama mostra o coeficiente de correlação entre o conjunto de dados observado e estimado, e a distância radial do ponto de referência representa a razão do desvio padrão normalizada dos dados estimados em relação aos dados observados (MOAZENZADEH et al., 2018; YASEEN et al., 2018). Logo, todos os modelos (ou métodos) utilizados podem ser visualizados, de acordo com a proximidade da variável de interesse (YASEEN et al., 2018). Assim, este diagrama possibilita a escolha de determinado modelo (ou método) (ZHANG et al., 2019).

### 3 CONSIDERAÇÕES GERAIS

A utilização de inteligência artificial constitui uma importante ferramenta para o manejo florestal, permitindo avanços em metodologias já difundidas na literatura. No entanto, a área carece de novos estudos que analisem a capacidade de resolução de problemas complexos e até mesmo superar os métodos tradicionais. Nesse sentido, este capítulo procurou fornecer conceitos relacionados ao estudo, abordando as temáticas de inteligência artificial associada ao ajuste de parâmetros da função de densidade probabilidade Weibull e a modelagem de estoque de carbono. Assim, estes conceitos foram essenciais para aprofundamento dos conhecimentos teóricos que darão subsídios aos próximos capítulos.

### REFERÊNCIAS

- ABBASI, B. et al. A hybrid variable neighborhood search and simulated annealing algorithm to estimate the three parameters of the Weibull distribution. **Expert Systems With Applications**, United States, v. 38, n. 1, p.700-708, 2011.
- ABBASI, B. et al. Estimating the parameters of Weibull distribution using simulated annealing algorithm. **Applied Mathematics And Computation**, New York, v. 183, n. 1, p.85-93, 2006.
- ABDULHAMED, A. A.; TAWFEEK, M. A.; KESHK, A. E. A genetic algorithm for service flow management with budget constraint in heterogeneous computing. **Future Computing And Informatics Journal**, [s.l.], v. 3, n. 2, p.341-347, 2018.
- ACITAS, S.; ALADAG, C. H.; SENOGLU, B. A new approach for estimating the parameters of Weibull distribution via particle swarm optimization: An application to the strengths of glass fibre data. **Reliability Engineering & System Safety**, England, v. 183, p.116-127, 2019.
- ADEYEMI, A. A.; ADESOYE, P. O. Evaluation of Some Weibull Parameter Estimation Methods for Characterizing Stem Diameter Distribution in a Tropical Mixed Forest of Southern Nigeria. **Journal Of Statistical Science And Application**, Germany, v. 4, n. 6, p.257-275, 2016.
- AHMED, O. S. et al. Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm. **Isprs Journal Of Photogrammetry And Remote Sensing**, Amsterdam, v. 101, p.89-101, mar. 2015.
- AMARO, M. A. et al. Estoque volumétrico, de biomassa e de carbono em uma floresta estacional semidecidual em viçosa, minas gerais. **Revista Árvore**, Viçosa, v. 37, n. 5, p.849-857, 2013.

- ANGELSEN, A. REDD+ as Result-based Aid: General Lessons and Bilateral Agreements of Norway. **Review Of Development Economics**, United Kingdom, v. 21, n. 2, p.237-264, 2016.
- ARCE, J. E. Modelagem da estrutura de florestas clonais de *Populus deltoides* Marsh. através de distribuições diamétricas probabilísticas. **Ciência Florestal**, Santa Maria, v. 14, n. 1, p.149-164, 2004.
- ARIAS-RODIL, M. et al. Modeling diameter distributions in radiata pine plantations in Spain with existing countrywide LiDAR data. **Annals Of Forest Science**, France, v. 75, n. 2, p.1-12, 2018.
- ASSAD, A.; DEEP, K. A Hybrid Harmony search and Simulated Annealing algorithm for continuous optimization. **Information Sciences**, New York, v. 450, p.246-266, 2018.
- BAILEY, R. L.; DELL, R. Quantifying Diameter Distributions with the Weibull Function. **Forest Science**, Bethesda, v. 19, n. 2, p. 97–104, 1972.
- BALIEIRO, A. et al. A multi-objective genetic optimization for spectrum sensing in cognitive radio. **Expert Systems With Applications**, United States, v. 41, n. 8, p.3640-3650, 2014.
- BELGIU, M.; DRAGUT, L. Random forest in remote sensing: A review of applications and future directions. **Isprs Journal Of Photogrammetry And Remote Sensing**, Amsterdam, v. 114, p.24-31, abr. 2016.
- BENJAMINSEN, G.; KAARHUS, R. Commodification of forest carbon: REDD+ and socially embedded forest practices in Zanzibar. **Geoforum**, United Kingdom, v. 93, p.48-56, 2018.
- BINOTI, D. H. B. et al. Probability density functions for description of diameter distribution in thinned stands of *Tectona grandis*. **Cerne**, Lavras, v. 18, n. 2, p.185-196, 2012.
- BISHOP, M. **Neural networks for pattern recognition Christopher**. Oxford: Oxford University, 1995. 482 p.
- BRAGA, A. de P.; CARVALHO, A. P. de L. F.; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: LTC, 2007. 262 p.
- BREIMAN, L. Random Forests. *Machine learning*, v. 45, n. 1, p. 5-32, 2001.
- BUI, D. T. et al. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. **Agricultural And Forest Meteorology**, New Haven, v. 233, p.32-44, 2017.
- BURKHART, H. E.; TOMÉ, M. **Modeling forest trees and stands**. London: Springer Science & Business Media, 2012.
- CAMPOS, J. C. C.; LEITE, H. G. **Mensuração florestal: perguntas e respostas**. 3.ed. atual. e ampl. Viçosa, MG: UFV, 2009. 548 p.

CAO, Q. V. Predicting parameters of a Weibull function for modeling diameter distribution. **Forest Science**, Bethesda, v. 50, n. 5, p. 682-685, 2004.

CAO, Q. V.; MCCARTY, S. M. New methods for estimating parameters of Weibull functions to characterize future diameter distributions in forest stands. In: **Proc. of the Southern Silviculture Research Conference**. 2006. p. 338-340.

CARRETERO, A. C.; ALVAREZ, T. E. Modelling diameter distributions of *Quercus suber* L. stands in “Los Alcornocales” Natural Park (Cádiz-Málaga, Spain) by using the two-parameter Weibull function. **Forest Systems**, Madrid, v. 22, n. 1, p.15-24, 2013.

CARGNELUTTI FILHO, A.; MATZENAUER, R.; TRINDADE, J. K. da. Ajustes de funções de distribuição de probabilidade à radiação solar global no Estado do Rio Grande do Sul. **Pesquisa Agropecuária Brasileira**, Brasília, v. 39, n. 12, p.1157-1166, 2004.

CARVALHO, M. C. et al. Modelagem da distribuição diamétrica de florestas tropicais. **Enciclopédia Biosfera**, Goiânia, v. 13, n. 24, p.731-745, 2016.

CASTRO, R. V. O. et al. Procedimentos de ajuste da função weibull para prognose da distribuição diamétrica de povoamentos de eucalipto. **Scientia Agraria Paranaensis**, Cascavel, v. 16, n. 4, p. 528-534, 2017.

CATALUNHA, J. C. et al. Avaliação de cinco funções densidade de probabilidade a séries de precipitação pluvial no estado de Minas Gerais. **Revista Brasileira de Agrometeorologia**, Santa Maria, v. 10, n. 1, p. 153-162, 2002.

ČERNÝ, V. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. **Journal Of Optimization Theory And Applications**, United States, v. 45, n. 1, p.41-51, 1985.

CHATTERJEE, S.; HADI, A. S. **Regression Analysis by Example**. Nova Jersey: Ed. John Wiley & Sons, 2006. 383p.

CHENG, C. et al. A Weibull model of the impact of thinning and a typhoon event on the stand structure of *Cryptomeria japonicain* Central Taiwan over 100 years. **Journal Of Forest Research**, Japan, v. 22, n. 1, p.22-29, 2016.

CHENG, C.; WANG, R.; JIANG, J. Variation of soil fertility and carbon sequestration by planting *Hevea brasiliensis* in Hainan Island, China. **Journal Of Environmental Sciences**, Edmonton, v. 19, n. 3, p.348-352, 2007.

CHEONG, K. H.; KOH, J. M. A hybrid genetic-Levenberg Marquardt algorithm for automated spectrometer design optimization. **Ultramicroscopy**, Netherlands, v. 202, p.100-106, 2019.

CHINEMBIRI, T. S. et al. The Precision of C Stock Estimation in the Ludhikola Watershed Using Model-Based and Design-Based Approaches. **Natural Resources Research**, United States, v. 22, n. 4, p.297-309, 2013.

CHOI, J. et al. Combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using ASTER images and GIS. **Engineering Geology**, Netherlands, v. 124, p.12-23, 2012.

CHUANG, Y.; CHEN, C.; HWANG, C. A real-coded genetic algorithm with a direction-based crossover operator. **Information Sciences**, New York, v. 305, p.320-348, 2015.

CONFORTI, M. et al. Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). **Catena**, Amsterdam, v. 113, p.236-250, 2014.

CORDEIRO, M. A. et al. Estimativa do volume de *Acacia mangium* utilizando técnicas de redes neurais artificiais e máquinas vetor de suporte. **Pesquisa Florestal Brasileira**, Colombo, v. 35, n. 83, p.255-261, 2015.

COURCHELLE, V. et al. A simulated annealing approach to 3D strategic aircraft deconfliction based on en-route speed changes under wind and temperature uncertainties. **Transportation Research Part C: Emerging Technologies**, England, v. 103, p.194-210, 2019.

DELPIERRE, N. et al. Quantifying the influence of climate and biological drivers on the interannual variability of carbon exchanges in European forests through process-based modelling. **Agricultural And Forest Meteorology**, New Haven, v. 154-155, p.99-112, 2012.

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

DÍAZ-URIARTE, R.; ANDRÉS, S. A. de. Gene selection and classification of microarray data using random forest. **Bmc Bioinformatics**, Londres, v. 7, n. 1, p.1-13, 2006.

DINIZ, A. R. et al. Biomassa, estoques de carbono e de nutrientes em estádios sucessionais da Floresta Atlântica, RJ. **Revista Brasileira de Ciências Agrárias - Brazilian Journal Of Agricultural Sciences**, Recife, v. 4, n. 2, p.443-451, 2015.

DONG, L. et al. A comparison of a neighborhood search technique for forest spatial harvest scheduling problems: A case study of the simulated annealing algorithm. **Forest Ecology And Management**, Amsterdam, v. 356, p.124-135, 2015.

FONTON, N. H. et al. Modeling the Distribution of Marketable Timber Products of Private Teak (*Tectona grandis* L.f.) Plantations. **Open Journal Of Forestry**, Irvine, v. 03, n. 04, p.115-121, 2013.

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS - FAO. **State of the World's Forests**. Rome, 2011. 179p. Disponível em: n<<http://www.fao.org/docrep/013/i2000e/i2000e00.htm> >. Acesso em: 05 set. 2018.

- FRANZIN, A.; STÜTZLE, T. Revisiting simulated annealing: A component-based analysis. **Computers & Operations Research**, England, v. 104, p.191-206, 2019.
- FREY, J. An exact Kolmogorov–Smirnov test for whether two finite populations are the same. **Statistics & Probability Letters**, Netherlands, v. 116, p.65-71, 2016.
- GENUER, R.; POGGI, J. M.; TULEAU-MALOT, C. Variable selection using random forests. **Pattern Recognition Letters**, Uppsala, v. 31, n. 14, p.2225-2236, 2010.
- GHAMISI, P.; BENEDIKTSSON, J. A. Feature Selection Based on Hybridization of Genetic Algorithm and Particle Swarm Optimization. **Ieee Geoscience And Remote Sensing Letters**, United States, v. 12, n. 2, p.309-313, 2015.
- GOMIDE, L. R.; ARCE, J. E.; SILVA, A. C. L. da. Comparação entre a meta-heurística simulated annealing e a programação linear inteira no agendamento da colheita florestal com restrições de adjacência. **Ciência Florestal**, Santa Maria, v. 23, n. 2, p.451-462, 2013.
- GONG, S.; ZHANG, K.; LIU, Y. Efficient test-based variable selection for high-dimensional linear models. **Journal Of Multivariate Analysis**, United States, v. 166, p.17-31, 2018.
- GÖRGENS, E. B.; MONTAGHI, A.; RODRIGUEZ, L. C. E. A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. **Computers And Electronics In Agriculture**, England, v. 116, p.221-227, 2015.
- GORGOSO-VARELA, J. J.; ROJO-ALBORECA, A. A comparison of estimation methods for fitting Weibull and Johnson's SB functions to pedunculate oak (*Quercus robur*) and birch (*Betula pubescens*) stands in northwest Spain. **Forest Systems**, Madrid, v. 23, n. 3, p.500-505, 2014a.
- GORGOSO-VARELA, J. J.; ROJO-ALBORECA, A. Use of Gumbel and Weibull functions to model extreme values of diameter distributions in forest stands. **Annals Of Forest Science**, France, v. 71, n. 7, p.741-750, 2014b.
- GUERA, O. G. M. et al. Evolução da distribuição diamétrica em plantios de *Pinus caribaea* Morelet var. *caribaea* Barrett & Golfari. **Biofix Scientific Journal**, Curitiba, v. 3, n. 1, p.161-171, 2018.
- GUO, P. et al. Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach. **Geoderma**, Amsterdam, v. 237-238, p.49-59, 2015.
- HAPFELMEIER, A.; ULM, K. A new variable selection approach using Random Forests. **Computational Statistics & Data Analysis**, Netherlands, v. 60, p.50-69, 2013.
- HONG, H. et al. Applying genetic algorithms to set the optimal combination of forest fire related variables and model forest fire susceptibility based on data mining models. The case of Dayu County, China. **Science Of The Total Environment**, New York, v. 630, p.1044-1056, 2018.

KABIR, Md. M.; SHAHJAHAN, Md.; MURASE, K. A new local search based hybrid genetic algorithm for feature selection. **Neurocomputing**, New York, v. 74, n. 17, p.2914-2928, 2011.

KARIMI, H.; KANI, I. M. Finding the worst imperfection pattern in shallow lattice domes using genetic algorithms. **Journal Of Building Engineering**, Netherlands, v. 23, p.107-113, 2019.

LEE, D.; SEO, Y.; CHOI, J. Estimation and validation of stem volume equations for *Pinus densiflora*, *Pinus koraiensis*, and *Larix kaempferi* in South Korea. **Forest Science And Technology**, United Kingdom, v. 13, n. 2, p.77-82, 2017.

LEI, Y. Evaluation of three methods for estimating the Weibull distribution parameters of Chinese pine (*Pinus tabulaeformis*). **Journal Of Forest Science**, Czech Republic, v. 54, n. 12, p.566-571, 2008.

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

LIMA, J. P. da C.; LEÃO, J. R. A. Dinâmica de crescimento e distribuição diamétrica de fragmentos de floresta nativa e plantada na Amazônia Sul Ocidental. **Floresta e Ambiente**, Seropédica, v. 20, n. 1, p.70-79, 2013.

LIMA FILHO, L. M. de A. et al. Modelagem do crescimento de clones de *Eucalyptus* usando o modelo de Chapman-Richards com diferentes distribuições simétricas dos erros. **Ciência Florestal**, Santa Maria, v. 22, n. 4, p.777-785, 2012.

LOETSCH, F.; ZÖHRER, F.; HALLER, K. E. **Forest inventory**. 2.ed. Munich: BLV Verlagsgesellschaft, 1973. 469 p.

LOPATIN, J. et al. Comparing Generalized Linear Models and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile. **Remote Sensing Of Environment**, Toronto, v. 173, p.200-210, 2016.

MACHADO, S. do A. et al. Distribuição diamétrica de *Araucaria angustifolia* (Bert.) O. Ktze. em um fragmento de floresta ombrófila mista. **Scientia Agraria**, Curitiba, v. 10, n. 2, p.103-110, 2009.

MACHADO, S. do A. et al. Modelagem da Distribuição Diamétrica de Quatro Espécies de Lauraceae em um Fragmento de Floresta Ombrófila Mista. **Revista Ciências Exatas e Naturais**, Guarapuava, v. 12, n. 1, p. 91-105, 2010.

MAESTRI, R.; SANQUETTA, C. R.; ARCE, J. E. Modelagem do crescimento de povoamentos de *Uucalyptus grandis* através de processos de difusão. **Floresta**, Curitiba, v. 33, n. 2, p.169-182, 2003.

MAFARJA, M. M.; MIRJALILI, S. Hybrid Whale Optimization Algorithm with simulated annealing for feature selection. **Neurocomputing**, New York, v. 260, p.302-312, 2017.

MAFARJA, M.; ABDULLAH, S. Investigating memetic algorithm in solving rough set attribute reduction. **International Journal Of Computer Applications In Technology**, Genebra, v. 48, n. 3, p.195-202, 2013.

MAFARJA, M.; MIRJALILI, S. Whale optimization approaches for wrapper feature selection. **Applied Soft Computing**, Netherlands, v. 62, p.441-453, 2018.

MALTAMO, M. et al. Airborne laser scanning for tree diameter distribution modelling: a comparison of different modelling alternatives in a tropical single-species plantation. **Forestry: An International Journal of Forest Research**, United Kingdom, v. 91, n. 1, p.121-131, 2017.

MANIMALA, K.; SELVI, K.; AHILA, R. Hybrid soft computing techniques for feature selection and parameter optimization in power quality data mining. **Applied Soft Computing**, Netherlands, v. 11, n. 8, p.5485-5497, 2011.

MARCO JÚNIOR, P.; SIQUEIRA, M. F. Como determinar a distribuição potencial de espécies sob uma abordagem conservacionista? **Megadiversidade**, São Paulo, v.5, n.1/2, p. 65-76, 2009.

MATTHEWS, H. D. et al. National contributions to observed global warming. **Environmental Research Letters**, England, v. 9, n. 1, p.1-9, 2014.

MAYRINCK, R. C. et al. A comparison of diameter distribution models for *Khaya ivorensis* A.Chev. plantations in Brazil. **Southern Forests: a Journal of Forest Science**, United Kingdom, v. 80, n. 4, p.373-380, 2018.

METAWA, N.; HASSAN, M. K.; ELHOSENY, M. Genetic algorithm based model for optimizing bank lending decisions. **Expert Systems With Applications**, United States, v. 80, p.75-82, 2017.

METROPOLIS, N. et al. Equation of State Calculations by Fast Computing Machines. **The Journal Of Chemical Physics**, United States, v. 21, n. 6, p.1087-1092, 1953.

MIGUEL, E. P. et al. Redes neurais artificiais para a modelagem do volume de madeira e biomassa do cerradão com dados de satélite. **Pesquisa Agropecuária Brasileira**, Brasília, v. 50, n. 9, p.829-839, 2015.

MIGUEL, E. P. et al. Using the Wweibull function for prognosis of yield by diameter class in *Eucalyptus urophylla* stands. **Cerne**, Lavras, v. 16, n. 1, p.94-104, 2010.

MOAZENZADEH, R. et al. Coupling a firefly algorithm with support vector regression to predict evaporation in northern Iran. **Engineering Applications Of Computational Fluid Mechanics**, Peoples R China, v. 12, n. 1, p.584-597, 2018.

MOHAMMADI, F.; FALLAHCHAI, M. M.; HASHEMI, S. A. Application of probability distribution in order to fit the diameter and height of oak species in two natural and man-made

stands in Hyrcanian forests. **International Journal of Biomathematics**, Singapore, v. 9, n. 03, p. 1-9, 2016.

MOHAMMADI, S. et al. Genetic algorithm to select a set of closure relationships in multiphase flow models. **Journal Of Petroleum Science And Engineering**, Netherlands, v. 181, p.1-10, 2019.

MORAIS, V. A. et al. Carbon and biomass stocks in a fragment of cerrado in Minas Gerais state, Brazil. **Cerne**, Lavras, v. 19, n. 2, p.237-245, 2013.

OGANA, F. N.; GORGOSO-VARELA, J. J. Comparison of estimation methods for fitting Weibull distribution to the natural stand of Oluwa Forest Reserve, Ondo State, Nigeria. **Journal of Research in Forestry, Wildlife and Environment**, Makurdi, v. 7, n. 2, p. 81-90, 2015.

OMER, G. et al. Mapping leaf nitrogen and carbon concentrations of intact and fragmented indigenous forest ecosystems using empirical modeling techniques and WorldView-2 data. **Isprs Journal Of Photogrammetry And Remote Sensing**, Amsterdam, v. 131, p.26-39, 2017.

ORELLANA, E. et al. Métodos de ajuste e procedimentos de seleção de funções probabilísticas para modelar a distribuição diamétrica em floresta nativa de Araucária. **Ciência Florestal**, Santa Maria, v. 27, n. 3, p.969-980, 2017.

ORELLANA, E. et al. Predicting the dynamics of a native Araucaria forest using a distance-independent individual tree-growth model. **Forest Ecosystems**, United Kingdom, v. 3, n. 1, p.1-11, 2016.

PACH, M.; PODLASKI, R. Tree diameter structural diversity in Central European forests with *Abies alba* and *Fagus sylvatica*: managed versus unmanaged forest stands. **Ecological Research**, Japan, v. 30, n. 2, p.367-384, 2014.

PAGIOLA, S.; GLEHN, H. C. V.; TAFFARELLO, D. **Experiências de pagamentos por serviços ambientais no Brasil**. São Paulo: SMA/CBRN, 2013.

PAIVA, A. O.; REZENDE, A. V.; PEREIRA, R. S. Estoque de carbono em cerrado sensu stricto do Distrito Federal. **Revista Árvore**, Viçosa, v. 35, n. 3, p.527-538, 2011.

PALAHÍ, M.; PUKKALA, T.; BLASCO, E. Comparison of beta, Johnson's SB, Weibull and truncated Weibull functions for modeling the diameter distribution of forest stands in Catalonia (north-east of Spain). **European Journal Of Forest Research**, United States, v. 126, n. 4, p.563-571, 2007.

PETERS, J. et al. Random forests as a tool for ecohydrological distribution modelling. **Ecological Modelling**, Towson, v. 207, n. 2-4, p.304-318, 2007.

POGODA, P.; OCHAŁ, W.; ORZEŁ, S. Modeling Diameter Distribution of Black Alder (*Alnus glutinosa* (L.) Gaertn.) Stands in Poland. **Forests**, Switzerland, v. 10, n. 5, p.1-16, 2019.

QIN, J.; CAO, Q. V.; BLOUIN, D. C. Projection of a diameter distribution through time. **Canadian Journal Of Forest Research**, Ottawa, v. 37, n. 1, p.188-194, 2007.

RAHMAN, I.; MOHAMAD-SALEH, J. Hybrid bio-Inspired computational intelligence techniques for solving power system optimization problems: A comprehensive survey. **Applied Soft Computing**, Netherlands, v. 69, p.72-130, 2018.

RATUCHNE, L. C. et al. Quantification of forest carbon on planted forests of *Araucaria angustifolia* in southwest of state Paraná. **Ambiência**, Guarapuava, v. 11, n. 2, p.321-335, 2015.

Razali, A. M.; Salih, A. A.; Mahdi, A. A. Precisão da estimativa dos parâmetros de distribuição Weibull. **Journal of Applied Sciences Research**, Faisalabad, v. 5, n. 7, p. 790-795, 2009.

REBOUÇAS FILHO, P. P. et al. Control of singularity trajectory tracking for robotic manipulator by genetic algorithms. **Journal Of Computational Science**, Amsterdam, v. 30, p.55-64, 2019.

REDDY, N.; GEBRESLASIE, M.; ISMAIL, R. A hybrid partial least squares and random forest approach to modelling forest structural attributes using multispectral remote sensing data. **South African Journal Of Geomatics**, Mowbray, v. 6, n. 3, p.377-394, 2017.

REIS, L. P. et al. Prognosis on the diameter of individual trees on the eastern region of the amazon using artificial neural networks. **Forest Ecology And Management**, Amsterdam, v. 382, p.161-167, 2016.

REYNOLDS JR., M. R.; BURK, T. E.; HUANG, W. C. Goodness-of-fit tests and model selection procedures for diameter distribution models. **Forest Science**, v.34, n.2, p.373-399, 1988.

RIBEIRO, A. et al. Estrutura da distribuição diamétrica em plantio experimental de candeia (*Eremanthus erythropappus* (DC.) MacLeish). **Ciência Florestal**, Santa Maria, v. 24, n. 4, p.1055-1065, 2014.

RIBEIRO, S. C. et al. Above- and belowground biomass in a Brazilian Cerrado. **Forest Ecology And Management**, Amsterdam, v. 262, n. 3, p.491-499, 2011.

RINNIE, H. **The Weibull distribution: a handbook**. Taylor & Francis Group, London, New York, p. 402-454, 2008.

ROBINSON, A. Preserving correlation while modelling diameter distributions. **Canadian Journal Of Forest Research**, Ottawa, v. 34, n. 1, p.221-232, 2004.

