**UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI**

**Programa de Pós-Graduação em Ciência Florestal**

**Leonardo Ippolito Rodrigues**

**Modelagem da Produção Florestal com Dados LiDAR Multitemporais**

**Diamantina**

**2025**

**Leonardo Ippolito Rodrigues**

**Modelagem da Produção Florestal com Dados LiDAR Multitemporais**

Dissertação apresentada ao programa de Pós-Graduação em Ciência Florestal da Universidade Federal dos Vales do Jequitinhonha e Mucuri, como requisito para obtenção do título de Mestre.

Orientador: [Eric Bastos Gorgens](mailto:eric.gorgens@ufvjm.edu.br)

**Diamantina**

**2025**

Catalogação na fonte - Sisbi/UFVJM

|  |  |
| --- | --- |
|  | Confeccionada pelo Sisbi/UFVJM |

**Leonardo Ippolito Rodrigues**

**Modelagem da Produção Florestal com Dados LiDAR Multitemporais**

Dissertação apresentada ao programa de Pós-Graduação em Ciência Florestal da Universidade Federal dos Vales do Jequitinhonha e Mucuri, como requisito para obtenção do título de Mestre.

Orientador: [Eric Bastos Gorgens](mailto:eric.gorgens@ufvjm.edu.br)

Data de aprovação \_\_\_\_ /\_\_\_\_/\_\_\_\_\_.

**Diamantina**

SOMENTE A ESTRUTURA DO CAPITULO 01 ESTÁ PRESENTE NESTE DOCUMENTO – ALGUMAS ETAPAS AINDA PRECISAM DE VALIDAÇÃO!

**CAPÍTULO 1**

**MODELOS PREDITIVOS**

**RESUMO**

A estimativa precisa do volume de madeira é fundamental para o planejamento e a gestão de povoamentos florestais. Nesse sentido, avanços tecnológicos têm possibilitado o uso de ferramentas mais eficientes e precisas para esse fim. Este estudo objetivou desenvolver e validar um modelo preditivo robusto para a estimativa do Volume Total com Casca (VTCC) em plantios de Eucalyptus sp., utilizando uma abordagem integrada de dados de LiDAR aerotransportado e variáveis auxiliares de cadastro. A metodologia envolveu a extração de métricas estruturais da nuvem de pontos LiDAR para 309 parcelas de inventário florestal, a seleção de variáveis preditivas por meio da técnica de Eliminação Recursiva de Atributos (RFE) e a otimização de um modelo Random Forest (RF). O modelo final, composto por cinco preditores (percentil 95 da altura, curtose da média cúbica, idade, regime de manejo e macrorregião) foi avaliado por validação cruzada. Os resultados demonstraram alta acurácia, com um coeficiente de determinação (R²) de 0,76 e uma raiz do erro quadrático médio (RMSE) de 42,06 m³.ha⁻¹. A combinação de métricas estruturais LiDAR com variáveis de cadastro mostrou-se crucial para capturar a complexidade dos povoamentos e maximizar o poder preditivo. O modelo desenvolvido representa uma ferramenta eficiente e acurada para o mapeamento de estoques de volume em larga escala, subsidiando a gestão florestal de precisão.

**Palavras-chave:** Inventário Florestal, LiDAR, Sensoriamento Remoto, Random Forest, Modelagem de Volume, *Eucalyptus*.

**ABSTRACT**

Accurate estimation of timber volume is fundamental for the planning and management of forest plantations. XXXX This study aimed to develop and validate a robust predictive model for estimating the Total Stem Volume (TSV) in Eucalyptus sp. plantations, using an integrated approach of airborne LiDAR data and ancillary stand information. The methodology involved the extraction of structural metrics from the LiDAR point cloud for 309 forest inventory plots, the selection of predictive variables through the Recursive Feature Elimination (RFE) technique, and the optimization of a Random Forest (RF) model. The final model, composed of five predictors (95th percentile of height, kurtosis of the cubic mean, age, management regime, and macro-region), was evaluated using cross-validation. The results demonstrated high accuracy, with a coefficient of determination (R²) of 0.76 and a root mean square error (RMSE) of 42.06 m³.ha⁻¹. The combination of LiDAR structural metrics with stand registry information proved crucial for capturing the complexity of the stands and maximizing predictive power. The developed model represents an efficient and accurate tool for large-scale mapping of timber stocks, supporting precision forest management.

**Keywords:** Forest Inventory, LiDAR, Remote Sensing, Random Forest, Volume Modeling, *Eucalyptus*.

**1 INTRODUÇÃO**

A gestão eficiente e sustentável de povoamentos florestais, especialmente de espécies de rápido crescimento como o Eucalyptus, depende fortemente da disponibilidade de informações precisas e atualizadas sobre o estoque de madeira. O inventário florestal tradicional, baseado em medições de campo — considerado o padrão-ouro para obtenção de dados em nível de parcela — ainda é a principal fonte de informações silviculturais. No entanto, esse método apresenta limitações quando se busca extrapolar resultados para grandes áreas. Os altos custos operacionais, a necessidade de intensa mão de obra e o tempo despendido na coleta de dados reduzem a densidade amostral, o que frequentemente resulta em estimativas de menor precisão e maior incerteza em escala de talhão ou propriedade (Poso, 1983; White et al., 2013).