RODRIGUEZ-GALIANO, V.F. et al. Feature selection approaches for predictive modelling of groundwater nitrate pollution: An evaluation of filters, embedded and wrapper methods. **Science Of The Total Environment**, New York, v. 624, p.661-672, 2018.

- SAMADIANFARD, S.; GHORBANI, M. A.; MOHAMMADI, B. Forecasting soil temperature at multiple-depth with a hybrid artificial neural network model coupled-hybrid firefly optimizer algorithm. **Information Processing In Agriculture**, Amsterdam, p.1-12, 2018.
- SAMIGULINA, G. A.; SAMIGULINA, Z. I. Modified immune network algorithm based on the Random Forest approach for the complex objects control. **Artificial Intelligence Review**, Netherlands, v. 52, n. 4, p.2457-2473, 2018.
- SANCHEZ-HERNANDEZ, C.; BOYD, D. S.; FOODY, G. M. Mapping specific habitats from remotely sensed imagery: Support vector machine and support vector data description based classification of coastal saltmarsh habitats. **Ecological Informatics**, Netherlands, v. 2, n. 2, p.83-88, 2007.
- SANQUETTA, C. R. et al. Estoques de biomassa e carbono em povoamentos de acácia negra em diferentes idades no Rio Grande do Sul. **Scientia Forestalis**, Piracicaba, v. 42, n. 103, p. 361-370, 2014.
- SANQUETTA, C. R. et al. Probabilistic modelling of diameter increment for *Pinus* sp. stands over a cycle of 28 years in Southern Brazil. **Australian Journal of Basic and Applied Sciences**, Punjab, v. 9, n. 23, p. 625-632, 2015.
- SCOLFORO, H. F. et al. Spatial Distribution of Aboveground Carbon Stock of the Arboreal Vegetation in Brazilian Biomes of Savanna, Atlantic Forest and Semi-Arid Woodland. **Plos One**, San Francisco, v. 10, n. 6, p.1-20, 2015.
- SGHAIER, T. et al. Modelling diameter distribution of *Tetraclinis articulata* in Tunisia using normal and Weibull distributions with parameters depending on stand variables. **Iforest - Biogeosciences And Forestry**, Italy, v. 9, n. 5, p.702-709, 2016.
- SHIFLEY, S.; LENTZ, E. Quick estimation of the three-parameter Weibull to describe tree size distributions. **Forest Ecology And Management**, Amsterdam, v. 13, n. 3-4, p.195-203, 1985.
- SILLETT, S. C. et al. Aboveground biomass dynamics and growth efficiency of *Sequoia sempervirens* forests. **Forest Ecology And Management**, Amsterdam, p.1-24, 2019.
- SOARES, T. S. et al. Projeção da distribuição diamétrica e produção de povoamentos de eucalipto empregando diferentes formas da função Weibull. **Revista Árvore**, Viçosa, v. 35, n. 5, p.1027-1032, 2011.
- SOUZA, L. A. de et al. Funções densidade de probabilidade para a descrição da distribuição de diâmetros em povoamentos florestais desbastados. **Revista Brasileira de Biometria**, Lavras, v. 34, n. 3, p. 421-434, 2016.
- TAYLOR, K. E. Summarizing multiple aspects of model performance in a single diagram. **Journal of Geophysical Research**, United States, v.106, n.7, p.7183-7192, 2001.

TÉO, S. J. et al. Desempenho de funções de densidade probabilísticas para descrever a distribuição diamétrica de *Pinus taeda*, na região de Caçador, SC. **Floresta**, Curitiba, v. 42, n. 4, p.741-754, 2012.

TIRELLI, T.; GAMBA, M.; PESSANI, D. Support vector machines to model presence/absence of *Alburnus alburnus alborella* (Teleostea, Cyprinidae) in North-Western Italy: Comparison with other machine learning techniques. **Comptes Rendus Biologies**, Paris, v. 335, n. 10-11, p.680-686, 2012.

TORMAN, V. B. L.; COSTER, R.; RIBOLDI, J. Normalidade de variáveis: métodos de verificação e comparação de alguns testes não-paramétricos por simulação. **HCPA**, Porto Alegre, v. 32, n. 2, p. 227-234, 2012.

TORRES-ROJO, J. M.; TORRES MARGAÑA, O. S.; MIRELES, M. A. Metodologia para mejorar la predicción de parametros de distribuciones diametricas. **Revista Agrociencia**, Texcoco-México, v. 34, n. 5, p. 627-637, 2000.

VAHEDI, A. A. Artificial neural network application in comparison with modeling allometric equations for predicting above-ground biomass in the Hyrcanian mixed-beech forests of Iran. **Biomass And Bioenergy**, United Kingdom, v. 88, p.66-76, 2016.

VAPNIK, V. N. **The nature of statistical learning theory**. Berlin: Springer Velag, 1995. 314 p.

VIBRANS, A. C. et al. Generic and specific stem volume models for three subtropical forest types in southern Brazil. **Annals Of Forest Science**, France, v. 72, n. 6, p.865-874, 2015.

VIEIRA, S. A. et al. Stocks of carbon and nitrogen and partitioning between above- and belowground pools in the Brazilian coastal Atlantic Forest elevation range. **Ecology And Evolution**, Agra, v. 1, n. 3, p.421-434, 2011.

WANG, M. et al. Remote sensing image classification based on the optimal support vector machine and modified binary coded ant colony optimization algorithm. **Information Sciences**, New York, v. 402, p.50-68, 2017.

WATTS, J. D. et al. Monitoring of cropland practices for carbon sequestration purposes in north central Montana by Landsat remote sensing. **Remote Sensing Of Environment**, Toronto, v. 113, n. 9, p.1843-1852, 2009.

WENDLING, W. T.; EMERENCIANO, D. B.; HOSOKAWA, R. T. Ajuste da função de distribuição diamétrica Weibull por planilha eletrônica. **Floresta**, Curitiba, v. 41, n. 2, p.205-220, 2011.

WERE, K. et al. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. **Ecological Indicators**, Amsterdam, v. 52, p.394-403, 2015.

YALÇdNKAYA, A.; ŐENOđLU, B.; YOLCU, U. Maximum likelihood estimation for the parameters of skew normal distribution using genetic algorithm. **Swarm And Evolutionary Computation**, Amsterdam, v. 38, p.127-138, 2018.

YASEEN, Z. M. et al. Predicting compressive strength of lightweight foamed concrete using extreme learning machine model. **Advances In Engineering Software**, New York, v. 115, p.112-125, 2018.

YU, F.; XU, X. A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network. **Applied Energy**, Amsterdam, v. 134, p.102-113, 2014.

ZASADA, M. Evaluation of the double normal distribution for tree diameter distribution modeling. **Silva Fennica**, Helsinki, v. 47, n. 2, p. 1-17, 2013.

ZHANG, C. et al. Developing Aboveground Biomass Equations Both Compatible with Tree Volume Equations and Additive Systems for Single-Trees in Poplar Plantations in Jiangsu Province, China. **Forests**, Switzerland, v. 7, n. 2, p.1-15, 2016.

ZHANG, H. et al. Prediction of soil organic carbon in an intensively managed reclamation zone of eastern China: A comparison of multiple linear regressions and the random forest model. **Science Of The Total Environment**, New York, v. 592, p.704-713, 2017.

ZHANG, J. et al. Modelling uniaxial compressive strength of light weight self-compacting concrete using random forest regression. **Construction And Building Materials**, New York, v. 210, p.713-719, 2019.

ZHANG, L.; PACKARD, K. C; LIU, C. A comparison of estimation methods for fitting Weibull and Johnson's SB distributions to mixed spruce fir stands in northeastern North America. **Canadian Journal Of Forest Research**, Ottawa, v. 33, n. 7, p.1340-1347, 2003.

ZHANG, X.; LEI, Y. A linkage among whole-stand model, individual-tree model and diameter-distribution model. **Journal Of Forest Science**, Czech Republic, v. 56, n. 12, p.600-608, 2010.

ZHENG, B. et al. A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. **International Journal Of Applied Earth Observation And Geoinformation**, Enschede, v. 34, p.103-112, 2015.

ZHENG, L.; ZHOU, X. Diameter distribution of trees in natural stands managed on polycyclic cutting system. **Forestry Studies In China**, Berlin, v. 12, n. 1, p.21-25, 2010.

ZHI, H.; LIU, S. Face recognition based on genetic algorithm. **Journal Of Visual Communication And Image Representation**, Orlando, v. 58, p.495-502, 2019.

ZORARPACI, E.; ŐZEL, S. A. A hybrid approach of differential evolution and artificial bee colony for feature selection. **Expert Systems With Applications**, Pergamon, v. 62, p.91-103, 2016.

**SEGUNDA PARTE – ARTIGOS**

**ARTIGO 1 – ESTIMAÇÃO DE PARÂMETROS DA FUNÇÃO WEIBULL  
UTILIZANDO META-HEURÍSTICAS**

**Artigo formatado conforme a NBR 6022 (ABNT, 2003) e adaptado às exigências do  
Manual de Normalização de Trabalhos Acadêmicos da UFLA.**

## RESUMO

A função Weibull tem sido utilizada de forma recorrente na descrição das distribuições diamétricas devido à sua flexibilidade e relativa simplicidade. Entretanto, a função depende de bons métodos de ajuste, visto que apresenta não linearidade em sua forma acumulada. Assim, os objetivos do estudo foram avaliar a viabilidade e desempenho do Algoritmo Genético (AG) e *Simulated Annealing* (SA) na predição dos parâmetros da função Weibull, identificar os percentis ótimos pelas meta-heurísticas (método híbrido), testar duas estratégias de construção da função a ser minimizada pelas meta-heurísticas, e analisar o desempenho e robustez dos métodos frente as condições de desbaste em plantios comerciais. Foram utilizados dados de povoamentos de *Pinus taeda* localizados em Campo Belo do Sul – SC. A função Weibull de 2 parâmetros foi então ajustada pelos seguintes métodos: Momentos, Percentis, SA<sub>1</sub>, SA<sub>2</sub>, AG<sub>1</sub>, AG<sub>2</sub>, SA<sub>1</sub> – Percentis, SA<sub>2</sub> – Percentis, AG<sub>1</sub> – Percentis e AG<sub>2</sub> – Percentis. Esses métodos foram avaliados pelo teste de Kolmogorov – Smirnov (KS), pelo Índice de Erro (EI), por um procedimento de classificação para estabelecer qual método é mais preciso pelo teste KS, além de análises gráficas. A partir dos resultados obtidos, todos os métodos de ajuste podem ser empregados, contudo os métodos que forneceram estimativas mais precisas foram SA<sub>2</sub>, AG<sub>2</sub>, SA<sub>2</sub> – Percentis para os dados submetidos aos diferentes regimes silviculturais avaliados. Os métodos híbridos resultaram em valores de 20 a 50 para o percentil 1 e 60 para o percentil 2, com a aplicação de desbaste e o avanço da idade. Dessa forma, recomenda-se o *Simulated Annealing* com a função de avaliação 2 devido os menores erros encontrados e menor tempo de processamento.

**Palavras-chave:** Manejo Florestal. Diâmetro. Algoritmo Genético. *Simulated Annealing*.

## ABSTRACT

The Weibull function has been used on a recurring basis in the description of diametric distributions due to its flexibility and relative simplicity. However, the function depends on good adjustment methods, as it presents non-linearity in its accumulated form. Thus, the objectives of this study were to evaluate the viability and performance of the Genetic Algorithm (AG) and *Simulated Annealing* (SA) in the prediction of the parameters of the Weibull function, to identify the great percentiles by the meta-heuristics (hybrid method), testing two construction strategies of the function to be minimized by the meta-heuristics, and to analyze the performance and robustness of the methods given the thinning conditions in commercial plantations. Data from *Pinus taeda* stands located in Campo Belo do Sul - SC were used. The 2-parameter Weibull function was adjusted by the following methods: Moments, Percentile, SA<sub>1</sub>, SA<sub>2</sub>, AG<sub>1</sub>, AG<sub>2</sub>, SA<sub>1</sub> - Percentile, SA<sub>2</sub> - Percentile, AG<sub>1</sub> - Percentile and AG<sub>2</sub> - Percentile. These methods were evaluated by the Kolmogorov - Smirnov (KS) test, by the Error Index (EI), by a classification procedure to establish which method is more accurate by the KS test, in addition to graphical analyses. From the results obtained, all adjustment methods can be used, however the methods that provided more accurate estimates were SA<sub>2</sub>, AG<sub>2</sub>, SA<sub>2</sub> - Percentiles for the data submitted to the different silvicultural regimes evaluated. The hybrid methods resulted in values of 20 to 50 for the 1 percentile and 60 for the 2 percentile, with the application of thinning and advancing age. Thus, *Simulated Annealing* with evaluation function 2 is recommended due to the smallest errors found and the shortest processing time.

**Keywords:** Forest management. Diameter. Genetic Algorithm. *Simulated Annealing*.

## 1. INTRODUÇÃO

A estrutura de um plantio comercial de floresta é uma formação complexa, que envolve a distribuição espacial entre as árvores, que disputam recursos para o seu crescimento. Nesse ambiente, suas interações competitivas são sintetizadas no porte e biomassa individual, que se adaptam as condições locais e de sítio para a manutenção das funções fisiológicas. O resultado dessas interações ecológicas são padrões heterogêneos de crescimento (OGANA; GORGOSO-VARELA, 2015), que moldam a ocupação do solo por estas árvores. Logo, o conhecimento da distribuição do diâmetro das árvores é uma importante ferramenta para caracterizar um povoamento, possibilitando o planejamento e o controle das atividades florestais (TÉO et al., 2012), seja no âmbito ecológico, ou econômico.

Usualmente, o diâmetro é uma variável universal e de fácil obtenção, sendo proveniente da medição direta das árvores, e apresenta boa correlação com outras variáveis importantes como volume, carbono, área basal e biomassa (BAILEY; DELL, 1973, POUDEL; CAO, 2013). Devido a sua forte relação funcional com uma gama de variáveis e aplicações no manejo florestal e silvicultura, há um interesse específico em analisar e descrever seu comportamento por meio de funções de densidade probabilidade.

Historicamente, essas funções foram introduzidas na área florestal em estudos de produtividade em termos de densidade, volume, biomassa e carbono (NASCIMENTO et al., 2014), permitindo assim auxiliar na obtenção da probabilidade de ocorrência do número de árvores em intervalos de classe predefinidos (LOETSCH; ZÖHRER; HALLER, 1973). Embora diversas funções de densidade probabilidade tenham sido utilizadas para modelar a distribuição do diâmetro, como Hiperbólica (LEITE et al., 2010),  $S_B$  de Jhonson (OGANA; ITAM; OSHO, 2017), Gama (ARAÚJO JÚNIOR et al., 2013), a mais utilizada e promissora delas é Weibull, devido à sua efetividade (ARIAS-RODIL et al., 2018). Essa função foi introduzida por Bailey e Dell (1973), sendo amplamente aplicada no meio florestal e incorporada em modelos de prognose em diversos trabalhos nacionais e internacionais.

Weibull apresenta algumas características desejáveis, como alta precisão, boa correlação entre os parâmetros e os atributos do povoamento florestal, e uma grande flexibilidade para assumir diferentes formas e assimetrias, de acordo com os valores de seus parâmetros, possibilitando dessa forma, sua utilização em diferentes tipologias florestais (NASCIMENTO et al., 2012). O número de parâmetros que a compõe pode ser observado com dois ou três, sendo que a formulação com dois parâmetros tende a ser mais precisa para

modelar diâmetros em diferentes espécies (DIAMANTOPOULOU et al., 2015; ARIAS-RODIL et al., 2018). Pode-se encontrar ainda funções segmentadas e truncadas, conforme trabalhos de BINOTI et al. (2012), PALAHÍ et al. (2006), PALAHÍ et al. (2007). Uma boa função ainda depende de bons métodos de ajuste, já que sua natureza tende a não linearidade em sua forma acumulada. Os métodos clássicos de ajuste incluem a máxima verossimilhança (MIRZAEI et al., 2016), métodos dos momentos (FERRAZ FILHO et al., 2015), método dos percentis (MIGUEL et al., 2010), não linear e outros (ACITAS; ALADAG; SENOGLU, 2019). Dessa forma, há espaço para a aplicação e teste de métodos não clássicos que envolve treinamento de máquina (DIAMANTOPOULOU et al., 2015) e meta-heurísticas (ABBASI et al., 2006; ACITAS; ALADAG; SENOGLU, 2019).

O uso de meta-heurísticas tem despertado interesse na predição de parâmetros, visto que apresenta fácil utilização e fornecem boas soluções em problemas reais. Diante dessas vantagens, a aplicação de meta-heurísticas pode propiciar um ganho na modelagem em termos de confiança de ajuste da distribuição, o que as tornam poderosas ferramentas. Como exemplo têm-se os trabalhos de Abassi et al. (2006) e Chang-Jun, Xin e Xin (2017) que utilizaram meta-heurísticas no ajuste dos parâmetros da função Weibull e encontraram bons resultados. Entretanto, um dos desafios dessa abordagem consiste em sua parametrização. De modo a verificar a eficiência dessas técnicas, este estudo tem como motivação averiguar as seguintes hipóteses: a) as meta-heurísticas possuem desempenho superior na estimativa de parâmetros para diferentes condições de desbaste quando comparadas aos métodos clássicos, e b) a performance das meta-heurísticas é determinada conforme a função de avaliação utilizada.

Nesse contexto, os objetivos do estudo consistem em: a) avaliar a viabilidade e desempenho do Algoritmo Genético e *Simulated Annealing* na predição dos parâmetros (escala e forma) da função Weibull, e comparar métodos clássicos (Momentos e Percentis); b) identificar os percentis ótimos pelas meta-heurísticas e aplicar o método dos percentis (método híbrido); c) testar duas estratégias de construção da função a ser minimizada pelas meta-heurísticas, envolvendo área basal e densidade; e d) avaliar o desempenho e robustez dos métodos frente as condições de desbaste em plantios comerciais.

## **2. MATERIAL E MÉTODOS**

### **2.1 Área de estudo e descrição dos dados**

As informações utilizadas neste estudo são provenientes de povoamentos de *Pinus taeda*, localizados no município de Campo Belo do Sul – SC, a 27°59'33" latitude sul e 50°54'16" longitude oeste, com cerca de 950 metros de altitude. De acordo com a classificação de Köppen, a região é caracterizada por um clima do tipo “Cfb”, subtropical úmido, com verões frescos, chuvas bem distribuídas durante todo o ano e ocorrência de geadas (ALVARES et al., 2013). As médias anuais de temperatura e precipitação estão em torno de 16°C e 1800 mm, respectivamente. A região apresenta solos do tipo Nitossolo Háplico, Neossolo Litólico e Latossolo Bruno (DOBNER Jr., 2013).

Os dados foram obtidos de parcelas com 1.000 m<sup>2</sup> de área útil, em experimento de desbaste, sendo estruturado em blocos casualizados, com três tratamentos e duas repetições, implantado com espaçamento de 2,5 x 1,6 m (densidade inicial de 2.500 árvores.ha<sup>-1</sup>). No experimento foi mensurado, anualmente, o diâmetro à altura do peito (DAP) de todas as árvores, entre os 5 e 16 anos, aos 18 anos e, finalmente, aos 30 anos de idade, dessa forma o número de medições feitas em cada tratamento são apresentados na Tabela 1. O tipo de desbaste implementado foi o seletivo pelo alto, em que os tratamentos são definidos pelo número de árvores remanescentes e o número de árvores concorrentes. As árvores concorrentes foram retiradas a fim de beneficiar os indivíduos com maior potencial de crescimento, sendo selecionadas 400 árvores na primeira intervenção, aos 5 anos de idade e essas, então submetidas aos tratamentos: testemunha (Sem desbaste - 0), 1 (Desbaste moderado) ou 2 (Desbaste pesado).

Tabela 1 - Tratamentos de desbaste com o número de intervenções e as medições realizadas em todas as parcelas para cada tratamento.

Tratamentos	Densidade (N.ha <sup>-1</sup> )		Desbastes executados	Nº de medições*	Nº de observações
	Inicial	Final			
Sem desbaste	2500	2500	0	39	6128
Desbaste Moderado	2500	400	4	44	4232
Desbaste Pesado	2500	500	4	42	3541

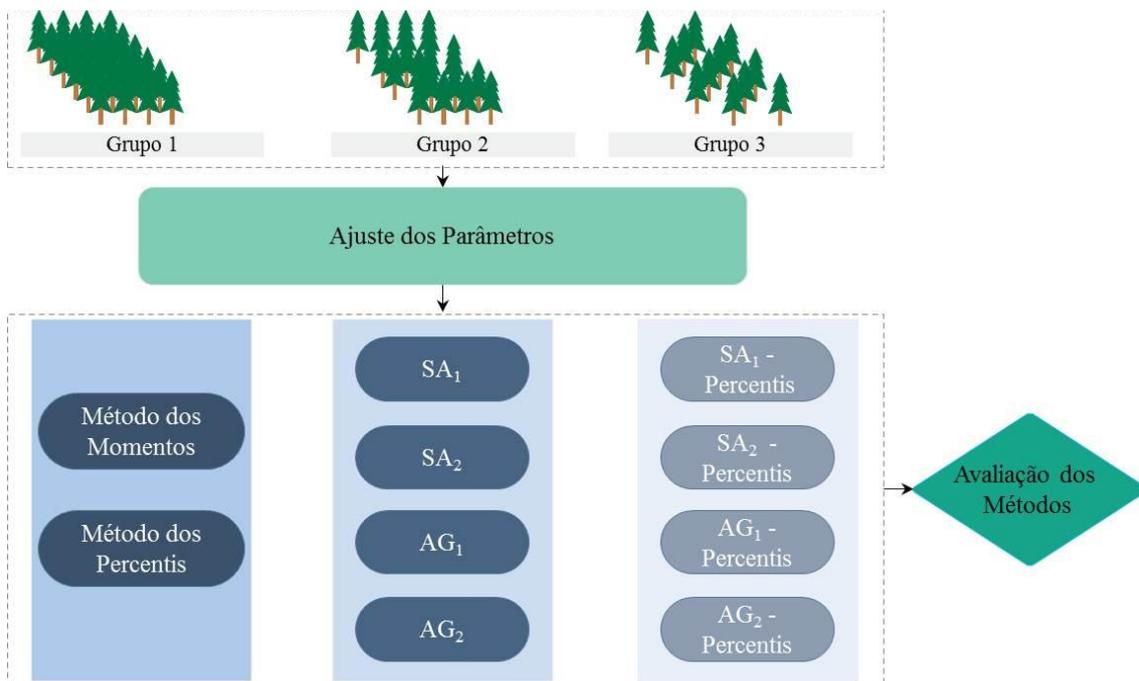
Em que: \*número de medições realizadas nas parcelas em cada tratamento.

Fonte: Dobner Jr. (2013).

## 2.2 Função de Distribuição Probabilística – Weibull

As funções de densidade probabilidade têm sido amplamente utilizadas, em destaque a distribuição Weibull, que apresenta relativa simplicidade e boa precisão para modelar a distribuição do diâmetro das árvores. Entretanto, a natureza não linear da distribuição diamétrica demanda uma simulação trabalhosa da distribuição Weibull (HUANG, 2000). Logo, diversos trabalhos foram direcionados para a obtenção dos parâmetros da mesma. Apesar do uso de métodos clássicos para a estimativa desses parâmetros, ainda são necessários esforços para o desenvolvimento de novas abordagens a fim de obter melhores estimativas. Nesse sentido, foram testados novos métodos envolvendo meta-heurísticas, em virtude do seu potencial de busca, para a estimativa dos parâmetros da distribuição Weibull 2 parâmetros (2P) e, assim, a estimativa da frequência das árvores em cada classe de diâmetro para todas as parcelas e idades estudadas. Para a realização dos ajustes, as informações de distribuição diamétrica observadas para cada parcela, em todas as idades, foram agrupadas em classes com amplitude de 1 cm, sendo as etapas descritas na Figura 1.

Figura 1 - Fluxograma simplificado das etapas metodológicas.



Em que: SA<sub>1</sub>: *Simulated Annealing* – Equação 7; SA<sub>2</sub>: *Simulated Annealing* – Equação 8; SA<sub>1</sub>-Percentis: *Simulated Annealing* – Equação 7 - Percentis; SA<sub>2</sub> - Percentis: *Simulated Annealing* – Equação 7 – Percentis; AG<sub>1</sub>: Algoritmo Genético – Equação 7; AG<sub>2</sub>: Algoritmo Genético – Equação 8; AG<sub>1</sub> - Percentis: Algoritmo Genético – Equação 7 – Percentis; AG<sub>2</sub> - Percentis: Algoritmo Genético – Equação 8 – Percentis.

Fonte: Da autora (2020).

A função de densidade probabilidade de Weibull 2P e sua respectiva distribuição acumulativa é apresentada pelas Equações 1 e 2, conforme Bailey e Dell (1973), em que  $x$ :

variável aleatória, no caso diâmetro central da classe ( $x \geq 0$ );  $b$  e  $c$ : parâmetros escala e forma da distribuição ( $b, c > 0$ ), respectivamente.

$$f(x) = \left(\frac{c}{b}\right) \left(\frac{x}{b}\right)^{c-1} \exp\left[-\left(\frac{x}{b}\right)^c\right] \quad (1)$$

$$f(x) = 1 - \exp\left[-\left(\frac{x}{b}\right)^c\right] \quad (2)$$

A função de Weibull 2P foi ajustada pelos métodos clássicos, Momentos e Percentis, os mais empregados na literatura. Além desses, novos métodos foram propostos envolvendo as meta-heurísticas Algoritmo Genético e o *Simulated Annealing*. Adicionalmente, para verificar a eficiência desses novos métodos foi utilizado um conjunto de dados que apresenta variações na estrutura florestal por meio de desbastes, uma vez que é fundamental que os novos métodos proporcionem estimativas precisas da densidade do povoamento por classe diamétrica tanto para sistemas com ou sem desbaste.

## 2.2.1 Métodos de ajuste da função

### 2.2.1.1 Método do Percentis

A estimativa dos parâmetros da função de Weibull por meio do método dos Percentis se relaciona com as Equações 3 e 4, em que  $a$ ,  $b$  e  $c$ : parâmetro locação, escala e forma, respectivamente;  $\ln()$ : logaritmo neperiano;  $p_1$  e  $p_2$ : percentis 1 e 2;  $Xp_1$  e  $Xp_2$ : diâmetros para os respectivos percentis 1 e 2. O parâmetro locação ( $a$ ) foi considerado como zero. Os valores dos percentis 1 e 2, empregados nas equações abaixo, foram 30 e 90 de acordo com Cunha Neto (1994).

$$c = \frac{\ln\left[\frac{-\ln(1-p_1)}{-\ln(1-p_2)}\right]}{\ln\left[\frac{(Xp_1-a)}{(Xp_2-a)}\right]} \quad (3)$$

$$b = \frac{Xp_1 - a}{(-\ln(1-p_1))^{1/c}} \quad (4)$$

### 2.2.1.2 Método dos Momentos

No método dos Momentos, foram utilizadas as informações do diâmetro médio e do desvio de cada parcela. Na estimativa dos parâmetros escala e forma aplicou-se um método iterativo, com o objetivo de reduzir a diferença entre o coeficiente de variação observado e o estimado por meio dos momentos amostrais. Os parâmetros foram estimados mediante as Equações 5 e 6, com auxílio do pacote EnvStats (MILLARD, 2018), em que  $b$ : parâmetro escala;  $\bar{D}$ : média aritmética dos diâmetros da parcela;  $\Gamma$ : função gamma;  $c$ : parâmetro forma;  $CV$ : coeficiente de variação;  $S_d$ : desvio padrão.

$$b = \frac{\bar{D}}{\Gamma\left(1 + \frac{1}{c}\right)} \quad (5)$$

$$CV = \frac{S_d}{\bar{D}} = \frac{\left[\Gamma\left(1 + \frac{2}{c}\right) - \Gamma^2\left(1 + \frac{1}{c}\right)\right]^{\frac{1}{2}}}{\Gamma\left(1 + \frac{1}{c}\right)} \quad (6)$$

### 2.2.1.3 Algoritmo Genético

A implementação do Algoritmo Genético (AG) foi realizada com o objetivo de encontrar o melhor conjunto de parâmetros da função de Weibull 2P. No AG foram definidos o tamanho da população, a função *fitness* e os operadores genéticos (seleção, cruzamento e mutação). O tamanho da população é representado por um grupo de indivíduos a serem avaliados em cada geração. Neste estudo, encontrou-se por meio de testes prévios, um tamanho populacional de 50 indivíduos e 100 gerações como critério de parada. O operador de seleção utilizado foi o torneio, no qual os indivíduos são comparados em busca dos mais aptos para a iteração subsequente. A taxa de mutação aplicada foi de 10%. Esse operador altera aleatoriamente um gene do cromossomo, de acordo com a probabilidade adotada. A avaliação dos indivíduos gerados foi realizada a partir de duas funções *fitness* utilizadas separadamente, ou seja, duas opções testadas de ajuste. A primeira (AG<sub>1</sub>) foi com base na soma residual, em módulo, das frequências de cada classe (Equação 7). A segunda (AG<sub>2</sub>) foi baseada na máxima diferença, em módulo, entre a probabilidade acumulativa observada e

estimada entre as classes de diâmetro (Equação 8) (APÊNDICE A e B), em que  $E$  : erro;  $fo_i$  : valor observado da frequência da classe  $i$ ;  $fe_i$  : valor estimado da frequência na classe  $i$ ;  $N$  : número total de classes;  $Máx$  : máximo;  $Po_i$  : valor da probabilidade acumulativa observada da classe  $i$ ;  $Pe_i$  : valor da probabilidade acumulativa estimada da classe  $i$ .

$$\text{Função Fitness 1: Minimizar } E = \sum_i^N |fo_i - fe_i| \quad (7)$$

$$\text{Função Fitness 2: Minimizar } E = Máx |Po_i - Pe_i| \quad (8)$$

Os valores preditos dos parâmetros de Weibull ( $b$  e  $c$ ) foram gerados de forma aleatória, porém dentro de um intervalo  $[0, +r]$ . No início, o valor  $r$  assume valores elevados (100), a fim de explorar um maior espaço de busca. Entretanto, no decorrer do processo de busca do algoritmo, esse intervalo reduz dinamicamente conforme valor da função fitness. Sendo assim, a atualização do limite superior ( $+r$ ) foi determinada como o dobro da média dos diâmetros da parcela, na próxima iteração do algoritmo. Devido à natureza estocástica do método, o processo foi realizado 50 vezes e então, retornou-se o indivíduo que alcançou o menor erro dentre as repetições.

#### 2.2.1.4 Simulated Annealing

O *Simulated Annealing* (SA) também foi implementado com objetivo de conseguir o melhor conjunto de parâmetros da função de Weibull 2P. No SA, previamente estabeleceu-se, por meio de testes, uma temperatura inicial ( $T_0$ ) igual a 50.000 e temperatura final ( $T_F$ ) igual a 25, sendo esse último o valor utilizado como critério de parada do algoritmo. A taxa de resfriamento utilizada foi igual a 1% e a função de desempenho empregada no SA foi idêntica à função *fitness* aplicada no AG. Assim, para avaliar as soluções foi utilizada a Equação 7 (SA<sub>1</sub>), como uma primeira forma de ajuste e a segunda opção empregando a Equação 8 (SA<sub>2</sub>) (APÊNDICE E e F). Os parâmetros ( $b$  e  $c$ ) foram gerados de forma aleatória dentro do intervalo de  $[0, +r]$ , sendo inicialmente o valor  $+r$  igual a 100, com a finalidade de investigar um maior espaço de busca para o SA. Posteriormente, os limites do intervalo foram modificados para acelerar o processo de busca do algoritmo, desta maneira, os limites do intervalo alteraram de forma dinâmica conforme a razão entre a temperatura corrente e a temperatura inicial. O SA apresenta dois mecanismos de refinamento para a escolha da

solução. O primeiro, mais robusto, seleciona somente a melhor solução como a corrente, já no segundo é empregado uma analogia da termodinâmica do recozimento de metais. Esse critério (metropolis) é um conceito probabilístico que permite aceitar soluções de pior qualidade (energia), que pode resultar em soluções ótimas com a progressão do processamento. O critério de metropolis também possibilita uma busca de forma sofisticada ao considerar o escape de ótimos locais. Todo esse processamento foi repetido 50 vezes a fim de reduzir o efeito aleatório da solução inicial e, então, a melhor solução é selecionada.

#### 2.2.1.5 Método percentis adaptado

Em alternativa a utilização de percentis fixos para todas as parcelas em cada idade, foi aplicada as duas meta-heurísticas descritas neste estudo para otimizar os valores dos percentis. Desta forma, a melhor combinação encontrada em cada meta-heurística foi utilizada para o ajuste da função de Weibull. A configuração das meta-heurísticas foi similar (APÊNDICE C, D, G e H), no entanto, os intervalos de busca aplicados para cada percentil foram de  $p1=\{1,\dots,50\}$  e  $p2=\{51,\dots,100\}$ , respectivamente para cada percentil fixo pelo método.

### 2.3 Critérios de avaliação dos ajustes

De modo a verificar a performance das meta-heurísticas frente aos métodos clássicos de ajuste, foram utilizadas estatísticas descritivas, tais como: o valor mínimo e máximo, desvio padrão e coeficiente de variação entre as repetições para cada técnica envolvendo a melhor solução/repetição. Além disso, uma análise gráfica do erro em cada iteração e o tempo de processamento das meta-heurísticas para cada teste aplicado (parâmetros e percentis) foram utilizados para fins comparativos. Na verificação da confiança do ajuste obtido por cada método, foi utilizado o teste de aderência não paramétrico Kolmogorov-Sminorv (KS) a 5% de significância (KOLMOGOROV, 1933; SMIRNOV, 1948). Esse teste compara a frequência acumulada estimada em relação à observada, sendo que o ponto de maior divergência entre as duas distribuições ( $D$ ) é a estatística do teste (Equação 9). Em que:  $F(x)$ : distribuição cumulativa teórica;  $S(x)$ : distribuição observada;  $D$ : diferença máxima entre  $F(x)$  e  $S(x)$ .

$$D = \text{MAX} |F(x) - S(x)| \quad (9)$$

O Índice de Erro (Equação 10) descrito por Reynolds et al. (1988) também foi utilizado para avaliar a compatibilidade entre a área basal ( $G$ ) observada e a estimada pelas técnicas empregadas neste estudo, em que  $EI_i$ : índice de erro;  $G_i$  e  $\hat{G}_i$ : são a área basal cumulativa observada e estimada na classe  $i$ . Dessa forma, foi calculada a média, o desvio padrão e o coeficiente de variação do índice de erro para o conjunto de dados.

$$EI_i = \sum |G_i - \hat{G}_i| \quad (10)$$

A fim de avaliar graficamente o erro obtido para todos os métodos ajustados foram elaborados os seguintes gráficos: histograma de frequência observada sob as curvas de frequência estimada por classe de diâmetro, um gráfico do erro acumulado de cada parcela para cada método e um gráfico de superfície de resposta para analisar as relações existentes entre os parâmetros e o erro obtido em uma repetição. Além desses gráficos, utilizou-se o gráfico de nível para melhor visualização das tendências dos percentis selecionados e o diagrama de Taylor para uma análise mais compreensível dos métodos de ajuste. Esse diagrama representa a proximidade das distribuições observadas e estimadas, sendo definida, então, por um ponto que relaciona o desvio padrão, a correlação e a raiz do erro quadrático médio (Equação 11) (TAYLOR, 2001), em que  $n$ : número de observações;  $e$ : erro;  $t$ : sub-índice das observações. Desta forma, o diagrama dá suporte a escolha do melhor método (ZHANG et al., 2019).

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \quad (11)$$

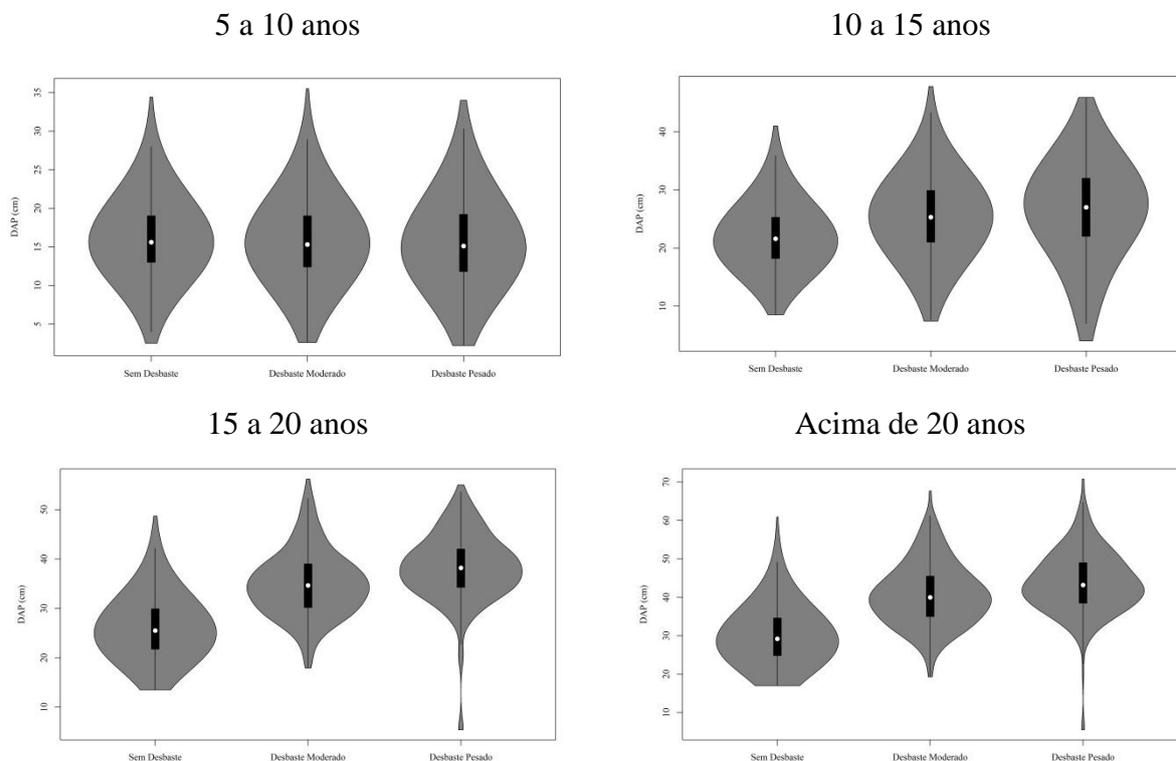
E por fim, com o objetivo de complementar os critérios de avaliação dos métodos de ajuste, realizou-se uma análise por meio de um *ranking* entre as distribuições. Esse *ranking* atribui pesos a cada um dos 10 métodos testados para cada parcela e idade. O método que apresentou menor valor de KS recebeu peso zero, o segundo menor valor recebeu peso um e assim sucessivamente até o método com maior valor de KS. Ao final, as pontuações de todas as parcelas independente da idade foram somadas e o método que apresentou o menor somatório foi indicado como o mais adequado ao conjunto de dados. O processamento dos dados e programação foi realizado no *software* R (*Version* 3.5.3 – © 2019 RStudio, Inc.), sendo utilizado com paralelismo para otimização do tempo por meio do pacote *doParallel*

(OOI et al., 2019). A metodologia foi processada numa CPU com processador Intel(R) Core™ i7-6700 CPU @ 3.40 GHz, com memória instalada (RAM) de 16,0 GB.

### 3. RESULTADOS

O comportamento dos dados mostra uma grande variação dos diâmetros independente do regime de desbaste adotado, sendo maior em parcelas sem a operação silvicultural, como esperado. Uma análise exploratória dos dados entre as medições realizadas (5 a 30 anos) comprova uma variação dos diâmetros entre 2,2 a 70,8 cm, com média aritmética de 21,3 cm e desvio padrão de  $\pm 8,99$  cm. A Figura 2 apresenta o comportamento dos dados sob diferentes idades e desbastes. Observa-se que as classes diamétricas médias apresentaram maior amplitude de frequência, contudo nas maiores idades nota-se maior variação com aplicação de desbaste pesado. Fato que permitiu explorar condições distintas aos métodos testados.

Figura 2- Gráfico de *violin plot* da variável diâmetro à altura do peito (cm) entre os diferentes tipos de desbaste. Em que: DAP: diâmetro à altura do peito em centímetros.



Fonte: Da autora (2020).

Do ponto de vista do processamento dos métodos observou-se uma variação na qualidade do ajuste, independentemente da idade e nível de desgaste. Na Tabela 2 está apresentado um resumo dos valores exploratórios obtidos pós-processamento. Assim, ao analisar essas estatísticas é constatado que no grupo das meta-heurísticas, o método SA apresenta altos valores de coeficiente de variação (CV), independente da estratégia de busca utilizada.

Tabela 2 - Estatísticas descritivas dos valores da função de avaliação obtidas entre as repetições e o tempo de processamento médio para cada técnica de ajuste.

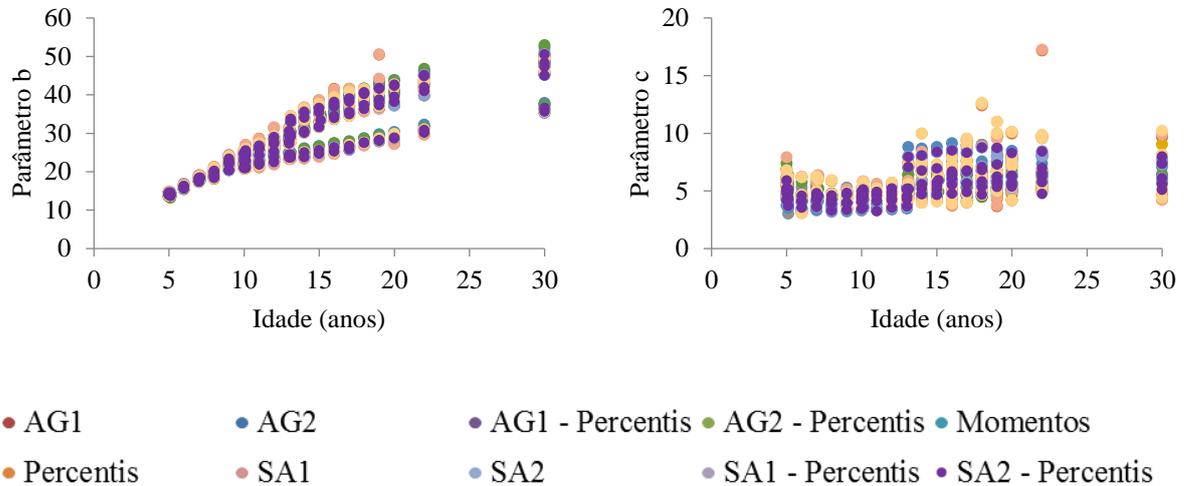
Natureza	Métodos	Função de avaliação				Tempo médio (segundos)
		Mínimo	Máximo	Média	CV (%)	
Uniobjetivo	SA <sub>1</sub>	16,492	236,000	51,045	46,041	2,57
	SA <sub>2</sub>	0,014	1,000	0,118	120,579	3,60
	AG <sub>1</sub>	16,494	219,148	44,233	22,736	145,90
	AG <sub>2</sub>	0,016	0,536	0,094	57,925	173,30
	SA <sub>1</sub> -Percentis	16,493	66,173	36,870	0,284	2,98
	SA <sub>2</sub> -Percentis	0,014	0,129	0,054	5,759	4,15
	AG <sub>1</sub> -Percentis	16,493	66,387	36,869	0,292	132,88
	AG <sub>2</sub> -Percentis	0,027	0,193	0,095	6,428	166,04

Em que: CV: coeficiente de variação em porcentagem.

Fonte: Da autora (2020).

Esse fato demonstra um comportamento aparentemente instável para as 50 repetições. Mesmo comportamento é observado para o algoritmo genético uni e multiobjetivo. Por outro lado, os métodos híbridos foram mais estáveis com baixos valores do coeficiente de variação e mesma tendência de valores da função entre repetições. Ao considerar o tempo de processamento na análise dos métodos, foi verificado que o SA demandou o menor tempo, como esperado. Esse resultado é atribuído ao tipo de exploração do espaço de soluções do SA que é local. A Figura 3 mostra os valores dos parâmetros escala e forma em diferentes idades. Verifica-se que para os dois parâmetros, no geral, os métodos estimaram valores bem próximos. Além disso, notou-se uma tendência do parâmetro escala apresentar valores superiores com o aumento da idade.

Figura 3 - Distribuição dos parâmetros obtidos por todos os métodos avaliados em diferentes idades.



Em que: SA<sub>1</sub>: *Simulated Annealing* – Equação 7; SA<sub>2</sub>: *Simulated Annealing* – Equação 8; SA<sub>1</sub>-Percentis: *Simulated Annealing* –Equação 7 – Percentis; SA<sub>2</sub> – Percentis: *Simulated Annealing* – Equação 7 – Percentis; AG<sub>1</sub>: Algoritmo Genético – Equação 7; AG<sub>2</sub>: Algoritmo Genético – Equação 8; AG<sub>1</sub> - Percentis: Algoritmo Genético – Equação 7 – Percentis; AG<sub>2</sub> - Percentis: Algoritmo Genético – Equação 8 – Percentis; b: parâmetro escala; c: parâmetro forma.

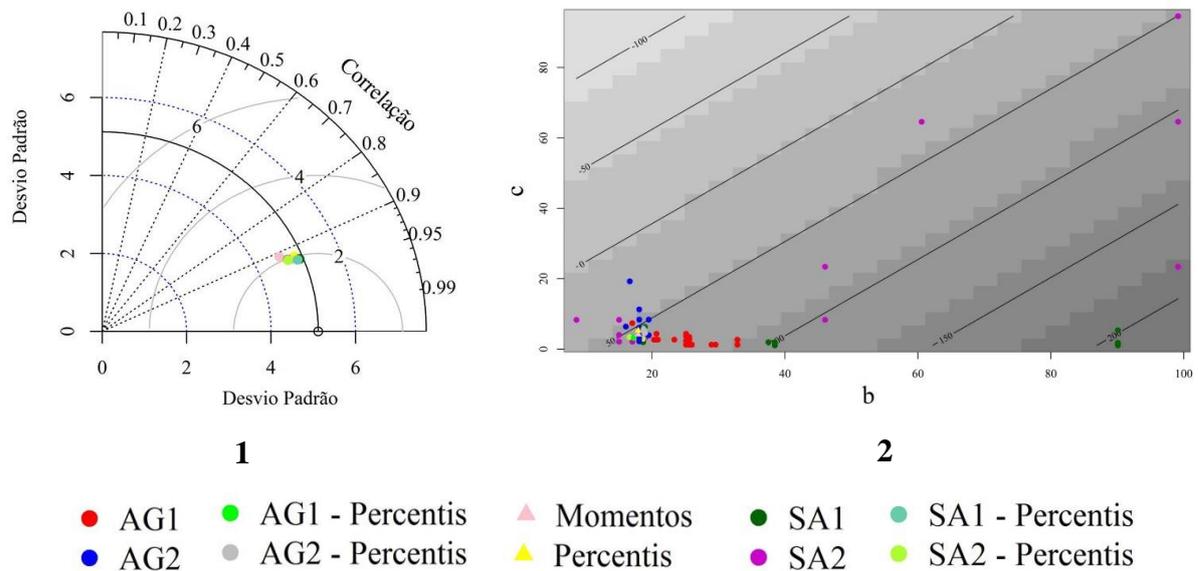
Fonte: Da autora (2020).

O Diagrama de Taylor (Figura 4 – 1) foi utilizado para permitir uma visualização da proximidade entre as distribuições observadas e estimadas. Nesse são apresentados todos os métodos de ajuste, inclusive os métodos clássicos, para fins comparativos. Dessa forma é possível verificar pequenas diferenças entre os métodos e uma boa proximidade com os dados observados. Além disso, observa-se que os métodos testados apresentaram correlação acima de 0,9 e valores de desvio padrão praticamente iguais ao da referência. A exceção foi o método dos Momentos que apresentou uma divergência em comparação com os demais métodos.

Adicionalmente, foi analisado o processo de melhoria da solução por meio do mapeamento do espaço amostral explorado (Figura 4 - 2). Na figura verifica-se que o grupo das meta-heurísticas converge após uma ampla pesquisa do espaço amostral, diferente dos métodos híbridos que ficaram mais concentrados em uma pequena região. No exemplo, foi adotada uma parcela com idade de 7 anos e 2 desbastes. Esse comportamento pode ser corroborado pelo CV médio dos métodos, apresentado na Tabela 2. Além disso, os métodos propostos apresentaram semelhanças e melhorias na estimativa dos parâmetros da função em relação aos métodos clássicos, devido aos mecanismos de alto nível para explorar o espaço de

busca. No entanto, os métodos propostos apresentam desvantagens ao consumir maior tempo para a sua convergência.

Figura 4 – Representação do diagrama de Taylor (1) envolvendo todos os ajustes e o comportamento do espaço amostral dos parâmetros b e c para uma parcela submetida a 2 desbastes gerados pelo gráfico de superfície de resposta (2).



Em que: SA<sub>1</sub>: *Simulated Annealing* – Equação 7; SA<sub>2</sub>: *Simulated Annealing* – Equação 8; SA<sub>1</sub>-Percentis: *Simulated Annealing* –Equação 7 - Percentis; SA<sub>2</sub> – Percentis: *Simulated Annealing* – Equação 7 – Percentis; AG<sub>1</sub>: Algoritmo Genético – Equação 7; AG<sub>2</sub>: Algoritmo Genético – Equação 8; AG<sub>1</sub> - Percentis: Algoritmo Genético – Equação 7 – Percentis; AG<sub>2</sub> - Percentis: Algoritmo Genético – Equação 8 – Percentis.