Neste cenário, o sensoriamento remoto consolidou-se como uma ferramenta indispensável para otimizar os processos de inventário florestal. Dentre as tecnologias disponíveis, o Light Detection and Ranging (LiDAR) aerotransportado se destaca por sua capacidade singular de mapear a estrutura tridimensional da floresta com altíssima resolução e precisão (Wulder et al., 2008). Ao emitir pulsos de laser que penetram o dossel, o LiDAR registra múltiplos retornos que permitem não apenas determinar a altura das árvores, mas também caracterizar a distribuição vertical da vegetação, a densidade do dossel e a topografia do terreno subjacente (Hyyppä et al., 2008). Essas informações tridimensionais, processadas na forma de métricas estatísticas, apresentam forte correlação com atributos biofísicos da floresta, como volume, biomassa e área basal (Næsset, 2004; Zhao et al., 2018).

A abordagem metodológica mais estabelecida e operacionalmente utilizada para derivar esses atributos é a Abordagem Baseada em Área (Area-Based Approach - ABA), formalizada por Næsset (2002). Esse método de duas etapas consiste, primeiramente, em desenvolver modelos estatísticos que relacionam as métricas LiDAR com os dados de inventário medidos em campo em parcelas georreferenciadas. Na segunda etapa, esses modelos são aplicados a toda a extensão da área de estudo para gerar mapas contínuos (wall-to-wall) dos atributos de interesse. O sucesso dessa abordagem depende fundamentalmente da qualidade dos dados de campo, da precisão do georreferenciamento das parcelas e, crucialmente, da capacidade do modelo estatístico em capturar a complexa relação entre a estrutura da floresta e as métricas LiDAR (Gobakken and Næsset, 2009; Frazer et al., 2011b).

Apesar de sua eficácia comprovada, a implementação da ABA enfrenta desafios complexos que influenciam diretamente a acurácia dos resultados. Estudos comparativos, como o de Fassnacht et al. (2014), demonstraram que a escolha do algoritmo de predição e o processo de seleção de variáveis podem ser tão ou mais importantes para a precisão final do que a própria fonte de dados. Nesse contexto, algoritmos de aprendizado de máquina, como o Random Forest (RF) (Breiman, 2001), ganharam proeminência. O RF se destaca por sua flexibilidade em modelar relações não-lineares complexas, sua robustez a ruídos e outliers, e por não exigir premissas rígidas sobre a distribuição dos dados, superando frequentemente os modelos de regressão linear tradicionais em aplicações de sensoriamento remoto (Belgiu & Drăguț, 2016).

No contexto específico de plantios florestais de Eucalyptus no Brasil, caracterizados por uma alta produtividade e ciclos de rotação curtos, a necessidade de inventários precisos para o planejamento da colheita e da logística industrial é ainda mais premente. A variabilidade espacial, decorrente de diferentes materiais genéticos, regimes de manejo e condições de sítio, exige modelos que sejam capazes de incorporar essas fontes de variação. Portanto, a integração de métricas estruturais LiDAR com variáveis auxiliares de cadastro (como idade, manejo e localização geográfica) por meio de um algoritmo robusto como o Random Forest apresenta-se como uma estratégia promissora.

Diante do exposto, o objetivo principal deste trabalho é desenvolver, otimizar e validar um modelo preditivo robusto para a estimativa do volume total com casca (VTCC) em povoamentos de Eucalyptus urograndis.

**2 MATERIAL E MÉTODOS**

**2.1 Área de estudo**

A área de estudo situa-se na porção leste do estado de Minas Gerais, Brasil, abrangendo propriedades florestais em três macrorregiões distintas, designadas neste trabalho como Região 01, Região 02 e Região 03 (Figura 01). O clima varia entre as macrorregiões, de acordo com a classificação climática de Köppen-Geiger (Alvares et al., 2013): a Região 01 apresenta clima Cwa (Subtropical úmido com inverno seco), com precipitação média anual de ~1.100 mm e temperatura média anual de ~20 °C; a Região 02 apresenta Cwa com presença de zonas Cwb (Temperado oceânico com inverno seco), precipitação de ~1.250 mm e temperatura de ~19 °C; e a Região 03 caracteriza-se por Aw (Tropical com inverno seco) com áreas de transição para Cwa, precipitação de ~900 mm e temperatura de ~23 °C. Todas as regiões caracterizam-se por uma topografia acidentada, com relevo predominantemente ondulado a montanhoso e elevado grau de declividade.