Fonte: Da autora (2020).

Após os ajustes das funções pelos métodos, observou-se uma superioridade dos métodos SA<sub>2</sub>, AG<sub>2</sub> e SA<sub>2</sub> – Percentis, obtendo aderência em todas as parcelas e tratamentos de desbaste (Tabela 3). Além disso, esses métodos também apresentaram as primeiras colocações no *ranking*. Fato não constatado pelos métodos clássicos Percentis e Momentos. Com isso nota-se que a meta-heurística SA, tanto em sua forma pura como em conjunto com os Percentis, demonstra maior capacidade em solucionar este tipo de problema. Em relação aos outros métodos, houve ocasiões cujos resultados não evidenciaram semelhança entre as distribuições estimada e observada, principalmente para as parcelas com aplicação de desbaste pesado. De acordo com os resultados, parcelas não submetidas a desbaste apresentaram os melhores ajustes na maioria dos métodos testados, com destaque para os novos métodos propostos. Em seguida, estão os dados com aplicação de desbaste moderado e

desbaste pesado. Esse comportamento pode ser atribuído à influência da dinâmica da densidade populacional por meio do manejo desses povoamentos, em que é conferida uma maior complexidade ao ajuste da distribuição diamétrica.

Tabela 3 – Síntese dos resultados obtidos pela avaliação estatística dos métodos de ajuste da distribuição Weibull nos diferentes tipos de regime de desbaste.

Métodos	Desbaste	Análise estatística				Ranking
		Kolomogorv – Smirnov <sup>a</sup>		Índice de Erro Área Basal		
		Aderência (%)	Média (%)	Média (m <sup>2</sup> /ha)	CV (%)	
Métodos Clássicos						
Momentos	0	97,44		0,86	25,49	(311)
	1	86,36	91,43	0,72	25,00	10 (281)
	2	90,48		0,72	22,86	(293)
Percentis	0	84,62		0,82	25,84	(270)
	1	100,00	94,08	0,69	24,39	7 (265)
	2	97,62		0,69	25,35	(247)
Meta-heurísticas						
SA <sub>1</sub>	0	100,00		0,76	24,43	(241)
	1	86,36	85,93	0,63	27,40	8 (309)
	2	71,43		0,63	26,82	(307)
SA <sub>2</sub>	0	100,00		0,81	24,67	(39)
	1	100,00	100,00	0,70	24,54	1 (44)
	2	100,00		0,68	24,25	(43)
GA <sub>1</sub>	0	100,00		0,76	24,29	(244)
	1	88,64	86,69	0,63	27,37	9 (319)
	2	71,43		0,63	26,73	(311)
GA <sub>2</sub>	0	100,00		0,81	24,62	(102)
	1	100,00	100,00	0,69	24,86	3 (119)
	2	100,00		0,68	24,69	(112)
Métodos Híbridos (meta-heurística + percentil)						
SA <sub>1</sub> -Percentis	0	100,00		0,77	25,17	(199)
	1	90,91	93,80	0,63	27,25	4 (270)
	2	90,48		0,64	26,90	(263)
SA <sub>2</sub> -Percentis	0	100,00		0,81	24,43	(96)
	1	100,00	100,00	0,69	24,54	2 (104)
	2	100,00		0,68	24,33	(97)
GA <sub>1</sub> -Percentis	0	100,00		0,77	25,17	(199)
	1	90,91	93,00	0,63	27,19	5 (271)
	2	88,10		0,64	26,90	(264)
GA <sub>2</sub> -Percentis	0	89,74		0,85	23,88	(253)
	1	100,00	93,41	0,69	25,66	6 (274)
	2	90,48		0,68	26,21	(222)

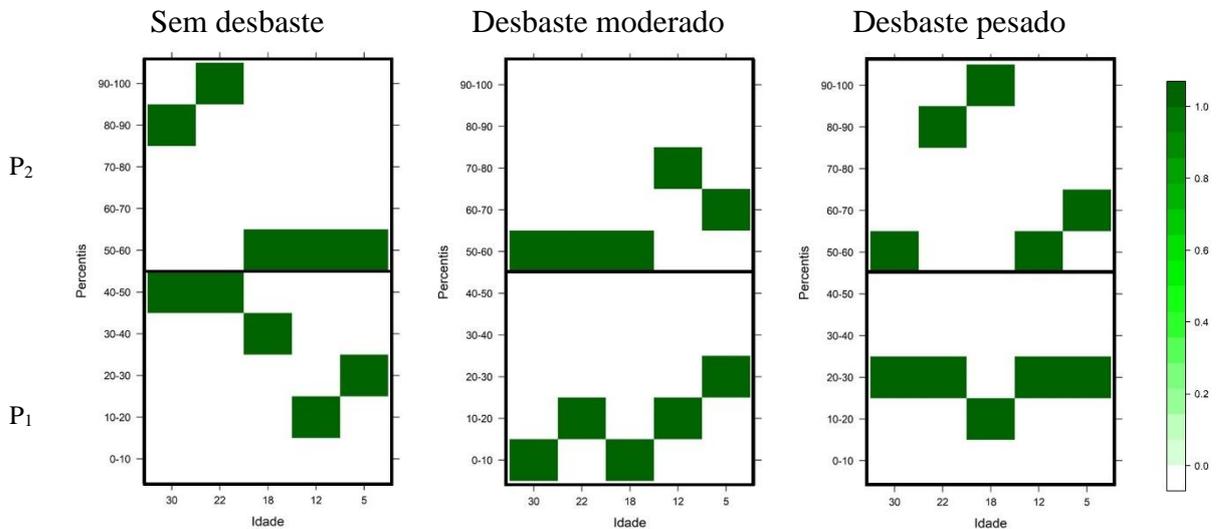
Em que: Níveis de desbaste: 0 - sem desbaste; 1 - desbaste moderado e 2 - desbaste pesado; a: teste de Kolmogorov – Smirnov a 5% de significância; \*: número de parcelas que apresentaram aderência ao teste de KS; DP: desvio padrão; CV: coeficiente de variação; () somatório correspondente a determinação do ranking.

Fonte: Da autora (2020).

Entre os métodos clássicos, os Percentis (30 e 90) recebe destaque, uma vez que obteve total aderência ao teste KS nas parcelas com aplicação de desbaste moderado e uma média geral entre tratamentos superior aos Momentos. Esse resultado é reforçado pelo Índice de Erro, em que os Percentis apresentaram os menores valores. Além disso, ao comparar o método dos Percentis com os demais métodos verificou-se por meio do ranking e da porcentagem de aderência que os Percentis se mostrou superior aos métodos SA<sub>1</sub> e AG<sub>1</sub>.

Os métodos híbridos obtiveram resultados consistentes de ajuste, em que, no geral, mais de 90% das parcelas apresentaram valores não significativos ao teste KS. De acordo com esse teste, o método SA<sub>2</sub> – Percentis se destacou entre os demais, no entanto, os métodos SA<sub>1</sub> – Percentis e o AG<sub>1</sub> – Percentis apresentaram estimativas mais precisas em área basal, conforme o Índice de Erro. De forma geral, ao observar os valores dos percentis encontrados pelos métodos híbridos para cada tratamento de desbaste e idade, é possível constatar algumas tendências para os percentis 1 e 2 (Figura 5). O percentil 1 demonstrou uma redução, variando de 50 a 20, com a aplicação de desbaste. Já o percentil 2 apresentou-se aproximadamente constante com o valor de 60. Ao analisar por idade foi verificado um aumento dos valores percentuais nas maiores idades. Dessa forma, essa análise explica os bons resultados encontrados pelo método dos Percentis com os valores fixos de 30 e 90. Além disso, é possível perceber que os diferentes valores de percentis variam com a dinâmica do povoamento e/ou aplicações de operações silviculturais.

Figura 5 - Os melhores percentis 1 e 2 selecionados pelo método híbrido SA<sub>2</sub> - Percentis para as idades de 5, 12, 18, 22 e 30 anos.

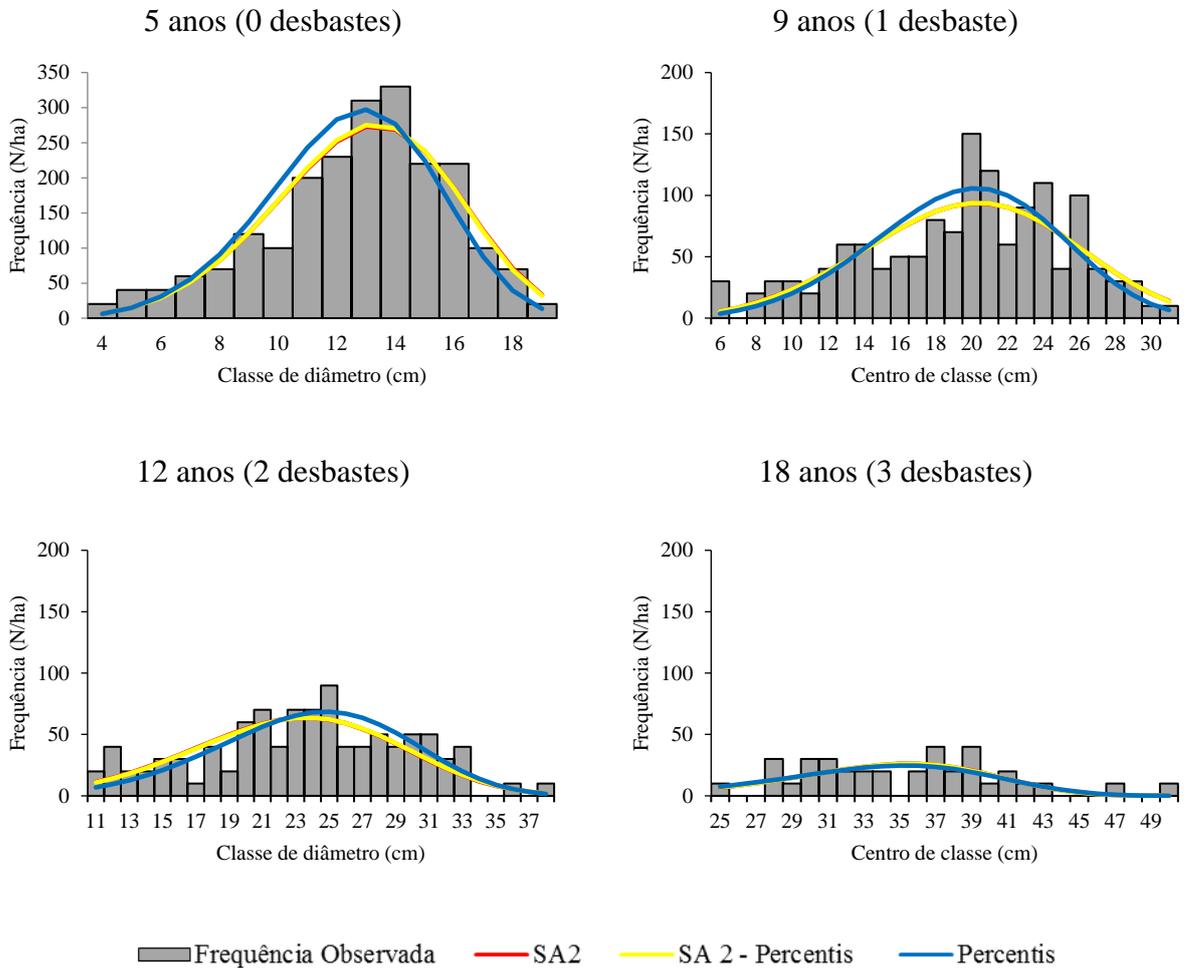


Em que: P<sub>1</sub> e P<sub>2</sub>: são percentil 1 e 2, respectivamente.

Fonte: Da autora (2020).

Em relação as meta-heurísticas, todos os ajustes realizados com os métodos SA<sub>2</sub> e o AG<sub>2</sub> resultaram em total aderência da função aos dados, pelo teste KS para todos os tratamentos de desbaste. Entretanto, estes métodos apresentaram valores maiores de Índice de Erro quando comparados com os métodos SA<sub>1</sub> e AG<sub>1</sub>. Dessa forma, é observada uma relação entre a função de avaliação 2 (Equação 8) e o desempenho do método utilizado, visto que a mesma busca minimizar a distância e aproximar a forma da curva de distribuição. Já a função de avaliação 1 (Equação 7) tem o objetivo de compatibilizar as frequências observada e estimada por classe. Dessa maneira, ao associar a frequência com a área basal da classe, conseqüentemente o erro em área basal é minimizado. Na Figura 6 é observado o comportamento das estimativas dos métodos com os melhores resultados por grupo (Percentis, SA<sub>2</sub> e SA<sub>2</sub> – Percentis) para uma parcela submetida a desbastes. Verifica-se que as maiores diferenças encontradas entre o método clássico e os propostos foi com a não aplicação de desbaste ou apenas um, principalmente nas classes médias e superiores.

Figura 6 - Histogramas de frequências observadas versus a distribuição Weibull ajustada pelo melhor método por grupo, em uma parcela submetida a desbaste moderado em diferentes idades.

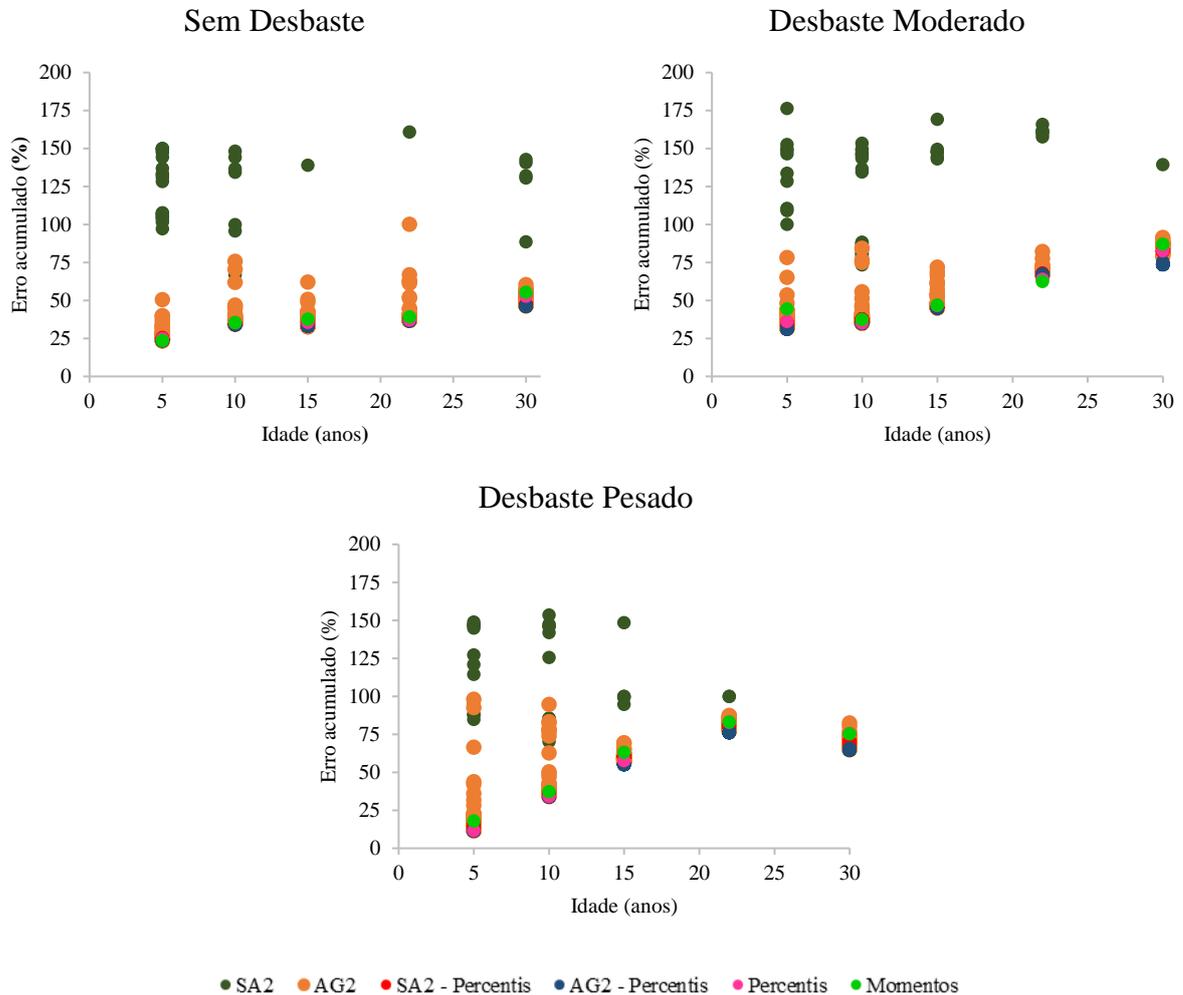


Em que: SA<sub>2</sub>: *Simulated Annealing* – Equação 8; SA<sub>2</sub> – Percentis: *Simulated Annealing* – Equação 7 – Percentis.

Fonte: Da autora (2020).

De forma complementar a essa análise, foram elaborados gráficos com o erro acumulado da parcela em diferentes idades e tratamentos, para todos os métodos testados (Figura 7). Assim, é verificado que os métodos propostos mostraram superioridade por apresentarem os menores erros em praticamente todas as idades e tratamentos. À exceção foi na idade de 22 anos com desbaste moderado, em que os métodos clássicos, tanto Momentos como Percentis, apresentaram erros menores. Dessa forma, também é observado que a diferença entre os erros dos métodos propostos é pequena.

Figura 7 - Erros das classes acumulados (%) de cada repetição para os métodos com bom desempenho, para os diferentes tipos de tratamentos nas idades de 5, 10, 15, 22 e 30 anos.



Em que: SA<sub>1</sub>: *Simulated Annealing* – Equação 7; SA<sub>2</sub>: *Simulated Annealing* – Equação 8; SA<sub>1</sub>-Percentis: *Simulated Annealing* –Equação 7 – Percentis; SA<sub>2</sub> – Percentis: *Simulated Annealing* – Equação 7 – Percentis; AG<sub>1</sub>: Algoritmo Genético – Equação 7; AG<sub>2</sub>: Algoritmo Genético – Equação 8; AG<sub>1</sub> - Percentis: Algoritmo Genético – Equação 7 – Percentis; AG<sub>2</sub> - Percentis: Algoritmo Genético – Equação 8 – Percentis.

Fonte: Da autora (2020).

#### 4. DISCUSSÃO

A função Weibull 2 parâmetros foi utilizada para a distribuição diamétrica em povoamentos de *Pinus taeda* submetidos a regimes de desbaste. Para tal, diferentes métodos de obtenção dos parâmetros e duas funções de avaliação foram testados. Os resultados do estudo constataram que os novos métodos propostos são uma excelente alternativa aos métodos clássicos na estimação dos parâmetros. Isso indica que qualquer um dos métodos

pode ser utilizado, enfatizando que uma maior consistência na qualidade de ajuste foi proporcionada pelos métodos SA<sub>2</sub>, AG<sub>2</sub> e SA<sub>2</sub> – Percentis (Tabela 3). Estudos anteriores denotaram a aplicabilidade de meta-heurísticas para a estimativa dos parâmetros da função Weibull, uma vez que também encontraram bons resultados. Dentre esses, estão Chang-Jun, Xin e Xin (2017) que utilizaram o AG junto com a função de máxima verossimilhança e Abbasi et al. (2006) que aplicaram o SA para maximizar a função de máxima verossimilhança. Esses autores afirmam ainda que outras meta-heurísticas podem ser usadas. Uma outra abordagem eficaz na qualidade dos ajustes foi o uso da função de avaliação 2 (Equação 8), que demonstrou superioridade frente as demais. Esse fato também foi constatado por Nosal e Nosal (2003), que encontraram um bom desempenho com a função de minimizar a distância de Kolmogorov – Smirnov para o ajuste da distribuição Weibull.

Além da estatística clássica de Kolmogorov – Smirnov, também foi utilizado o Índice de Erro (EI) em área basal (Reynolds et al., 1988). De acordo com o EI, o método dos Momentos foi o que apresentou os maiores valores de erro em área basal em comparação com os demais métodos. Esse resultado corrobora com o trabalho de Liu et al. (2009), que encontraram que o método dos Percentis produziu estatísticas de EI e KS menores que o método dos Momentos e outros métodos avaliados. No entanto, contraria o trabalho de Lei (2008) para a mesma função, já que nesse caso o método dos Momentos produziu a melhor estimativa considerando o *Pinus tabulaeformis*. O resultado também diverge do estudo de Fonton et al. (2013), que observaram para o método dos Momentos menores valores de EI e KS em diferentes regimes silviculturais (talhadia e alto fuste).

Em relação a abordagem do método dos Percentis, Adeyemi e Adesoye (2016) relatam que o desempenho obtido por esse método é influenciado pela escolha dos percentis a serem utilizados. Também, é apontada uma dificuldade de se compararem diferentes combinações de percentis para cada conjunto de dados. Esse problema pode ser solucionado com o método híbrido proposto neste estudo, em que as meta-heurísticas avaliam diferentes valores de percentis para cada parcela. Os resultados encontrados por esse método mostraram uma ligeira melhora nas estatísticas avaliadas, principalmente com o uso do SA (Tabela 3). A eficiência dessa meta-heurística se torna desejável, uma vez que apresenta boa capacidade de exploração na procura da solução e boa adaptação para vários tipos de problemas (ASSAD; DEEP, 2018; COURCHELLE et al., 2019). Com isso, a utilização desses algoritmos simplifica o processo de obtenção dos parâmetros, sendo possível testar várias combinações tanto para percentis, quanto para os próprios valores dos parâmetros.

Apesar dos métodos avaliados produzirem estimativas precisas para todas ou a maioria das parcelas, em outros casos, os resultados previstos eram significativamente diferentes dos valores observados. Tal situação pode ser ocasionada pelo desbaste, mas também pelo avanço da idade do povoamento e outros fatores naturais, como condições heterogêneas no local e densidade irregular (POGODA, OCHAL, ORZEL, 2019). Esses fatores afetam a distribuição do diâmetro que provocam uma redução do pico da distribuição tornando-o pouco distorcido, porém com maior variação (NORD-LARSEN; CAO, 2006). Além disso, uma análise visual mostrou que em alguns casos as distribuições observadas tendiam a ser bimodais (Figura 6). No entanto, a distribuição Weibull é unimodal e, dessa forma, é inadequada para modelar essas distribuições, sendo outras formas mais indicadas. Contudo, os métodos testados são confiáveis e aplicáveis.

## 5. CONCLUSÃO

Os resultados apresentados pelos métodos propostos são adequados, demonstrando um potencial de aplicação das meta-heurísticas na estimativa dos parâmetros da distribuição Weibull 2 parâmetros, em povoamentos de *Pinus taeda* destacando os métodos SA<sub>2</sub>, AG<sub>2</sub> e SA<sub>2</sub> – Percentis. Os métodos híbridos destacaram a importância na identificação dos percentis ótimos da função, tendo uma tendência com a aplicação de desbaste e o avanço da idade. A eficiência das meta-heurísticas utilizadas foram influenciadas pela função de avaliação, em que a função 2 (Equação 8) mostrou superioridade quando analisado o valor crítico de Kolmogorov – Smirnov, por outro lado a função 1 (Equação 7) apresentou resultados mais precisos pelo teste do Índice de Erro.

## REFERÊNCIAS

ABBASI, B. et al. Estimating the parameters of Weibull distribution using simulated annealing algorithm. **Applied Mathematics And Computation**, New York, v. 183, n. 1, p.85-93, 2006.

ACITAS, S.; ALADAG, C. H.; SENOGLU, B.. A new approach for estimating the parameters of Weibull distribution via particle swarm optimization: An application to the strengths of glass fibre data. **Reliability Engineering & System Safety**, Barking, v. 183, p.116-127, 2019.

ADEYEMI, A. A.; ADESOYE, P. O.. Evaluation of Some Weibull Parameter Estimation Methods for Characterizing Stem Diameter Distribution in a Tropical Mixed Forest of

Southern Nigeria. **Journal Of Statistical Science And Application**, Germany, v. 4, n. 6, p.257-275, 2016.

ALVARES, C. A. et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, Stuttgart, v.22, n.6, p.711-728, 2013.

ARAÚJO JÚNIOR, C. A. et al. Modelagem da distribuição diamétrica de povoamentos de eucalipto utilizando a função Gama. **Cerne**, Lavras, v. 19, n. 2, p.307-314, 2013.

ARIAS-RODIL, M. et al. Modeling diameter distributions in radiata pine plantations in Spain with existing countrywide LiDAR data. **Annals Of Forest Science**, France, v. 75, n. 2, p.1-12, 2018.

ASSAD, A.; DEEP, K. A Hybrid Harmony search and Simulated Annealing algorithm for continuous optimization. **Information Sciences**, New York, v. 450, p.246-266, 2018.

BAILEY, R.; DELL, T. Quantifying diameter distributions with the Weibull function. **Forest Science**. Bethesda, v.19, n.2. p.97-104, 1973.

BARRA, O. S. V. et al. Proposta metodológica para o ajuste ótimo da distribuição diamétrica Weibull 3P. **Floresta**, Curitiba, v. 34, n. 3, p.387-393, 2004.

BINOTI, D.H.B; BINOTI, M.L.M.S.; LEITE, H.G.; MELIDO, R.C.N.; SANTOS, F.L. Descrição da distribuição diamétrica de sistemas agrossilvipastoris utilizando as funções Weibull e hiperbólica. **Revista Árvore**, Viçosa, v.36, n.2, p.349-356, 2012.

CHANG-JUN, W.; XIN, L.; XIN, C. Parameter Evaluation of 3-parameter Weibull Distribution based on Adaptive Genetic Algorithm. **Advances in Engineering Research**, France, v. 138, p. 246-431, 2017.

COURCHELLE, V. et al. A simulated annealing approach to 3D strategic aircraft deconfliction based on en-route speed changes under wind and temperature uncertainties. **Transportation Research Part C: Emerging Technologies**, Amsterdam, v. 103, p.194-210, 2019.

CUNHA NETO, F. R. da et al. Modelo para predição da produção por classe de diâmetro para *Eucalyptus grandis*. **Cerne**, Lavras, MG, v.1, n.1, p.108-122, jun. 1994.

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

DOBNER Jr, M. **Crown thinning effects on growth and wood quality of Pinus taeda stands in southern Brazil**. 2013. 221 p. Tese (Doutor em Ciências Naturais) – Universidade Albert-Ludwigs Freiburg im Breisgau, Alemanha, 2013.

FERRAZ FILHO, A. C. et al. Modeling growth and yield of loblolly pinestands under intensive management. **Pesquisa Agropecuária Brasileira**, Brasília, v. 50, n. 8, p.707-717, 2015.

FONTON, N. H. et al. Modeling the Distribution of Marketable Timber Products of Private Teak (*Tectona grandis* L.f.) Plantations. **Open Journal Of Forestry**, Irvine, v. 03, n. 04, p.115-121, 2013.

OOI, H. et al. doParallel: Foreach Parallel Adaptor for the 'parallel' Package. R package version, v. 1, n. 15, 2019.

Huang J. R. Studies on the Suitability of Weibull Distribution in Masson Pine Plantations. **Guizhou Forestry Science and Technology**, Guiyang, v.28, n.1, p.7-13, 2000.

KOLMOGOROV, A. Sulla determinazione empirica di una legge di distribuzione. **Giornale dell'Istituto Italiano degli Attuari**, Roma, v. 4, p. 83–91, 1933.

LEI, Y. Evaluation of three methods for estimating the Weibull distribution parameters of Chinese pine (*Pinus tabulaeformis*). **Journal Of Forest Science**, Czech Republic, v. 54, n. 12, p.566-571, 2008.

LIU, C. et al. Applications and comparison of six methods for predicting parameters of the Weibull function in unthinned *Picea glauca* plantations. **Scandinavian Journal Of Forest Research**, Basingstoke, v. 24, n. 1, p.67-75, 2009.

LEITE, H. G. et al. Avaliação do ajuste das funções Weibull e hiperbólica a dados de povoamentos de eucalipto submetidos a desbaste. **Revista Árvore**, Viçosa, v. 34, n. 2, p.305-311, 2010.

LOETSCH, F.; ZÖHRER, F.; HALLER, K. E. **Forest inventory**. 2.ed. Munich: BLV Verlagsgesellschaft, 1973. 469 p.

MIGUEL, E. P. et al. Using the Weibull function for prognosis of yield by diameter class in *Eucalyptus urophylla* stands. **Cerne**, Lavras, v. 16, n. 1, p.94-104, 2010.

MILLARD, S. P. 2018. **EnvStats**. Disponível em: <https://cran.r-project.org/web/packages/EnvStats/index.html>. Acesso em 20 de jun. 2019.

MIRZAEI, M. et al. Modeling frequency distributions of tree height, diameter and crown area by six probability functions for open forests of *Quercus persica* in Iran. **Journal Of Forestry Research**, Berlin, v. 27, n. 4, p.901-906, 2015.

NASCIMENTO, R. G. M. et al. Modelo de projeção por classe diamétrica para florestas nativas: enfoque na função probabilística de Weibull. **Pesquisa Florestal Brasileira**, Colombo, v. 32, n. 70, p.93-103, 2012.

NASCIMENTO, R. G. M. et al. A growth and yield projection system for a tropical rainforest in the Central Amazon, Brazil. **Forest Ecology And Management**, Amsterdam, v. 327, p.201-208, 2014.

NORD-LARSEN, T.; CAO, Q. V. A diameter distribution model for even-aged beech in Denmark. **Forest Ecology And Management**, Amsterdam, v. 231, n. 1-3, p.218-225, 2006.

Nosal, M.; Nosal, E.M. Three-parameter Weibull generator for replacing missing observations, in: Proceedings of the Hawaii International Conference on Statistics and Related Fields, Honolulu, Hawaii, p. 1-9, 2003.

OGANA, F. N.; GORGOSO-VARELA, J.J. Comparison of estimation methods for fitting Weibull distribution to the natural stand of Oluwa forest reserve, Ondo State, Nigeria. **Journal of Research in Forestry, Wildlife and Environment**, Makurdi, v. 7, n. 2, p. 81-90, 2015.

OGANA, F. N.; ITAM, E. S.; OSHO, J. S. A.. Modeling diameter distributions of *Gmelina arborea* plantation in Omo Forest Reserve, Nigeria with Johnson's SB. **Journal Of Sustainable Forestry**, Binghamton, v. 36, n. 2, p.121-133, 23 nov. 2016.

PALAHÍ, M.; PUKKALA, T.; TRASOBARES, A. Calibrating Predicted Tree Diameter Distributions in Catalonia, Spain. **Silva Fennica**, Helsinki, n. 40, v. 3, p. 487-500, 2006.

PALAHÍ, M.; PUKKALA, T.; BLASCO, E.;TRASOBARES, A. Comparison of beta, Johnson's SB, Weibull and truncated Weibull functions for modeling the diameter distribution of forest stands in Catalonia (north-east of Spain). **European Journal of Forest Research**, Berlin, n. 126, p. 563-571, 2007.

POGODA, P.; OCHAŁ, W.; ORZEŁ, S. Modeling Diameter Distribution of Black Alder (*Alnus glutinosa* (L.) Gaertn.) Stands in Poland. **Forests**, Switzerland, v. 10, n. 5, p.1-16, 2019.

POUDEL, K. P.; CAO, Q. V. Evaluation of Methods to Predict Weibull Parameters for Characterizing Diameter Distributions. **Forest Science**, Switzerland, v. 59, n. 2, p.243-252, 2013.

R Core Team (2019). **R**: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

SMIRNOV, N. Table for estimating the goodness of fit of empirical distributions. **Annals of Mathematical Statistics**, Ann Arbor, v. 19, p. 279–281, 1948.

TAYLOR, K. E. Summarizing multiple aspects of model performance in a single diagram. **Journal of Geophysical Research**, United States, v.106, n.7, p.7183-7192, 2001.

TÉO, S. J. et al. Desempenho de funções de densidade probabilísticas para descrever a distribuição diamétrica de pinus taeda, na região de Caçador, SC. **Floresta**, Curitiba, v. 42, n. 4, p.741-754, 2012.

ZHANG, J. et al. Modelling uniaxial compressive strength of lightweight self-compacting concrete using random forest regression. **Construction And Building Materials**, New York, v. 210, p.713-719, 2019.

**ARTIGO 2 – ESTOQUE DE CARBONO ACIMA DO SOLO: UM ESTUDO  
SOBRE A OTIMIZAÇÃO NA SELEÇÃO DE VARIÁVEIS EXPLICATIVAS E  
MODELAGEM**

**Artigo formatado conforme a NBR 6022 (ABNT, 2003) e adaptado as exigências do  
Manual de Normalização de Trabalhos Acadêmicos da UFLA.**

## RESUMO

Os ecossistemas florestais tropicais são importantes no processo de armazenamento de carbono. Diante disso, torna-se necessário desenvolver relações alométricas mais precisas. Assim, o objetivo do estudo foi comparar diferentes técnicas de modelagem na predição do estoque de carbono acima do solo, com o propósito de obter estimativas mais precisas e selecionar variáveis que expliquem o comportamento do acúmulo de carbono nas árvores. Para isso, estudos sobre o estoque de carbono em florestas ineqüíneas foram direcionados para a Bacia do Rio Grande – Minas Gerais, onde foram inventariados 29 fragmentos de forma representativa da área e mensuradas 227 árvores das mais diferentes espécies e fisionomias existentes dentro da Bacia. Os modelos de análise de regressão, uma equação pan-tropical, o RF e o método híbrido proposto - *Simulated Annealing* com RF (SARF) foram utilizados para estimar o estoque de carbono com base na seleção de variáveis para os compartimentos folha, galho, fuste e total. A abordagem do modelo SARF consistiu no uso da metaheurística SA para a busca de um conjunto de variáveis a serem utilizadas como entradas do *Random Forest*. Esses métodos foram avaliados pela raiz do erro quadrático médio absoluto e percentual (RMSE) e coeficiente de determinação ( $R^2$ ), além de análise gráfica. A partir dos resultados obtidos, verificaram ajustes superiores da equação pan-tropical em relação aos modelos adaptados, devido a maior homogeneidade dos resíduos. A técnica *Stepwise* permitiu reduzir o número de variáveis, com predominância da variável do peso seco, e proporcionou um bom desempenho das estimativas para os diferentes compartimentos com valores de RMSE (%) variando de 4 a 40% para o treino e até 76% para a validação. O SARF apresentou métricas de avaliação superiores ao RF puro. Além disso, o SA mostrou alta capacidade de redução de variáveis e do erro das estimativas com reduções percentuais de RMSE% de até 29% para o treino e 42% para validação em relação ao RF simples. Dentre as variáveis selecionadas pelo modelo SARF, verifica-se maior contribuição do peso seco, volume, densidade básica, área de copa, área basal, diâmetro a 0%, diâmetro à altura do peito, altura e latitude. A partir desses resultados, é observada uma associação do estoque de carbono da árvore e fatores ecológicos e ambientais. Os métodos, *Stepwise* e SARF, proporcionaram bons ajustes, no entanto, o SARF seria uma alternativa ao método tradicional quando envolve grande número de variáveis e observações.

**Palavras-chave:** Mineração de dados. Biomassa. *Simulated Annealing*. *Random Forest*.

## ABSTRACT

Tropical forest ecosystems are important in the carbon storage process. Therefore, it becomes necessary to develop more precise allometric relationships. So, the objective of the study was to compare different modeling techniques in the prediction of the carbon stock above ground, in order to obtain more accurate estimates and select variables that explain the behavior of carbon accumulation in trees. For this, studies on the carbon stock in unequal forests were directed to the Rio Grande Basin - Minas Gerais, where 29 fragments representing the area were inventoried and 227 trees of the most different species and physiognomies existing within the Basin were measured. The regression analysis models, a pan-tropical equation, the RF and the proposed hybrid method - Simulated Annealing with RF (SARF) were used to estimate the carbon stock based on the selection of variables for the leaf, twig compartments, stem and total. The approach of the SARF model consisted of using the metaheuristics SA to search for a set of variables to be used as inputs of RF. These methods were evaluated by the root of the mean absolute and percentage square error (RMSE) and coefficient of determination ( $R^2$ ), in addition to graphical analysis. From the results obtained, superior adjustments of the pan-tropical equation were verified in relation to the adapted models, due to the greater homogeneity of the residues. The Stepwise technique allowed to reduce the number of variables, with a predominance of the dry weight variable, and provided a good performance of the estimates for the different compartments with RMSE values (%) ranging from 4 to 40% for training and up to 76% for validation. SARF presented evaluation metrics superior to pure RF. In addition, SA showed a high capacity for reducing variables and error in estimates with percentage reductions of RMSE% of up to 29% for training and 42% for validation in relation to simple RF. Among the variables selected by the SARF model, there is a greater contribution of dry weight, volume, basic density, crown area, basal area, diameter at 0%, diameter at breast height, height and latitude. From these results, an association of the tree's carbon stock and ecological and environmental factors is observed. The methods, Stepwise and SARF, provided good adjustments, however, SARF would be an alternative to the traditional method when it involves a large number of variables and observations.

**Keywords:** Data mining. Biomass. *Simulated Annealing*. *Random Forest*.

## 1. INTRODUÇÃO

Os ecossistemas florestais tropicais evidenciam a importância do armazenamento de carbono, ao regularem o clima da Terra com a adição e remoção dos gases de efeito estufa, como o CO<sub>2</sub> atmosférico (BAKER et al., 2010; PECHANEC et al., 2018). No entanto, a mudança de cobertura do uso da terra tem grande influência nas mudanças ambientais, o que impacta tanto os serviços ecossistêmicos quanto a biodiversidade (MENDOZA-PONCE et al., 2018). O desmatamento e a degradação florestal colaboram com 15% a 20% das emissões globais de carbono, sendo que as regiões tropicais correspondem com a maior parte dessa contribuição (VICHARNAKORN et al., 2014). No Brasil, as alterações do uso da terra e o desmatamento são as práticas principais que ajudam para o deslocamento do carbono do solo para a atmosfera, e que coloca o país em quarto lugar no *ranking* dos países que mais liberam CO<sub>2</sub> para a atmosfera (MATTEWS et al., 2014).

Apesar desse panorama negativo, o Brasil possui importância mundial, tendo em vista ainda seu vasto recobrimento do solo pelas florestas. Esses ambientes consistem em um dos maiores reservatórios de carbono, atuando de forma significativa no ciclo global do carbono. No componente arbóreo, por exemplo, para a sua quantificação, carece inicialmente da determinação do estoque de biomassa e assim chegar ao estoque de carbono. Desta forma, a estimativa da biomassa e do teor de carbono tornam importantes na avaliação dos estoques e fluxos desse elemento nas florestas tropicais, auxiliando no conhecimento do ciclo global de carbono e possíveis ações na mitigação das mudanças climáticas (CHINEMBIRI et al., 2013; VICHARNAKORN et al., 2014).

No entanto, apesar da importância de se quantificar o estoque de carbono nas florestas, os métodos tradicionais dependem de campanhas de campo, que usualmente são dispendiosas em tempo e recursos financeiros (VICHARNAKORN et al., 2014). Diante dessas condições, há uma carência desses estudos nas florestas localizadas em regiões tropicais, gerando informações pontuais, no nível de árvore individual, e usualmente associados a variáveis dendrométricas, como observados nos trabalhos de Ribeiro et al., (2011); Sanquetta et al., (2014).

Dentro do contexto da estatística, os modelos de regressão são os mais difundidos por produzir boas estimativas e acurácia. No entanto, o uso da regressão quando envolve banco de dados com um grande número de variáveis e observações, pode apresentar limitações, uma vez que, sua aplicação demanda o atendimento a pressuposições estatísticas e também o

conhecimento prévio do conjunto de dados. Assim, em busca de técnicas mais eficientes, os métodos de inteligência computacional estão sendo utilizados para superar essas limitações (DRAKE; RAIDIN; GUIBAN, 2006; WERE et al., 2015). Esses métodos apresentam como vantagens a independência de conhecimento e suposições sobre as relações entre as variáveis, conseguem uma maior generalização das estimativas, menor susceptibilidade a ruídos e *outliers* (CORDEIRO et al., 2015; VIEIRA et al., 2018). Essas vantagens têm incentivado a aplicação de inteligência computacional na modelagem do estoque de carbono, como redes neurais artificiais (VAHEDI, 2016) e *support vector machine* (GLEASON; IM, 2012; VAFAEI et al., 2018). Técnicas promissoras como o algoritmo genético, *random forest*, *simulated annealing* ou a hibridização desses, também estão sendo exploradas e testadas na área do conhecimento chamada de *data mining* (mineração de dados), que envolve a redução de informações redundantes e irrelevantes do conjunto de dados com o propósito de aumentar a eficiência dos modelos de predição (BAGHERZADEH-KHIABANI et al., 2016).

Nesse sentido, a hibridização de métodos consiste em aprimorar o desempenho de técnicas já conhecidas, por meio da separação de problemas complexos em subproblemas. Assim, a resolução desses problemas é realizada a partir de diferentes técnicas com o objetivo de procurar as melhores soluções. Conforme exposto, é fundamental dispor de estimativas mais precisas e com maior qualidade sobre o estoque de carbono nas florestas, para garantir ações mitigadoras efetivas nas mudanças climáticas (SCOLFORO et al., 2015; VAHEDI, 2016). Além disso, é importante também compreender a influência de outras variáveis sobre o estoque de carbono total ou parcial (por compartimentos), visto que a maioria dos estudos se concentram apenas no diâmetro, altura e densidade da madeira. Portanto, um caminho promissor tem sido a aplicação de algoritmos híbridos na resolução de problemas complexos.

Neste caso, aplicando técnicas automatizadas de inteligência computacional tem-se vantagens na modelagem, quando há um enorme espaço amostral de possibilidades e combinação entre variáveis. A premissa principal do estudo está na abordagem de seleção automática de variáveis pelo método híbrido, identificando vantagens e desvantagens em seu uso. Logo, objetivou-se a integração da meta-heurística *Simulated Annealing* para a seleção de variáveis que melhor explicam o comportamento do carbono acima do solo da floresta nos diferentes compartimentos, juntamente com a aplicação do algoritmo *Random Forest* para predição do carbono em nível de árvore. Abordando-o de forma comparativa com a metodologia desenvolvida pelos métodos clássicos, envolvendo análise de regressão e também a equação pan-tropical, considerando aspectos como ganho na precisão dos modelos

e redução do número de variáveis utilizadas. De forma complementar, buscou-se analisar as tendências entre as variáveis selecionadas, e se essas afetam significativamente o poder explicativo dos métodos avaliados.

## 2. MATERIAL E MÉTODOS

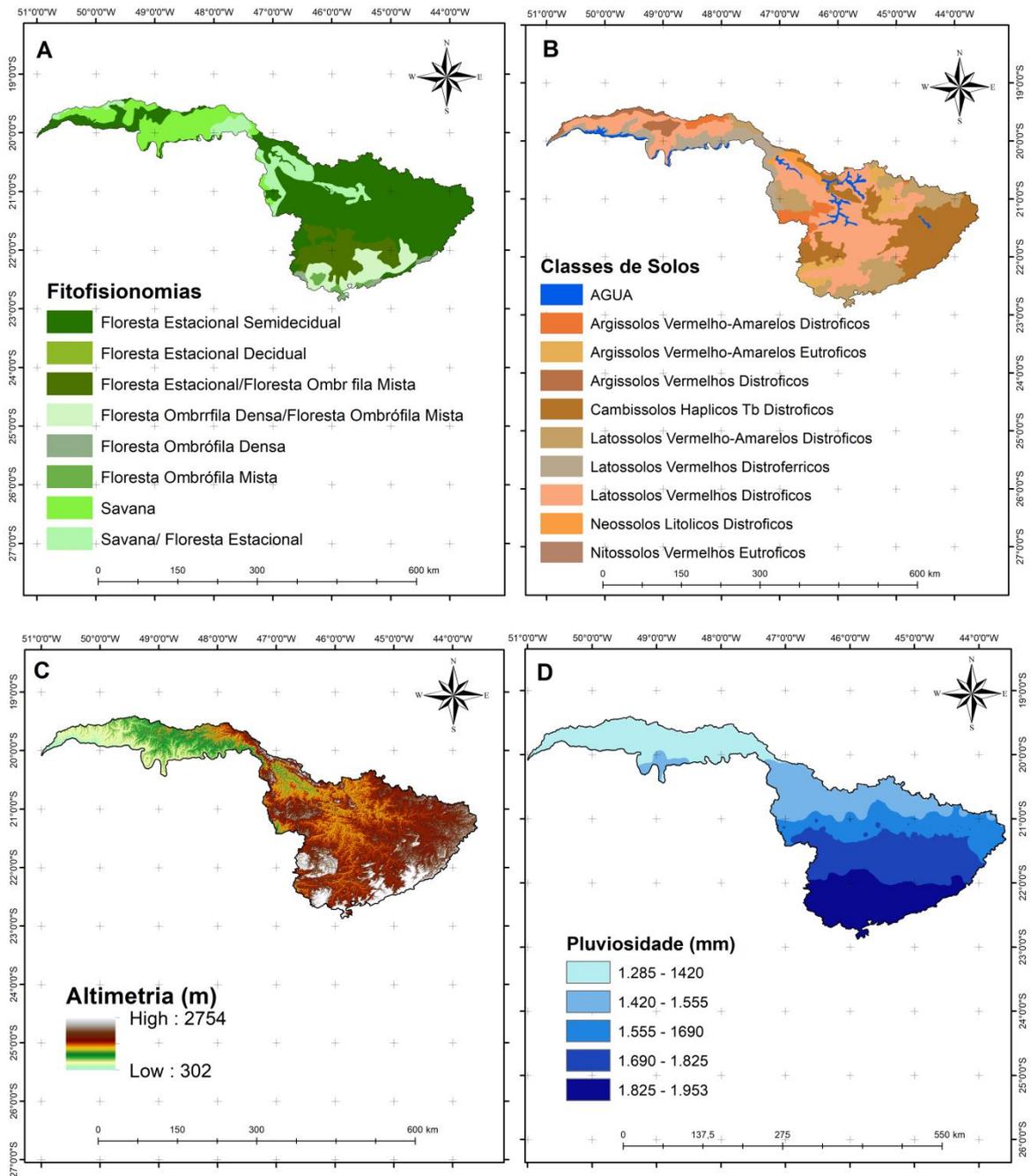
### 2.1 Área de estudo

A área de estudo compreende a Bacia Hidrográfica do Rio Grande, localizada ao Sul do estado de Minas Gerais, ocupando uma área em torno de 86.110 km<sup>2</sup>, o que representa 14,7% do território de Minas Gerais (Figura 3). Na bacia são encontradas as maiores altitudes, variando de 300 a 2.700 m, aproximadamente. As maiores cotas altimétricas estão localizadas na parte sul da Bacia e as menores cotas, no Triângulo Mineiro (Figura 3 - C). Segundo a classificação climática, baseada no Índice de Umidade de Thornthwaite, a classe climática Úmido B2 é encontrada na região Leste da Bacia do Rio Grande, em que o índice de umidade varia de 40 a 60%, com temperatura e precipitação médias anuais variando de 19 a 20°C e 1.500 a 1.600 mm, respectivamente. A caracterização climática do tipo Úmido B3 é identificada na parte Central da Bacia com umidade entre 60 e 80%, com índice de chuvas anual superior a 1.600 mm e temperatura média inferior a 18°C. Na região Sul da Bacia verificam duas classes climáticas: o tipo Úmido B4 e Superúmido. A classe Úmido B4 apresenta características de umidade mais elevada associada a temperaturas mais baixas, sendo influenciadas pelas regiões serranas, com umidade variando de 80 a 100%, índice pluviométrico maior que 1.700 mm e temperaturas amenas. O clima Superúmido tem índices climáticos com valor superior a 100, temperaturas médias anuais inferiores a 14°C e média acumulada de precipitação acima de 1.750 mm (Figura 3 - D) (CARVALHO et al., 2008).

A Bacia do Rio Grande está situada numa área de transição entre dois biomas, Cerrado e Mata Atlântica (Figura 3 - A), com elevada antropização. A bacia apresenta formações vegetais que variam desde espécies gramíneo-lenhosas do Cerrado a espécies lenhosas de área de Floresta Estacional Decidual. Os tipos principais de solos existentes na Bacia do Rio Grande são Argissolo, Cambissolo, Neossolo e Latossolo, sendo este último com maior ocorrência (Figura 3 - B) (CARVALHO et al., 2008). Na bacia existem três fisionomias bem características que podem ser destacadas, floresta ombrófila, floresta estacional semidecidual e cerrado. Por isso, foram selecionadas três regiões referentes a cada fisionomia para realizar

a coleta dos dados, sendo as regiões Frutal e Comendador Gomes (Triângulo Mineiro), Campanha (Centro-Sul) e Passa Quatro (Serra da Mantiqueira).

Figura 1 – Mapa de localização da Bacia Hidrográfica do Rio Grande no estado de Minas Gerais, sendo (A) Classificação da vegetação, (B) Classificação dos tipos de solos, (C) Classificação altimétrica, e (D) índices de pluviosidade.



Sistema de Coordenadas Geográficas. Datum: WGS 1984 / 23S

Fonte: Da autora (2020).

## 2.2 Amostragem e seleção das árvores individuais

Inicialmente, um inventário florestal foi realizado envolvendo um total de 29 fragmentos de vegetação nativa distribuídos ao longo da Bacia. A seleção das áreas buscou representar as variações edafo-climáticas e tipos vegetacionais presentes nos limites da Bacia.

As unidades amostrais foram sistematizadas nas áreas, variando conforme dimensão dos fragmentos. As unidades amostrais constituíam os conglomerados, sendo estes formados por três elementos ou sub-parcelas, com dimensão de 10 x 25 m (250 m<sup>2</sup>) cada, distantes 50 m um do outro, sendo cada conglomerado separado por uma distância de 75 m. Após a realização e análise do inventário florestal, dividiu-se a bacia em três estratos ou grupos florísticos, considerando uma análise em cluster e a diversidade florística. Finalmente, considerando os padrões fitossociológicos selecionaram-se as espécies mais importantes para cada estrato, concentrando a amostragem nestes indivíduos. Um total de 227 árvores foram cubadas envolvendo os três grupos florísticos, sendo: I (cerrado) - 77 árvores, II (semidecidual) - 69 árvores e III (ombrófila mista) - 81 árvores. A seleção das árvores buscou representar bem o padrão de cada grupo, amostrando todas as classes de diâmetro (Tabela 1), quando possível.

Tabela 1 - Distribuição diamétrica dos indivíduos selecionados na amostragem para o procedimento de cubagem rigorosa e coleta das demais variáveis morfométricas.

Classes de diâmetro	Faixa de DAP	Grupo florístico		
		Cerrado	Semidecídua	Ombrófila
1	5-10	10	15	10
2	10-15	9	10	9
3	15-20	10	10	12
4	20-25	9	10	10
5	25-30	11	10	11
6	30-35	8	9	10
7	35-40	7	4	8
8	>40	13	1	11
Total		77	69	81
Total geral		227		

DAP: diâmetro à altura do peito (cm).

Fonte: Da autora (2020).

## 2.3 Variáveis biométricas das árvores

Uma árvore pode apresentar uma série de variáveis úteis na quantificação e explicação de atributos estratégicos como volume e carbono, através de relações alométricas. Nesse estudo foram coletadas as variáveis clássicas, como o diâmetro à altura do peito (DAP) e altura total (HT), bem como a área de copa (AC), altura comercial (Hc) – altura até o diâmetro de 3 cm, altura da base da copa (Hbc) – altura até a primeira bifurcação, sendo discriminado a espécie e sua classe de diâmetro. Além dessas variáveis, também foram utilizadas a latitude, longitude, classe de diâmetro e altura, volume do fuste, dos galhos e total, densidade básica da madeira do fuste, dos galhos e total, peso seco das folhas, dos galhos, do fuste e total, diâmetro nas posições 0%, 25%, 50%, 75% e 100% e área basal de cada espécie. Essas variáveis puras foram aplicadas como entrada nos métodos de predição, assim como as transformações logarítmicas, inverso, raiz quadrada, segunda e terceira potências (132 variáveis). As combinações entre as variáveis puras com os operadores aritméticos multiplicação e divisão também foram utilizadas (922 variáveis), com o propósito de gerar um maior número de opções de entradas e também avaliar o poder explicativo das variáveis.

Por outro lado, a variável dependente foi o carbono presente separadamente nos compartimentos: fuste, galhos, folhas e total da árvore. Assim, a quantificação da biomassa foi feita a partir de uma amostragem destrutiva (cubagem rigorosa) sendo a biomassa aérea estratificada por meio dos componentes (fuste, galho e folha), para assim se obter no final do processo a massa em carbono. Para tal, foi utilizada a metodologia proposta pela Food and Agricultural Organization of the United Nations - FAO (PICARD; SAINT-ANDRÉ; HENRY, 2012).

#### 2.4 Estratégias de modelagem do carbono

Os métodos de predição utilizados neste estudo foram discriminados em três grupos: 1) modelos clássicos e a equação pan-tropical (CHAVE et al., 2005), 2) aplicação da técnica de *Stepwise* e 3) uso de técnicas de inteligência computacional. À exceção da equação pan-tropical, os demais métodos foram empregados separadamente para os compartimentos folha, galho, fuste e total. A aplicação desses métodos considerou 5 estratégias distintas para estimar o estoque de carbono em cada compartimento, com base em: a) diâmetro à altura do peito (DAP) e altura total (Ht), b) DAP, Ht e a densidade básica da madeira, c) todas as variáveis puras, d) variáveis puras e transformadas e, por último, e) com todas as variáveis (puras, transformadas e combinadas). Em seguida, as variáveis selecionadas, indiferentemente dos

métodos, foram correlacionadas com o carbono para os diversos compartimentos utilizando a correlação de Pearson.

Inicialmente, a base de dados foi dividida de forma aleatória e estratificada em dois conjuntos independentes e representativos. O primeiro, chamado de conjunto de treinamento, contendo 174 observações (80%), o segundo, conjunto de validação, com 53 observações (20%). O conjunto de treinamento foi utilizado para ajustar/treinar os métodos e o conjunto de validação para avaliar a capacidade preditiva dos métodos.

#### 2.4.1 Estatística clássica via modelos de regressão

A partir de testes prévios com os modelos de regressão mais recorrentes na literatura, selecionou-se o modelo de Spurr Linear para ser comparado com os demais métodos por meio de suas estatísticas (análise gráfica dos resíduos e o erro padrão residual (Syx), nas suas formas absoluta e relativa). A estimativa dos coeficientes desse modelo linear foi obtida por meio do método dos mínimos quadrados ordinários (MQO). Além desses modelos, adotou-se o procedimento *Stepwise* para gerar um novo modelo linear múltiplo. O método é amplamente aplicado para encontrar um número de variáveis na inclusão nos modelos de regressão. Dessa forma, foram construídos dois modelos, o primeiro, STEP 1, elaborado somente com as variáveis puras, e o segundo, STEP 2, considerando as variáveis puras e transformadas, devido a limitação do método que não consegue convergir com um número de variáveis maior que o número de observações. Adotou-se o fator de inflação da variância (VIF) para identificar multicolinearidade caso existisse, considerando valores inferiores a 10 unidades. Para tal foi utilizado o pacote car do R (FOX et al., 2019).

A equação (1) de abrangência pan-tropical (CHAVE et al., 2005) foi testada para estimar a biomassa acima do solo de cada árvore e, posteriormente, o estoque de carbono. Devido à natureza do modelo, foi utilizado um fator de 0,417 para conversão da biomassa para estoque de carbono. Esse fator foi obtido a partir da média dos valores de teor de carbono de cada árvore obtido em campo. A equação inclui as variáveis diâmetro à altura do peito (DAP), altura total (HT) e densidade básica da madeira (DB), sua validação foi apenas para o compartimento total.

$$Y = (0,0559 * DAP^2 * HT * DB) * 0,417 \quad (1)$$

## 2.4.2 Inteligência computacional

### 2.4.2.1 *Random Forest*

O algoritmo *Random Forest* (RF) (BREIMAN, 2001) é um método estatístico, não paramétrico, muito popular e eficaz tanto para problemas de regressão quanto classificação, que pode ser atribuída pela sua parametrização simples, o alto desempenho preditivo e a capacidade de trabalhar com valores faltantes, ruídos e alta dimensão de dados (GENUER, POGGI, TULEAU-MALOT, 2010; HAPFELMEIER; ULM, 2014). Dessa forma, o algoritmo foi escolhido para modelar o estoque de carbono neste estudo. Assim, a partir de testes prévios, considerou-se o número de árvores (*ntrees*) em 500 unidades, o número de atributos a serem sorteados (*mtry*) como 2 e o número de observações nos nós terminais de cada árvore (*nodesize*) como 5. Na ocasião em que foi utilizado somente DAP e HT o *mtry* utilizado foi igual a 1. O algoritmo RF foi aplicado com 50 repetições e o erro médio quadrático dessas repetições foi utilizado para filtrar o modelo com menor valor de erro.

### 2.4.2.2 *Simulated Annealing e Random Forest (SARF)*

Este método híbrido é uma associação entre a metaheurística *Simulated Annealing* e o algoritmo *Random Forest*. Ele tem o propósito multiobjectivo, ou seja, redução no número de variáveis selecionadas pelo algoritmo de treinamento, bem como a busca pela melhor precisão (erro). A meta-heurística tem como tarefa o gerenciamento do algoritmo de predição, identificando combinações mais viáveis do conjunto envolvendo algoritmo e variáveis. Essa hibridação de métodos permite explorar melhor o potencial dos métodos de predição, pois possibilita gerenciar de forma autônoma e inteligente estes algoritmos.

A implementação da meta-heurística *Simulated Annealing* (SA) foi utilizada para encontrar o conjunto ideal de variáveis, ou seja, trabalha no gerenciamento de variáveis a serem utilizadas pelo *Random Forest*. No contexto da seleção de variáveis, o procedimento torna-se um problema de otimização multiobjectivo, em que dois objetivos são observados, sendo um número mínimo de variáveis na solução e uma maior precisão na predição. Os objetivos foram normalizados para que tenham o mesmo peso e efeito sobre a função resultante  $f(x)$ . Dessa forma, foram estabelecidos previamente, por meio de testes, uma

temperatura inicial ( $T_0$ ) igual 100.000 e temperatura final ( $T_F$ ) igual a 25, sendo esse último valor utilizado como critério de parada do algoritmo. A taxa de resfriamento utilizada foi igual a 1% e a função objetivo empregada consistiu na minimização do erro médio quadrático e o número de variáveis. Nesse sentido, a função de avaliação é formada pela soma de dois termos, o primeiro, a razão entre o *erroOOB* obtido pelo RF e o seu máximo observado via testes, sendo esse determinado como uma constante. A segunda parte da equação representa uma razão entre o número de variáveis selecionadas ( $n$ ) e o número total de variáveis candidatas testadas ( $N$ ). O sentido da função multiobjectivo foi minimizar  $f(x)$  (Equação 2).

$$f(x) = \frac{\text{erroOOB}}{\text{erroOOB}_{\max}} + \frac{n}{N} \quad (2)$$

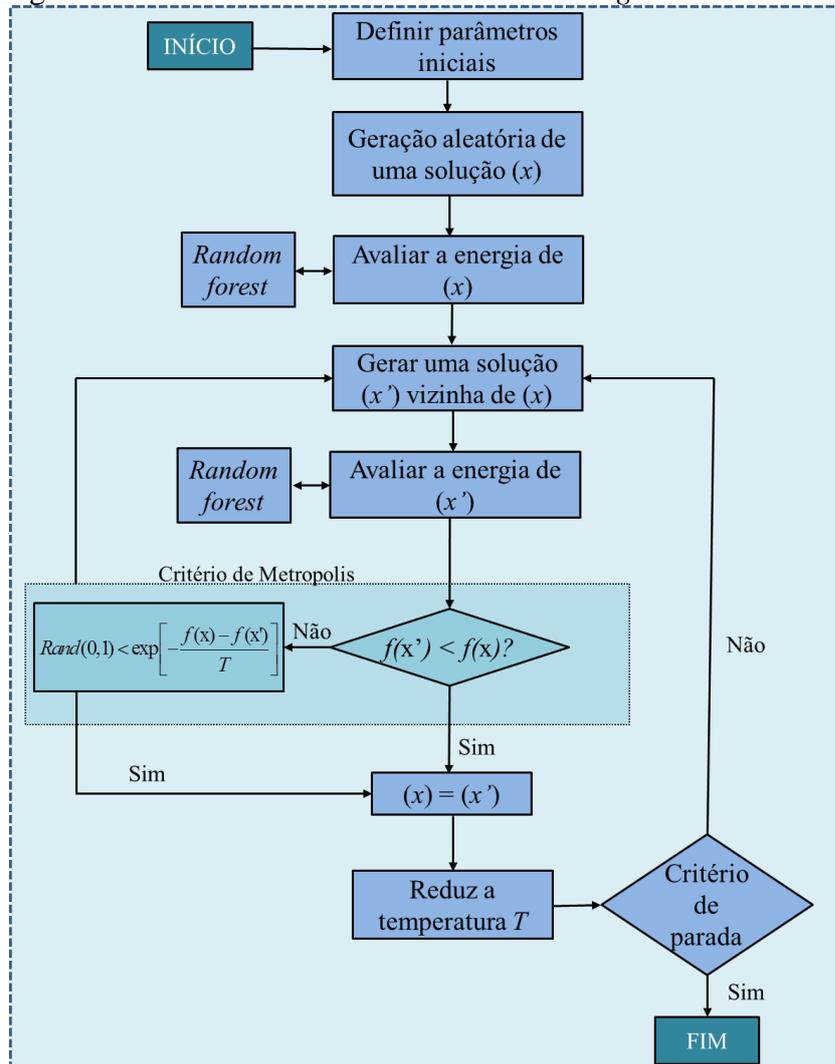
No primeiro estágio, o algoritmo SA inicia com uma solução inicial ( $x$ ) gerada aleatoriamente, sendo representada por um vetor dimensional com comprimento de acordo com o número de variáveis selecionadas. A dimensão do mesmo está associada ao número de variáveis independentes do conjunto de dados, sendo o estoque de carbono a variável fixa investigada. A cada iteração uma nova solução  $x'$  é gerada, conforme probabilidades de troca (0,1), inclusão (0,9), exclusão (0,1), transformação (0,7) e combinação (0,7), definidas a partir de testes. Assim, um valor aleatório é gerado a cada iteração e caso esse valor fosse maior que as probabilidades mencionadas, realizariam algum ou todos os procedimentos (troca, inclusão, exclusão, transformação e/ou combinação).

A solução  $x'$  será aceita para ser a nova solução atual dependendo de  $T$  e dos valores da função de avaliação, se a  $f(x') < f(x)$  então  $x'$  é aceito e substitui  $x$ . No entanto, se esta condição não for atendida,  $x'$  ainda poderá ser aceita com uma determinada probabilidade estabelecida pela probabilidade de Boltzmann do algoritmo de Metropolis (Equação 3), no qual permite que o SA não fique concentrado somente em uma região do espaço de busca. Durante o processo de busca, a temperatura  $T$  é reduzida e a probabilidade de aceitar soluções inferiores diminui gradualmente (APÊNDICE I) (MAFARJA e MIRJALILI, 2017). Em que:  $Rand(0,1)$ : número aleatório entre 0 e 1 conforme distribuição de probabilidade

uniforme;  $T$ : temperatura atual do sistema;  $x$ : solução atual;  $x'$ : solução vizinha;  $\exp$ : exponencial.

$$\left\{ \begin{array}{ll} \text{Rand}(0,1) < \exp\left[-\frac{f(x)-f(x')}{T}\right] & \text{Aceita a nova solução } (x') \\ \text{Caso contrário} & \text{Não aceita a nova solução } (x') \end{array} \right. \quad (3)$$

Figura 2 - Fluxograma do modelo híbrido *Simulated Annealing* e *Random Forest*.



Fonte: Adaptado de Gomide; Arce, Silva (2013).

No segundo estágio, as variáveis selecionadas pelo algoritmo SA foram utilizadas como predictoras no *Random Forest*. Esse consiste em um conjunto de árvores que são formadas a partir de uma amostra *bootstrap* definida aleatoriamente para os dados de treinamento. A divisão dos nós de cada árvore ocorre por meio da seleção de uma variável de

forma aleatória, sendo assim, obtêm-se diferentes árvores. O modelo *Random Forest* utiliza um conjunto de validação de dados *out-of-bag* (OOB) para cada árvore, em que os valores das variáveis selecionadas para a divisão dos nós são permutados. Para tal, esse algoritmo requer uma parametrização inicial, sendo essa a mesma adotada para o *Random Forest* puro. Ao final do processo, a predição é calculada com base na média de todas as árvores. E por fim, para reduzir o efeito aleatório da solução inicial todo o processamento foi repetido 50 vezes e, então, a melhor solução foi selecionada.

## 2.5 Avaliação comparativa de desempenho entre métodos

Devido à natureza distinta dos métodos, a avaliação entre eles seguiu uma abordagem considerando as análises de resíduo, conforme as métricas estatísticas: a raiz do erro quadrático médio absoluto e percentual (RMSE) (Equação 4 e 5) e coeficiente de determinação ( $R^2$ ) (Equação 6), sendo  $n$ : número de observações;  $i$ : sub-índice das observações;  $Y$ : valor observado do estoque de carbono (kg),  $\hat{Y}$ : valor estimado do estoque de carbono (kg). Os dados de validação foram usados nessa etapa para quantificar as métricas.

$$RMSE = \frac{1}{n} \sum_{i=1}^n \sqrt{(Y_i - \hat{Y}_i)^2} \quad (4)$$

$$RMSE\% = \frac{RMSE}{\bar{Y}} * 100 \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2} \quad (6)$$

O processamento dos dados foi realizado no software R (*Version 3.5.3* – © 2019 RStudio, Inc.), com o auxílio dos pacotes *randomForest* (LIAW; WIENER, 2002), bem como o desenvolvimento do método híbrido. A metodologia foi processada numa CPU com processador Intel(R) Core™ i3-2100 CPU @ 3.10 GHz, com memória instalada (RAM) de 8,0 GB.

### 3. RESULTADOS

A análise exploratória dos valores do estoque de carbono (kg/árvore) para os diferentes compartimentos da árvore está apresentada na Tabela 2. O comportamento dos dados mostra, a partir dos valores de coeficiente de variação, uma alta heterogeneidade dos valores de estoque de carbono independente do compartimento. Esse fato é comprovado por meio das características da área estudada, em que exibe uma transição dos biomas Cerrado e Mata Atlântica. Esses biomas apresentam regiões desde florestas com alta densidade e árvores de grande porte até regiões com baixo número de indivíduos e árvores pequenas.

Tabela 2 - Análise exploratória do teor de carbono (kg) nos diferentes compartimentos da árvore.

Compartimento	Média (kg/árvore)	Mínimo (kg/árvore)	Máximo (kg/árvore)	Desvio padrão (kg/árvore)	CV (%)
Folha	6,44	0,14	53,18	7,10	110,27
Galho	74,04	0,51	658,37	102,24	138,09
Fuste	80,05	1,51	652,67	96,68	120,78
Total	147,24	2,62	1005,23	188,84	128,25

Em que: CV (%): coeficiente de variação em porcentagem.

Fonte: Da autora (2020).

O estoque de carbono (kg/árvore) foi estimado a partir do ajuste do modelo Spurr linearizado (escolhido dentre os modelos testados) e pela técnica *Stepwise*. Os parâmetros ajustados para o modelo de Spurr e os novos modelos desenvolvidos (via *Stepwise*) são apresentados na Tabela 3. Na tabela também são observadas todas as variáveis selecionadas em cada modelo (via *Stepwise*) com valores de VIF entre 1,15 e 9,05.

Tabela 3 - Equações ajustadas via método *Stepwise* e do modelo de Spurr linearizado para a estimativa do estoque de carbono total, no fuste, nos galhos e nas folhas. (continua)

Modelos		Equação
Spurr Linear	Total	$\ln(Y) = -4,203 + 1,008 \cdot \ln(DAP^2 \cdot HT)$
	Fuste	$\ln(Y) = -4,299 + 0,942 \cdot \ln(DAP^2 \cdot HT)$
	Galho	$\ln(Y) = -6,360 + 1,143 \cdot \ln(DAP^2 \cdot HT)$
	Folha	$\ln(Y) = -4,228 + 0,649 \cdot \ln(DAP^2 \cdot HT)$
STEP 1	Total	$Y = 0,127 + 0,431 \cdot PST - 0,256 \cdot PSFo$
	Fuste	$Y = -5,811 + 0,437 \cdot PSFu - 0,141 \cdot PSFo + 9,95 \cdot 10^{-6} \cdot X$
	Galho	$Y = 0,431 \cdot PSG - 0,142 \cdot PSFo$
	Folha	$Y = 0,437 \cdot PSFo + 0,014 \cdot AC - 0,017 \cdot D25$

Tabela 3 - Equações ajustadas via método Stepwise e do modelo de Spurr linearizado para a estimativa do estoque de carbono total, no fuste, nos galhos e nas folhas. (conclusão)

<b>STEP 2</b>	Total	$Y = 0,410.PST - 1,905.VT^3 + 6,82.10^{-4}.D50^3 + 7.19.10^{-09}.PST^3$
	Fuste	$Y = 13,740 + 0,001.DAP^3 + 0,008.C\_HT^3 - 24,070.VTG^2 + 2,72.10^{-7}.PSFu^3$
	Galho	$Y = 0,420.PSG + 7,53.10^{-9}.PSG^3 + 4,98.10^{-5}.D25^3 + 5,15.10^{-4} \cdot \frac{1}{g} + 28,240.g^3 - 0,012.C\_HT^2 + 0,001.HC^3 - 0,003.PSFo^2$
	Folha	$Y = -0,415 + 0,428.PSFo + 2,12.10^{-4}.AC^2 - 3,14.10^{-10}.PST^3 + 1,01.10^{-4}.HC^3 + 1,418.DBMP\_T^3$

Em que: STEP 1: modelo construído pelo *Stepwise* somente com variáveis puras; STEP 2: modelo construído pelo *Stepwise* somente com variáveis puras e transformadas; VIF: fator de inflação da variância; *PST*: peso seco total (kg); *PSFo*: peso seco de folhas (kg); *X*: longitude; *PSG*: peso seco dos galhos (kg); *AC*: área de copa; *D25*: diâmetro aos 25% da árvore; *VT*: volume total; *D50*: diâmetro aos 25% da árvore; *VTFu*: volume total do fuste; *DBMP\_T*: densidade básica total da madeira; *N*: densidade de indivíduos de cada espécie; *DAP*: diâmetro à altura do peito; *C\_HT*: classe de altura total; *HT*: altura total; *VTG*: volume total dos galhos; *PSFu*: peso seco do fuste; *g*: área seccional; *HC*: altura comercial.

Fonte: Da autora (2020).

As métricas estatísticas de avaliação para todas as equações utilizadas e testadas são apresentadas na Tabela 4. Nessa tabela verifica-se para o grupo 1, de forma geral, que as equações mostraram baixo desempenho para estimar o estoque de carbono dos compartimentos (kg/árvore), não acompanhando eficientemente os valores observados. Esse resultado evidencia que somente o uso do DAP e a altura não consegue estimar com precisão o estoque de carbono. A equação pan-tropical foi a que proporcionou maior precisão nas estimativas para o estoque de carbono total em comparação com o modelo de Spurr. O modelo de Spurr linearizado para os demais compartimentos mostraram alta tendenciosidade em superestimar os indivíduos com menores estoques de carbono. O maior desempenho da equação de Chave et al. (2005) pode ser atribuída à utilização da variável densidade básica, que geralmente apresenta correlação com o estoque de carbono, melhorando assim as estimativas.

Tabela 4 - Estatísticas de ajuste dos modelos ajustados para gerar equações para estimar o estoque de carbono (kg) para os compartimentos total, fuste, galho e folha.

Est	Grupo	Métodos	N	Treino	Validação
-----	-------	---------	---	--------	-----------

			Total	Fuste	Galho	Folha	Total	Fuste	Galho	Folha	
RMSE	1	Spurr	2	76,294	31,820	59,323	5,623	88,415	82,353	52,689	5,186
		Chave	3	70,627	-	-	-	90,608	-	-	-
	2	STEP 1	22	7,629	4,438	4,119	0,668	7,935	3,633	4,006	0,803
		STEP 2	132	6,720	30,311	3,301	0,606	14,786	67,445	3,637	0,686
	3	RF <sub>1</sub>	2	44,658	17,356	35,032	3,154	102,465	82,217	40,954	4,528
		RF <sub>2</sub>	3	38,207	14,853	29,247	2,976	99,963	84,463	40,959	4,311
		RF <sub>3</sub>	22	22,558	11,334	17,703	1,966	61,564	66,868	28,167	2,918
		RF <sub>4</sub>	132	23,162	10,415	19,304	2,006	62,834	67,313	27,926	2,944
		RF <sub>5</sub>	1054	22,151	10,430	16,110	1,779	54,952	62,563	27,900	2,714
		SARF	1054	16,234	8,781	13,025	1,496	29,998	57,353	20,157	1,757
RMSE (%)	1	Spurr	2	51,942	40,966	79,918	88,118	57,871	94,013	71,761	78,319
		Chave	3	48,084	-	-	-	59,307	-	-	-
	2	STEP 1	22	5,194	5,714	5,548	10,467	5,194	4,147	5,456	12,120
		STEP 2	132	4,575	39,024	4,447	9,490	9,678	76,994	4,954	10,355
	3	RF <sub>1</sub>	2	28,214	22,345	47,195	49,423	61,121	93,858	55,778	68,381
		RF <sub>2</sub>	3	24,138	19,122	39,401	46,630	59,629	96,421	55,785	65,108
		RF <sub>3</sub>	22	14,252	14,592	23,850	30,806	36,742	76,335	38,363	44,063
		RF <sub>4</sub>	132	14,633	13,409	26,006	31,441	37,481	76,843	38,035	44,460
		RF <sub>5</sub>	1054	15,081	13,428	21,703	27,875	35,968	71,421	38,000	40,995
		SARF	1054	11,053	11,306	17,547	23,450	19,635	65,473	27,453	26,540
R <sup>2</sup>	1	Spurr	2	0,831	0,876	0,671	0,387	0,805	0,480	0,713	0,425
		Chave	3	0,856	-	-	-	0,795	-	-	-
	2	STEP 1	22	0,998	0,998	0,998	0,991	0,998	0,999	0,998	0,986
		STEP 2	132	0,999	0,887	0,999	0,993	0,995	0,650	0,999	0,990
	3	RF <sub>1</sub>	2	0,945	0,963	0,885	0,807	0,740	0,482	0,827	0,562
		RF <sub>2</sub>	3	0,959	0,973	0,920	0,828	0,753	0,453	0,827	0,603
		RF <sub>3</sub>	22	0,986	0,984	0,971	0,925	0,906	0,657	0,918	0,818
		RF <sub>4</sub>	132	0,985	0,987	0,965	0,922	0,902	0,653	0,920	0,815
		RF <sub>5</sub>	1054	0,986	0,987	0,976	0,936	0,925	0,700	0,920	0,842
		SARF	1054	0,992	0,991	0,984	0,957	0,978	0,748	0,958	0,934

Em que: RMSE: erro quadrático médio absoluto; RMSE%: erro quadrático médio percentual; R<sup>2</sup>: coeficiente de determinação; STEP 1: modelo construído pelo *Stepwise* somente com variáveis puras; STEP 2: modelo construído pelo *Stepwise* somente com variáveis puras e transformadas; RF<sub>1</sub>: *Random Forest* utilizando DAP e HT; RF<sub>2</sub>: *Random Forest* utilizando DAP, HT e Densidade básica; RF<sub>3</sub>: *Random Forest* utilizando todas as variáveis puras; RF<sub>4</sub>: *Random Forest* utilizando todas as variáveis puras e transformadas; RF<sub>5</sub>: *Random Forest* utilizando todas as variáveis puras, transformadas e combinadas; SARF: *Simulated Annealing* com *Random Forest* com todas as variáveis puras, transformadas e combinadas.

Fonte: Da autora (2020).

A partir da técnica *Stepwise* (Grupo 2) foram desenvolvidas duas equações para cada compartimento da árvore (folha, galho, fuste e o total), visando à predição do estoque de carbono (kg) em nível árvore. Além disso, objetivou-se também encontrar as variáveis mais

explicativas sobre o comportamento do estoque de carbono nas árvores. Assim, ao observar as estatísticas dos modelos gerados para os dados de treinamento e validação constatou-se que a estratégia STEP 2, construída com variáveis puras e transformadas, apresentou modelos ligeiramente mais eficientes na estimativa do estoque de carbono dos compartimentos galho e folha. Já a estratégia STEP 1, formada por variáveis puras, demonstrou maior superioridade nas estimativas do estoque de carbono do fuste e total.

Sobre as variáveis selecionadas, no geral, é observado que o peso seco total, do fuste, dos galhos e das folhas obtiveram elevada capacidade de explicação para todos os compartimentos. Uma vez que sempre uma ou mais dessas variáveis foram escolhidas na construção dos modelos, tanto em sua forma pura quanto transformada. Esse resultado evidencia a relação do estoque de carbono com a biomassa, em que essa é utilizada na elaboração de um fator de conversão para estoque de carbono. A densidade básica total e a área de copa mostraram maior relação com a estimativa do estoque de carbono total e da folha. Já a altura e o volume dos compartimentos de galho, fuste e total, apresentaram-se importantes para estimar o estoque de carbono total, do fuste e do galho. As demais variáveis selecionadas para cada compartimento podem ser verificadas na Tabela 3.

Alternativamente, o *Simulated Annealing* (SA) e o *Random Forest* (RF) (Grupo 3) foram utilizados para selecionar variáveis explicativas e estimar o estoque de Carbono (kg/árvore) dos diferentes compartimentos da árvore com maior precisão, respectivamente. A primeira abordagem refere-se à aplicação do algoritmo RF na sua forma pura (Grupo 3). Dessa maneira, foi possível estimar os estoques de carbono como também avaliar a contribuição de cada variável. E, posteriormente, realizar comparações com a metodologia híbrida empregada para seleção de variáveis e predição.

Ao analisar a importância das variáveis selecionadas/utilizadas para os modelos do Grupo 3, verificou-se que a variável diâmetro foi a mais importante para o modelo RF<sub>1</sub> (somente DAP e HT) e RF<sub>2</sub> (somente DAP, HT e DB). Entre os modelos RF<sub>1</sub> e RF<sub>2</sub> observou-se que a introdução da densidade básica não gerou melhorias significativas. Já no modelo RF<sub>3</sub> que utilizou todas as variáveis puras, pode-se constatar que o peso seco, de forma geral, apresentou-se mais frequente entre as primeiras variáveis que mais contribuem para o modelo. Além dessa, as variáveis DAP, volume total, do fuste e dos galhos e os diâmetros a 0% e a 25% foram as variáveis encontradas com maior valor de importância entre os compartimentos. A estratégia que utilizou as variáveis puras e transformadas (RF<sub>4</sub>) selecionou o peso seco e o volume, na forma pura e transformada, como as primeiras variáveis com

maior contribuição. O modelo que utilizou todas as variáveis disponíveis (RF<sub>5</sub>) apresentou, dentre todos os compartimentos (folha, galho, fuste e total), as variáveis transformadas e combinadas como mais representativas nas primeiras posições. Entre essas variáveis encontram-se as mesmas também selecionadas pelos modelos anteriores, e outras como a longitude, latitude, densidade básica, altura total, altura da base da copa e diâmetro.

O compartimento folha, a metodologia híbrida SARF, envolvendo o *Simulated Annealing*, encontrou um conjunto ótimo de 16 variáveis. A variável DAP, diâmetro à altura do peito, destacou-se com maior valor de importância. As demais variáveis foram peso seco das folhas (PSFo), razão entre peso seco total e o diâmetro à altura do peito (PST/DAP), longitude (X), raiz do diâmetro a 0% (D0\_raiz), raiz do volume total do fuste (VTFu\_raiz), altura total vezes a altura comercial (Ht\*HC), altura total (Ht), peso seco dos galhos (PSG), latitude (Y), área seccional (g), peso seco total (PST), logaritmo de volume total (VT\_log), raiz do peso seco das folhas (PSFo\_raiz), densidade básica total (DBMP\_T) e diâmetro a 25%. Quanto ao compartimento galho, o SARF resultou em um conjunto de menor erro com 12 variáveis. Esse conjunto é formado pelo peso seco dos galhos ao quadrado (PSG\_quad) com maior contribuição, seguido da raiz do volume total dos galhos (VTG\_raiz), inverso do peso seco total (PST\_inv), diâmetro à altura do peito (DAP), diâmetro a 0% (D0), área de copa (AC), raiz do diâmetro a 25% (D25\_raiz), classe de diâmetro vezes altura total (C\_DAP\*Ht), diâmetro a 75% (D75), área seccional (G), razão entre a classe de diâmetro e peso seco das folhas (C\_DAP/PSFo) e latitude (Y).

Em relação ao compartimento fuste, um total de 10 variáveis foram selecionadas, sendo essas: o peso seco do fuste (PSFu) com maior valor de importância, o volume total do fuste (VTFu), peso seco total (PST), diâmetro a 0% (D0), peso seco dos galhos (PSG), volume total dos galhos (VTG), altura da base da copa (HBC), logaritmo da densidade básica total (DBMP\_T\_log), latitude (Y) e área seccional (G). Por fim, para o estoque de carbono total 8 variáveis foram escolhidas. A variável de maior contribuição foi o peso seco total (PST), sucedido do volume total (VT), logaritmo do peso seco do fuste (PSFu\_log), classe de diâmetro (C\_DAP), área de copa (AC), densidade básica total (DBMP\_T), raiz da área basal (G\_raiz) e diâmetro a 100% (D100). Em vista disso, observa-se que algumas variáveis apresentam maior frequência dentre os compartimentos, sendo elas o peso seco e o volume tanto na sua forma pura quanto as transformações e combinações com outras variáveis. Essas evidenciam uma relação mais direta com o estoque de carbono. Outras variáveis como a

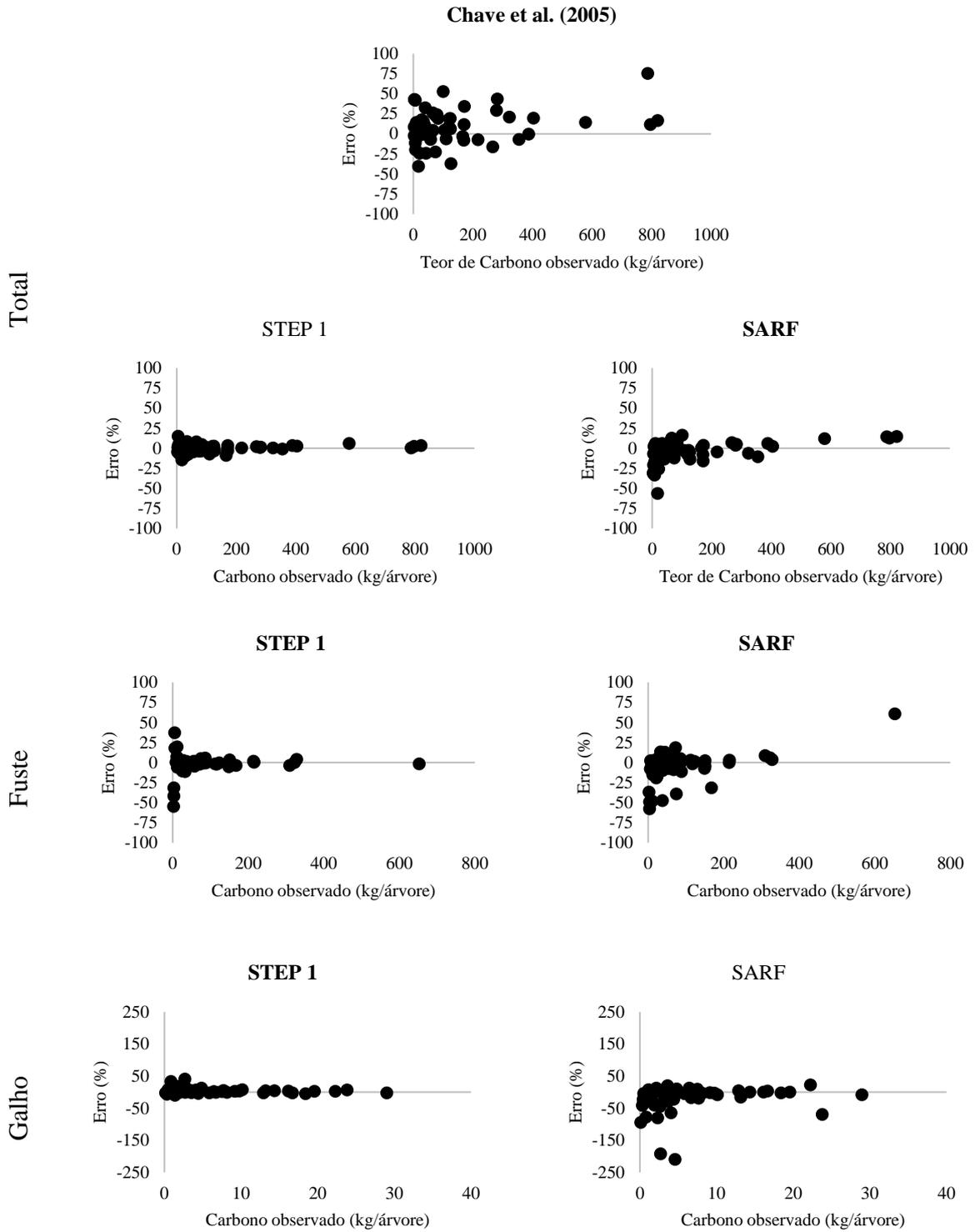
latitude, área de copa, densidade e diâmetro a 0% mostraram-se bem representativas na maioria dos compartimentos analisados.

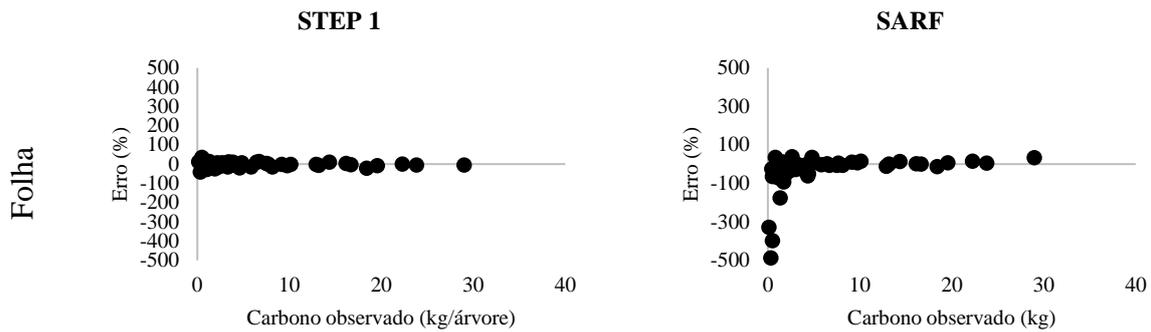
A partir desses resultados, nota-se que a metodologia proposta com a abordagem multiobjetivo, que visa a diminuir o número de variáveis e o erro, permitiu ao algoritmo SA reduzir satisfatoriamente o número de variáveis para os diferentes compartimentos. No que se refere à avaliação do erro obtido pelos modelos de predição, constata-se a partir da Tabela 4, a contribuição da seleção de variáveis feita pelo SA por meio da redução do erro das estimativas. O SARF apresentou melhores resultados de acordo com o RMSE% em comparação ao RF<sub>1</sub> (somente com DAP e HT), com reduções percentuais de até 29% para os dados de treinamento e 42% para os dados de validação. Esse resultado mostra que apenas as variáveis DAP e HT não conseguem explicar bem o comportamento do estoque de carbono na árvore. Para os valores de coeficiente de determinação ( $R^2$ ) os modelos apresentaram respostas próximas para os dados de treinamento, já para a validação o SARF se sobressaiu com valores ligeiramente superiores.

Embora o RF tenha apresentado desempenho próximo ao SARF no conjunto de treinamento, a redução do número de variáveis para os dados de validação proporcionou uma melhora significativa das estimativas para os diferentes compartimentos da árvore. É observada também a melhoria das estimativas à medida que mais variáveis eram utilizadas como entrada no RF. Pois a utilização de poucas variáveis no RF traz baixa diversificação que pode provocar um sobreajuste do modelo, como verificado nos modelos de estoque de carbono total e do fuste. No entanto, ao comparar o RF e o modelo de Spurr e de Chave et al. (2005) com o mesmo conjunto de variáveis, verifica-se que o RF apresentou estatísticas superiores.

A dispersão dos resíduos foi analisada a partir dos métodos que mais se destacaram entre cada grupo. Assim, foi observado que para o conjunto de treinamento todos os métodos apresentaram uma tendência em superestimar os indivíduos com menores estoques de carbono. Esse comportamento também pode ser verificado para o conjunto de validação (Figura 3). Dentre os métodos, o modelo STEP 1 mostrou uma distribuição residual mais homogênea para os compartimentos total, galho e folha.

Figura 3 – Distribuição residual dos melhores modelos gerados em cada grupo para estimar o estoque de carbono (kg/árvore).





Em que: SARF: *Simulated Annealing* com *Random Forest*; STEP 1: modelo construído pelo *Stepwise* somente com variáveis puras; Chave et al. (2005): equação pan-tropical.

Fonte: Da autora (2020).

Os métodos de seleção *Stepwise* e SARF não selecionaram o mesmo conjunto de variáveis para cada compartimento, contudo, algumas variáveis foram iguais. Sendo elas, o peso seco das folhas, galhos, fuste e total. Essas variáveis apresentam altos valores de correlação variando entre 0,66 e 1, como pode ser verificado na Tabela 5. Nessa tabela também são apresentados os valores das correlações lineares de Pearson para as demais variáveis selecionadas.

Tabela 5 - Coeficientes de correlação linear de Pearson entre o estoque de Carbono dos compartimentos total, fuste, galho e folha e as variáveis selecionadas entre os melhores métodos de seleção utilizados.

Variáveis	Compartimento				Variáveis	Compartimento			
	Total	Fuste	Galho	Folha		Total	Fuste	Galho	Folha
DAP <sup>3</sup>	0,86	0,79	0,83	0,69	D0	0,85	0,79	0,80	0,69
C_HT	0,55	0,58	0,45	0,44	VTFu	0,87	0,96	0,68	0,64
HC	0,65	0,70	0,53	0,46	DAP	0,86	0,81	0,81	0,70
D25/VTFu	-0,46	-0,46	-0,40	-0,40	Ht	0,57	0,63	0,45	0,39
C_DAP/PSFo	-0,21	-0,17	-0,20	-0,38	D25	0,75	0,75	0,66	0,57
D75	0,72	0,74	0,62	0,49	PSG	0,94	0,76	1,00	0,79
C_DAP*Ht	0,80	0,82	0,69	0,59	X	0,18	0,09	0,24	0,25
D25 <sup>0,5</sup>	0,72	0,72	0,64	0,56	PSFu	0,93	1,00	0,76	0,67
1/PST	-0,39	-0,38	-0,35	-0,37	D100	0,16	0,15	0,15	0,10
VTG <sup>0,5</sup>	0,91	0,79	0,92	0,77	g <sup>0,5</sup>	0,09	0,11	0,06	0,02
PSG <sup>2</sup>	0,79	0,58	0,90	0,68	DBMP_T	0,36	0,31	0,36	0,20
g	0,12	0,13	0,09	0,06	C_DAP	0,76	0,73	0,70	0,62
Y	0,11	0,00	0,20	0,18	ln(PSFu)	0,79	0,80	0,69	0,63
ln(DBMP_T)	0,32	0,28	0,33	0,19	VT	0,96	0,93	0,88	0,76
HBC	0,23	0,34	0,10	0,00	AC	0,84	0,67	0,88	0,80
VTG	0,92	0,75	0,96	0,78	PSFo	0,78	0,66	0,77	0,99
DAP <sup>3</sup>	0,86	0,79	0,83	0,69	PST	1,00	0,93	0,95	0,79

Em que: C\_HT: classe de altura; HC: altura comercial; D25: diâmetro a 25%; VTFu: volume total do fuste; C\_DAP: classe de diâmetro; PSFo: peso seco das folhas; D75: diâmetro a 75%; HT: altura total; PST: peso seco total; VTG: volume total dos galhos; PSG: peso seco dos galhos; g: área seccional; Y: latitude; DBMP\_T: densidade básica total; HBC: altura da base da copa; D0: diâmetro a 0%; AC: área de copa; PSFu: peso seco do fuste; DAP: diâmetro à altura do peito; X: longitude; D100: diâmetro a 100%; VT: volume total; DBMP\_Fu: densidade básica do fuste; D50: diâmetro a 50%; Ht: altura total.

Fonte: Da autora (2020).

Ao comparar os modelos, STEP 1 e SARF, para cada compartimento verifica-se uma maior dificuldade em modelar folha e galho, visto que apresentaram maior amplitude dos resíduos como evidenciado na Figura 3. A modelagem do estoque de carbono total destacou-se frente aos compartimentos devido a distribuição mais homogênea, concentrando em um menor intervalo (-60% e 20%). Portanto, o resultado apresentado justifica a aplicação do SARF para estimar o estoque de carbono, além de contribuir para a compreensão das variáveis que mais influenciam no estoque carbono. No contexto da seleção das variáveis explicativas, as técnicas analisadas encontraram um conjunto relativamente similar, no entanto, o *Stepwise* resultou em menores erros concernentes as estimativas.

#### 4. DISCUSSÃO

As estimativas do estoque de carbono presentes nas florestas nativas são importantes para compreender o papel dos ecossistemas no ciclo global de carbono, além de garantir o manejo sustentável dos recursos florestais. Diante das várias abordagens existentes para estimar o estoque de carbono, a quantificação exata e a avaliação dos fatores influentes, nos incrementos e decréscimos do armazenamento de carbono nas árvores e na floresta como um todo torna-se necessária (HENRY et al., 2011; SANQUETTA; DALLA CORTE; SILVA, 2011). No entanto, os métodos diretos de quantificação são difíceis e onerosos, já os métodos indiretos empregando equações alométricas e variáveis mensuráveis na árvore, são mais aplicados. Marziliano et al. (2017) ressaltam que geralmente os fatores padrão de conversão da biomassa são mais utilizados. Contudo, esses valores podem variar de acordo com a estrutura da floresta e as condições ambientais, portanto seu uso pode resultar em avaliações não confiáveis.

Nesse sentido, o presente estudo sugeriu testar diferentes metodologias para selecionar variáveis e estimar o estoque de carbono acima do solo da vegetação nativa. De acordo com os resultados, o modelo de Spurr não apresentou estimativas precisas do estoque de Carbono

dos compartimentos (Tabela 4). Já a equação pan-tropical de Chave et al. (2005) também avaliada para o Carbono total, destacou-se ao exibir maior homogeneidade dos resíduos. Porém, mesmo assim, essa equação apresentou um alto valor de  $Syx$  (%). Resultado similar foi encontrado por Segura, Acuña e Andrade (2017), que avaliaram e compararam a aplicação de modelos locais e globais em florestas tropicais no pacífico colombiano. Os autores constataram que a equação de Chave et al. (2005) e as demais equações globais analisadas mostraram comportamento de subestimar a biomassa das árvores. Com isso, os autores apontam a importância de desenvolver ferramentas para geração de modelos locais. Alguns estudos relatam que a grande variabilidade dos modelos alométricos existentes ocorre em virtude do pequeno tamanho das amostras das árvores utilizadas (CHAVE et al., 2014; PAUL et al., 2016).

Além desses modelos, também foram avaliados o uso do *Stepwise* e de algoritmos para construção de novos modelos, com o objetivo de estimar o estoque de carbono entre os compartimentos da árvore. Os resultados obtidos (Tabela 4) indicaram que, no geral, os modelos desenvolvidos utilizando diversas variáveis explicativas mostraram respostas muito diferentes. O RF com poucas variáveis, DAP, HT e densidade básica, apresentaram ajustes inferiores. O acréscimo de mais variáveis (RF<sub>3</sub>, RF<sub>4</sub> e RF<sub>5</sub>) demonstrou melhoras nos ajustes, em que juntamente com as demais técnicas (*Stepwise* e SARF) conseguiram explicar a variância dos dados ao apresentarem valores de  $R^2$  entre 65% e 99%. No entanto, nota-se que para alguns compartimentos os modelos gerados apresentaram altos valores de RMSE%. Esse resultado pode ser atribuído as diferentes fitofisionomias presentes na área, uma vez que os estoques de biomassa e carbono variam entre espécies e tipos florestais. Pois reflete as características morfológicas que diferem na maneira de usar a luz, nas relações de competição entre as árvores e das condições locais que, geralmente, levam a uma variação substancial nas estimativas (COLETTA et al., 2016; MARZILIANO et al., 2013; SANQUETTA; DALLA CORTE; SILVA, 2011). Dessa forma, Henry et al. (2011) evidenciam que o desenvolvimento de modelos alométricos com a estratificação por tipos ecológicos de florestas é uma maneira altamente eficaz para melhorar as estimativas.

Em relação ao uso do algoritmo RF foi observada a robustez do método que não foi muito afetado pela inclusão de muitas variáveis relacionadas entre si, e ao utilizar o mesmo conjunto de variáveis que o modelo de Spurr e Chave et al. (2005) apresentou resultados superiores. Wu et al. (2017) avaliaram a utilização do RF e da máquina vetor de suporte (SVM) para estimar a biomassa acima do solo usando imagens de Landsat em uma floresta na

China. Os autores identificaram que o RF obteve estimativas precisas e satisfatórias com  $R^2$  maior e RMSE baixo. Resultado semelhante foi encontrado por Wu et al. (2016) em que avaliaram diferentes abordagens de regressão, dentre eles o *Stepwise*, vizinho mais próximo, SVM e RF para estimar biomassa acima do solo. O algoritmo RF também apresentou resultados superiores de acordo com os valores de  $R^2$  e RMSE. Ao contrário do resultado obtido nesse trabalho, em que o uso do *Stepwise* mostrou melhores estimativas. Contudo, ressalta-se que essa técnica tem seu desempenho afetado por variáveis correlacionadas e ruidosas (ZIMMER; ANZANELLO, 2014).

Apesar disso, a utilização da técnica de seleção de variáveis feita pela meta-heurística SA conseguiu contribuir consideravelmente na redução do número de variáveis e no erro das estimativas. Esse comportamento era esperado visto que foi estabelecida uma condição multiobjectivo. Manimala, Selvi e Ahila (2011) avaliaram a utilização do algoritmo genético (AG) e *Simulated Annealing* (SA) como métodos de seleção de características e otimização dos parâmetros da máquina vetor de suporte (SVM). Os autores encontraram bons resultados para os dois algoritmos, contudo, ressaltam que o SA produz boas soluções em um curto espaço de tempo.

Sobre as variáveis que mais contribuem para estimar o estoque de carbono acima do solo, verifica-se maior contribuição do peso seco, volume, densidade básica, área de copa, área basal, diâmetro a 0%, diâmetro à altura do peito (DAP), altura e latitude. Na Tabela 5 constata-se que a maioria dessas variáveis apresenta altos valores de correlação com o estoque de carbono. O peso seco ou biomassa e o volume são diretamente relacionados ao estoque de carbono, visto que são utilizados para geração de fatores de conversão e expansão da biomassa. Magalhães e Seifert (2015) avaliaram as diferenças nas estimativas com o valor padrão para conteúdo de carbono (50%), usado pelo Painel Intergovernamental de Mudanças Climáticas. Segundo estes autores, pequenas diferenças nas estimativas foram encontradas. Entretanto, embora pequenas as diferenças, esses erros podem se propagar à medida que as estimativas fossem ampliadas, por exemplo, no nível de povoamento.

As variáveis DAP e altura são comumente utilizadas para estimar a biomassa e/ou estoque de carbono devido à sua alta relação (SEGURA; ACUÑA; ANDRADE, 2017; SANQUETTA et al., 2018; VARGAS-LARRETA et al., 2017), fato este também observado neste estudo. Além dessas, a densidade básica da madeira tem grande aplicação na explicação do estoque de carbono, uma vez que essa variável representa aspectos relacionados à estrutura da floresta, como a taxa de crescimento e o estado de sucessão da área (RIBEIRO et al.,

2011). De acordo com o estudo de Stas et al. (2017), o melhor modelo encontrado para estimar biomassa indicou a inclusão do DAP, altura e densidade. Os autores citam que, segundo Chave et al. (2005), essas variáveis possuem grande poder explicativo em virtude de refletirem os diferentes tamanhos entre as árvores e as espécies.

Dentre as variáveis selecionadas, a área seccional (g) e o diâmetro a 0% (D0) mostraram-se presentes para a maioria dos compartimentos analisados. O diâmetro a 0% também foi encontrado como preditor do estoque de biomassa nos estudos de Henry et al. (2011) e Ziemmer, Behling e Dalla Corte (2016). Sobre a área basal, Burrows et al. (2000), indicam que essa variável é um bom preditor de biomassa e carbono, visto que inclui o efeito do número e tamanho das árvores. Além disso, os autores ressaltam que a relação entre biomassa e área basal pode ser aplicada para auxiliar nas estimativas, pois essa variável pode ser rapidamente mensurada. Magalhães e Seifert (2015) afirmam que o estoque de biomassa, e consequentemente o estoque de carbono, é uma função da densidade do fuste e do DAP, logo também tem associação com a área basal. Segundo os autores as árvores que possuem maior proporção de área basal também possuem maior proporção de estoque de biomassa.

A área de copa (AC) explica variações no crescimento individual das árvores, pois o seu desenvolvimento é limitado pelo tamanho das copas e influenciado pela competição. Alterações no dossel das árvores causadas por intervenções e variações de fatores climáticos e edáficos, é geralmente seguido por um aumento ou redução do estoque de biomassa (COLETTA et al., 2016; KUYAH et al., 2012). Como exemplo, as árvores mais altas que se sobressaem no dossel da floresta têm seu crescimento reduzido em altura, enquanto o crescimento horizontalmente dos galhos aumenta. Nesse sentido, adicionar variáveis relacionadas à copa pode melhorar às estimativas de estoque de carbono ao capturar as variações de biomassa da árvore (GOODMAN; PHILLIPS; BAKER, 2014), como verificado neste estudo.

A latitude é uma variável geográfica que auxilia na explicação dos diferentes biomas e das características fitogeográficas. Além disso, é um dos elementos geográficos que mais influencia no clima (SCOLFORO et al., 2015; MELLO et al., 2013). De acordo com o estudo de Scolforo et al. (2015), foi encontrado que a latitude e a variação da distribuição do estoque de carbono da vegetação arbórea possuem forte associação, em alguns biomas no estado de Minas Gerais. Os autores explicam que à medida que a latitude diminui, maior é o estoque de carbono. Portanto, embora a maior parte das variáveis selecionadas neste estudo não sejam usualmente abordada nos inventários florestais, o ganho obtido com a inclusão dessas

variáveis nas estimativas deve ser avaliado, visto que apresentam uma associação do estoque de carbono das árvores e os processos ecofisiológicos e ambientais.

## 5. CONCLUSÃO

Diante dos diversos métodos avaliados, esses apresentaram diferentes resultados de ajuste. O modelo de Chave et al. (2005) apresentou resultado superior ao modelo de Spurr linear, o que mostra a importância da inclusão da densidade básica. O modelo desenvolvido pelo *Stepwise* demonstrou estatisticamente maior desempenho ao utilizar as variáveis puras. O método híbrido SARF sobressaiu-se em comparação ao RF na sua forma pura. Os métodos de seleção de variáveis, *Stepwise* e o SARF, proporcionaram bons ajustes, com  $R^2$  similares e menores erros para o *Stepwise*. Sendo assim, o SARF é uma alternativa ao método tradicional quando envolve um grande número de variáveis e observações. Em geral, as variáveis mais importantes para a modelagem do estoque de carbono em cada compartimento são o peso seco, volume, densidade básica, área de copa, área basal, diâmetro a 0%, diâmetro à altura do peito, altura e latitude, sendo essas puras, transformadas ou combinadas. O SARF aliado à função multiobjetivo consegue identificar as variáveis de maior importância, reduzir o erro das estimativas e mostrar a relevância da interação das variáveis biométricas, os fatores ecofisiológicos e ambientais na estimativa do estoque de Carbono. Além disso, denotam a aplicabilidade do *Simulated Annealing* na seleção de variáveis para modelagem do estoque de carbono acima do solo com o *Random Forest*.

## REFERENCIAS

BAGHERZADEH-KHIABANI, F. et al. A tutorial on variable selection for clinical prediction models: feature selection methods in data mining could improve the results. **Journal Of Clinical Epidemiology**, Pergamon, v. 71, p.76-85, 2016.

BAKER, D. J. et al. Achieving forest carbon information with higher certainty: A five-part plan. **Environmental Science & Policy**, New York, v. 13, n. 3, p.249-260, 2010.

BREIMAN, L. **Random Forests**. Machine learning, v. 45, n. 1, p. 5-32, 2001.

Burrows, WH et al. Relações alométricas e estimativas de biomassa comunitária para alguns eucaliptos dominantes em florestas do centro de Queensland. **Jornal Australiano de Botânica**, Melbourne, v. 48, n. 6, p. 707-714, 2000.

CARVALHO, L. G. et al. Clima. In: SCOLFORO, J. R. S.; CARVALHO, L. M. T.; OLIVEIRA, A. D. (Ed.). **Zoneamento ecológico -econômico do estado de Minas Gerais:** componentes geofísico e biótico. Lavras: UFLA, 2008. P. 89-101.

CHAVE, J. et al. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. **Oecologia**, Berlin, v. 145, n. 1, p.87-99, 22 jun. 2005.

CHAVE, J. et al. Improved allometric models to estimate the aboveground biomass of tropical trees. **Global Change Biology**, Blackwell, v. 20, n. 10, p.3177-3190, 2014.

CHEN, W. et al. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. **Catena**, Amsterdam, v. 151, p.147-160, 2017.

CHINEMBIRI, T. S. et al. The Precision of C Stock Estimation in the Ludhikola Watershed Using Model-Based and Design-Based Approaches. **Natural Resources Research**, New York, v. 22, n. 4, p.297-309, 2013.

COLETTA, V. et al. Effect of thinning on above-ground biomass accumulation in a Douglas-fir plantation in southern Italy. **Journal Of Forestry Research**, Berlin, v. 27, n. 6, p.1313-1320, 2016.

CORDEIRO, M. A. et al. Estimativa do volume de *Acacia mangium* utilizando técnicas de redes neurais artificiais e máquinas vetor de suporte. **Pesquisa Florestal Brasileira**, Colombo, v. 35, n. 83, p.255-261, 2015.

DRAKE, J. M.; RANDIN, C.; GUISAN, A. Modelling ecological niches with support vector machines. **Journal Of Applied Ecology**, Blackwell, v. 43, n. 3, p.424-432, 2006.

ELZHOV, T. V. et al. 2016. **minpack.lm**: Levenberg-Marquardt nonlinear least-squares. Disponível em: <<https://CRAN.R-project.org/package=minpack.lm>>. Acesso em 13 de nov. 2018.

FOX, et al. 2019. Car: Companion to Applied Regression. Disponível em: <<https://cran.r-project.org/web/packages/car/index.html>>. Acesso em 09 de fev. 2020.

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

GLEASON, C. J.; IM, J. Forest biomass estimation from airborne LiDAR data using machine learning approaches. **Remote Sensing Of Environment**, Toronto, v. 125, p.80-91, 2012.

GOMIDE, L. R.; ARCE, J. E.; SILVA, A. C. L. da. Comparação entre a meta-heurística simulated annealing e a programação linear inteira no agendamento da colheita florestal com restrições de adjacência. **Ciência Florestal**, Santa Maria, v. 23, n. 2, p.451-462, 2013.

GOODMAN, R. C.; PHILLIPS, O. L.; BAKER, T. R. The importance of crown dimensions to improve tropical tree biomass estimates. **Ecological Applications**, Tempe, v. 24, n. 4, p.680-698, 2014.

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

HAPFELMEIER, A.; ULM, K.. Variable selection by Random Forests using data with missing values. **Computational Statistics & Data Analysis**, New York, v. 80, p.129-139, 2014.

HENRY, M. et al. Estimating tree biomass of sub-Saharan African forests: a review of available allometric equations. **Silva Fennica**, Helsinki, v. 45, n. 3, p. 477-569, 2011.

KUYAH, S. et al. Crown area allometries for estimation of aboveground tree biomass in agricultural landscapes of western Kenya. **Agroforestry Systems**, Boston, v. 86, n. 2, p.267-277, 2012.

LIAW, A.; WIENER, M. 2002. **randomforest**: random forests for classification and regression. Disponível em: < <https://CRAN.R-project.org/package=randomForest>>. Acesso em 13 de nov. 2018.

MAFARJA, M. M.; MIRJALILI, S. Hybrid Whale Optimization Algorithm with simulated annealing for feature selection. **Neurocomputing**, New York, v. 260, p.302-312, 2017.

MAGALHÃES, T. M.; SEIFERT, T. Estimation of Tree Biomass, Carbon Stocks, and Error Propagation in Mecrusse Woodlands. **Open Journal Of Forestry**, Irvine, v. 05, n. 04, p.471-488, 2015.

MANIMALA, K.; SELVI, K.; AHILA, R. Hybrid soft computing techniques for feature selection and parameter optimization in power quality data mining. **Applied Soft Computing**, Netherlands, v. 11, n. 8, p.5485-5497, 2011.

MATTHEWS, H. D. et al. National contributions to observed global warming. **Environmental Research Letters**, England, v. 9, n. 1, p.1-9, 2014.

MARZILIANO, P. A. et al. Structural diversity and height growth models in urban forest plantations: A case-study in northern Italy. **Urban Forestry & Urban Greening**, Jena, v. 12, n. 2, p.246-254, 2013.

Marziliano, P. A., Menguzzato, G., Scuderi, A., Scalise, C., & Coletta, V. (2017). Biomass conversion and expansion factors in Douglas-fir stands of different planting density: variation according to individual growth and prediction equations. **Forest systems**, Madrid, v. 26, n. 1, p. 1-10, 2017.

MELLO, C. R. et al. Multivariate models for annual rainfall erosivity in Brazil. **Geoderma**, Amsterdam, v. 202-203, p.88-102, 2013.

MENDOZA-PONCE, A. et al. Identifying effects of land use cover changes and climate change on terrestrial ecosystems and carbon stocks in Mexico. **Global Environmental Change**, England, v. 53, p.12-23, 2018.

PAUL, K. I. et al. Testing the generality of above-ground biomass allometry across plant functional types at the continent scale. **Global Change Biology**, Blackwell, v. 22, n. 6, p.2106-2124, 2016.

PECHANEC, V. et al. Modelling of the carbon sequestration and its prediction under climate change. **Ecological Informatics**, Netherlands, v. 47, p.50-54, 2018.

PICARD, N.; SAINT-ANDRÉ, L.; HENRY, M. **Manual for building tree volume and biomass allometric equations: from field measurement to prediction**. Montpellier: FAO, 2012.

R Core Team, 2018. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

RIBEIRO, S. C. et al. Above- and belowground biomass in a Brazilian Cerrado. **Forest Ecology And Management**, Amsterdam, v. 262, n. 3, p.491-499, 2011.

SANQUETTA, C. R.; DALLA CORTE, A. P. ; SILVA, F. da. Biomass expansion factor and root-to-shoot ratio for Pinus in Brazil. **Carbon Balance and Management**, United Kingdom, v. 6, n. 6, p.1-8, 2011.

SANQUETTA, C. R. et al. Estoques de biomassa e carbono em povoamentos de acácia negra em diferentes idades no Rio Grande do Sul. **Scientia Forestalis**, Piracicaba, v. 42, n. 103, p. 361-370, 2014.

SANQUETTA, C. R. et al. Selection criteria for linear regression models to estimate individual tree biomasses in the Atlantic Rain Forest, Brazil. **Carbon Balance And Management**, United Kingdom, v. 13, n. 1, p.1-15, 2018.

SCOLFORO, H. F. et al. Spatial Distribution of Aboveground Carbon Stock of the Arboreal Vegetation in Brazilian Biomes of Savanna, Atlantic Forest and Semi-Arid Woodland. **Plos One**, San Francisco, v. 10, n. 6, p.1-20, 2015.

SEGURA, M. A.; ACUÑA, L. M.; ANDRADE, H. J. Allometric models to estimate aboveground biomass of small trees in wet tropical forests of colombian pacific area. **Revista Árvore**, Viçosa, v. 42, n. 2, p.1-8, 2018.

STAS, S. M. et al. Estimating the aboveground biomass in an old secondary forest on limestone in the Moluccas, Indonesia: Comparing locally developed versus existing allometric models. **Forest Ecology And Management**, Amsterdam, v. 389, p.27-34, 2017.

VAFAEI, S. et al. Improving Accuracy Estimation of Forest Aboveground Biomass Based on Incorporation of ALOS-2 PALSAR-2 and Sentinel-2A Imagery and Machine Learning: A Case Study of the Hyrcanian Forest Area (Iran). **Remote Sensing**, Beijing, v. 10, n. 2, p.1-21, 2018.

VAHEDI, A. A. Artificial neural network application in comparison with modeling allometric equations for predicting above-ground biomass in the Hyrcanian mixed-beech forests of Iran. **Biomass And Bioenergy**, United Kingdom, v. 88, p.66-76, 2016.

VARGAS-LARRETA, B. et al. Allometric Equations for Estimating Biomass and Carbon Stocks in the Temperate Forests of North-Western Mexico. **Forests**, Switzerland, v. 8, n. 8, p.1-20, 2017.

VICHARNAKORN, P. et al. Carbon Stock Assessment Using Remote Sensing and Forest Inventory Data in Savannakhet, Lao PDR. **Remote Sensing**, Beijing, v. 6, n. 6, p.5452-5479, 2014.

VIEIRA, G. C. et al. Prognoses of diameter and height of trees of eucalyptus using artificial intelligence. **Science Of The Total Environment**, New York, v. 619-620, p.1473-1481, 2018.

ZIMMER, J.; ANZANELLO, M. J. Um novo método para seleção de variáveis preditivas com base em índices de importância. **Production**, Porto Alegre, v. 24, n. 1, p.84-93, 2013.

ZIEMMER, J. K.; BEHLING, A.; DALLA CORTE, A. P. Quantificação da biomassa e dos teores de carbono de pteridófitas arborescentes em floresta ombrófila mista. **BIOFIX Scientific Journal**, Curitiba, v. 1, n. 1, 2016.

WERE, K. et al. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afrotropical landscape. **Ecological Indicators**, Amsterdam, v. 52, p.394-403, 2015.

WU, C. et al. Comparison of machine-learning methods for above-ground biomass estimation based on Landsat imagery. **Journal of Applied Remote Sensing**, Bellingham, v. 10, n. 3, p. 1-18, 2016.

WU, C. et al. Using nonparametric modeling approaches and remote sensing imagery to estimate ecological welfare forest biomass. **Journal Of Forestry Research**, Berlin, v. 29, n. 1, p.151-161, 2017.

**APÊNDICE A – Pseudocódigo do Algoritmo Genético 1**

Algoritmo Genético

1: **Begin**

2:     Tamanho da população  $V = X$

3:     Solução inicial/avaliação do *fitness* {Equação 7 – Artigo 1}

4:     Início do processo iterativo:

5:         **While** geração  $n \leq 100$ , **Do:**

6:             Função de Seleção

7:             Função de *Crossover*

8:             Função de Mutação

9:             Avaliação do *fitness* { Equação 7 – Artigo 1 }

10:         **Loop**

11:         retorna a melhor solução  $f^*$

12: **End**

**APÊNDICE B – Pseudocódigo do Algoritmo Genético 2**

Algoritmo Genético

1: **Begin**

2:     Tamanho da população  $V = X$

3:     Solução inicial/avaliação do *fitness* {Equação 8 – Artigo 1}

4:     Início do processo iterativo:

5:         **While** geração  $n \leq 100$ , **Do:**

6:             Função de Seleção

7:             Função de *Crossover*

8:             Função de Mutação

9:             Avaliação do *fitness* { Equação 8 – Artigo 1 }

10:         **Loop**

11:         retorna a melhor solução  $f^*$

12: **End**

**APÊNDICE C – Pseudocódigo do Algoritmo Genético 1 - Percentis**

Algoritmo Genético - Percentis

1: **Begin**

2:     Tamanho da população  $V = X$

3:     Solução inicial { Equação 3 e 4 – Artigo 1 }/avaliação do *fitness* {Equação 7 – Artigo 1 }

4:     Início do processo iterativo:

5:         **While** geração  $n \leq 100$ , **Do:**

6:             Função de Seleção

7:             Função de *Crossover*

8:             Função de Mutação

9:             Avaliação do *fitness* { Equação 7 – Artigo 1 }

10:         **Loop**

11:         retorna a melhor solução  $f^*$

12: **End**

**APÊNDICE D – Pseudocódigo do Algoritmo Genético 2 - Percentis**

Algoritmo Genético - Percentis

1: **Begin**

2:     Tamanho da população  $V = X$

3:     Solução inicial { Equação 3 e 4 – Artigo 1 }/avaliação do *fitness* {Equação 8 – Artigo 1 }

4:     Início do processo iterativo:

5:         **While** geração  $n \leq 100$ , **Do:**

6:             Função de Seleção

7:             Função de *Crossover*

8:             Função de Mutação

9:             Avaliação do *fitness* { Equação 8 – Artigo 1 }

10:         **Loop**

11:         retorna a melhor solução  $f^*$

12: **End**

## APÊNDICE E – Pseudocódigo do *Simulated Annealing* 1

### *Simulated Annealing*

```

1: Begin:
2:    $x^* \leftarrow x$  {melhor solução obtida até então} /avaliação da energia {Equação 7 – Artigo 1}
3:    $T_0 \leftarrow 50.000$ ; {temperatura inicial}
4:    $T_F \leftarrow 25$ ; {temperatura final}
5:    $T \leftarrow T_0$ ; {temperatura corrente}
6:   Início do processo iterativo:
7:     While  $T > 25$ , Do:
8:       Gerar um vizinho  $x' \in N(x)$  /avaliação da energia {Equação 7 – Artigo 1};
9:       if  $f(x') \leq f(x)$  then:
10:         $x \leftarrow x'$ ;
11:       else if
12:        gerar  $u \in \text{random}[0,1]$ ;
13:        se  $u < \exp\left[-\frac{f(x) - f(x')}{T}\right]$  then
14:           $x \leftarrow x'$ ;
15:        End
16:       End
17:       reduzir temperatura  $T$ 
18:     End
19:     retorna a melhor função  $f^*$ 
20: End

```

## APÊNDICE F – Pseudocódigo do *Simulated Annealing* 2

### *Simulated Annealing*

```

1: Begin:
2:    $x^* \leftarrow x$  { melhor solução obtida até então } /avaliação da energia {Equação 8 – Artigo 1}
3:    $T_0 \leftarrow 50.000$ ; {temperatura inicial}
4:    $T_F \leftarrow 25$ ; {temperatura final}
5:    $T \leftarrow T_0$ ; {temperatura corrente}
6:   Início do processo iterativo:
7:     While  $T > 25$ , Do:
8:       Gerar um vizinho  $s' \in N(x)$  /avaliação da energia {Equação 8 – Artigo 1};
9:       if  $f(x') \leq f(x)$  then:
10:         $x \leftarrow x'$ ;
11:       else if
12:        gerar  $u \in \text{random}[0,1]$ ;
13:        se  $u < \exp\left[-\frac{f(x) - f(x')}{T}\right]$  then
14:           $x \leftarrow x'$ ;
15:        End
16:       End
17:       reduzir temperatura  $T$ 
18:     End
19:     retorna a melhor função  $f^*$ 
20: End

```

## APÊNDICE G – Pseudocódigo do *Simulated Annealing 1 - Percentis*

### *Simulated Annealing - Percentis*

```

1: Begin:
2:    $x^* \leftarrow x$  { Equação 3 e 4 – Artigo 1 } / { melhor solução obtida até então } / avaliação da energia
   {Equação 7 – Artigo 1}
3:    $T_0 \leftarrow 50.000$ ; {temperatura inicial}
4:    $T_F \leftarrow 25$ ; {temperatura final}
5:    $T \leftarrow T_0$ ; {temperatura corrente}
6:   Início do processo iterativo:
7:     While  $T > 25$ , Do:
8:       Gerar um vizinho  $x' \in N(x)$  { Equação 3 e 4 – Artigo 1 } / avaliação da energia
       {Equação 7 – Artigo 1};
9:       if  $f(x') \leq f(x)$  then:
10:         $x \leftarrow x'$ ;
11:       else if
12:        gerar  $u \in \text{random}[0,1]$ ;
13:        se  $u < \exp\left[-\frac{f(x) - f(x')}{T}\right]$  then
14:           $x \leftarrow x'$ ;
15:        End
16:       End
17:       reduzir temperatura  $T$ 
18:     End
19:     retorna a melhor função  $f^*$ 
20: End

```

## APÊNDICE H – Pseudocódigo do *Simulated Annealing 2 - Percentis*

*Simulated Annealing - Percentis*

1: **Begin:**

2:  $x^* \leftarrow x$  { Equação 3 e 4 – Artigo 1 } / { melhor solução obtida até então } / avaliação da energia { Equação 8 – Artigo 1 }

3:  $T_0 \leftarrow 50.000$ ; { temperatura inicial }

4:  $T_F \leftarrow 25$ ; { temperatura final }

5:  $T_F \leftarrow T_0$ ; { temperatura corrente }

6: Início do processo iterativo:

7: **While**  $T_F > 25$ , **Do:**

8: Gerar um vizinho  $x' \in N(x)$  { Equação 3 e 4 – Artigo 1 } // avaliação da energia { Equação 8 – Artigo 1 };

9: **if**  $f(x') \leq f(x)$  **then:**

10:  $x \leftarrow x'$ ;

11: **else if**

12: gerar  $u \in \text{random}[0,1]$ ;

13: **se**  $u < \exp\left[-\frac{f(x) - f(x')}{T}\right]$  **then**

14:  $x \leftarrow x'$ ;

15: **End**

16: **End**

17: reduzir temperatura  $T_F$

18: **End**

19: retorna a melhor função  $f^*$

20: **End**

## APÊNDICE I – Pseudocódigo do *Simulated Annealing* e *Random Forest*

### *Simulated Annealing* - Percentis

```

1: Begin:
2:    $x^* \leftarrow x$  {melhor solução obtida até então} /avaliação da energia {Random Forest}
3:    $T_0 \leftarrow 100.000$ ; {temperatura inicial}
4:    $T_F \leftarrow 25$ ; {temperatura final}
5:    $T \leftarrow T_0$ ; {temperatura corrente}
6:   Início do processo iterativo:
7:     While  $T > T_F$ , Do:
8:       Gerar um vizinho  $x' \in N(x)$  /avaliação da energia { Random Forest };
9:       if  $f(x') \leq f(x)$  then:
10:         $x \leftarrow x'$ ;
11:       else if
12:        gerar  $u \in \text{random}[0,1]$ ;
13:        se  $u < \exp\left[-\frac{f(x) - f(x')}{T}\right]$  then
14:           $x \leftarrow x'$ ;
15:        End
16:       End
17:       reduzir temperatura  $T$ 
18:     End
19:     retorna a melhor função  $f^*$ 
20: End
21: Random Forest: predição.

```