O estudo foi conduzido em 200 talhões comerciais de Eucalyptus urophylla x Eucalyptus grandis, totalizando 6.403 hectares, manejados primariamente para a produção de celulose. Estes povoamentos abrangeram considerável variabilidade de sítio e manejo, incluindo: (i) 31 diferentes materiais genéticos; (ii) manejo silvicultural sob regimes de alto fuste e talhadia; (iii) diversos espaçamentos de plantio (2,5x3,0 m, 3,0x2,5 m, 3,0x3,0 m, 3,0x3,33 m, 3,33x3,0 m), sendo as configurações de 2,5x3,0 m e 3,0x2,5 m específicas da Região 03. No momento da coleta dos dados de inventário florestal, a idade dos povoamentos variava entre 36 e 168 meses.

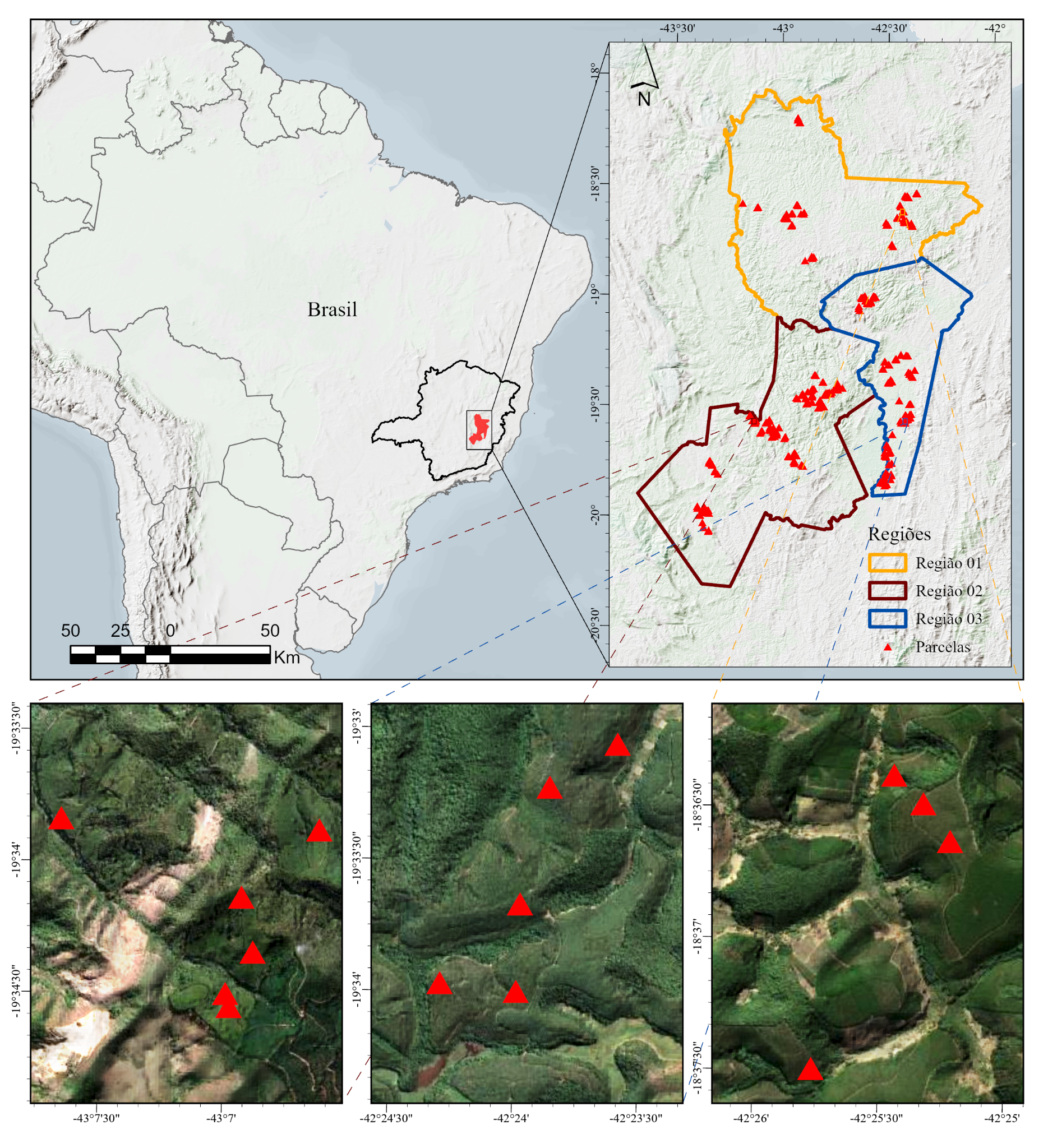


Figura 1. Mapa da localização da área de estudo, destacando as macrorregiões e a distribuição espacial das parcelas.

**2.2 Visão geral da metodologia**

O fluxo de trabalho metodológico completo adotado neste estudo é apresentado esquematicamente na Figura 02. Este fluxograma ilustra as principais etapas sequenciais, desde o pré-processamento dos dados de campo e LiDAR até o desenvolvimento, aplicação e validação do modelo preditivo de volume florestal. As etapas específicas são detalhadas nas subseções subsequentes deste capítulo.

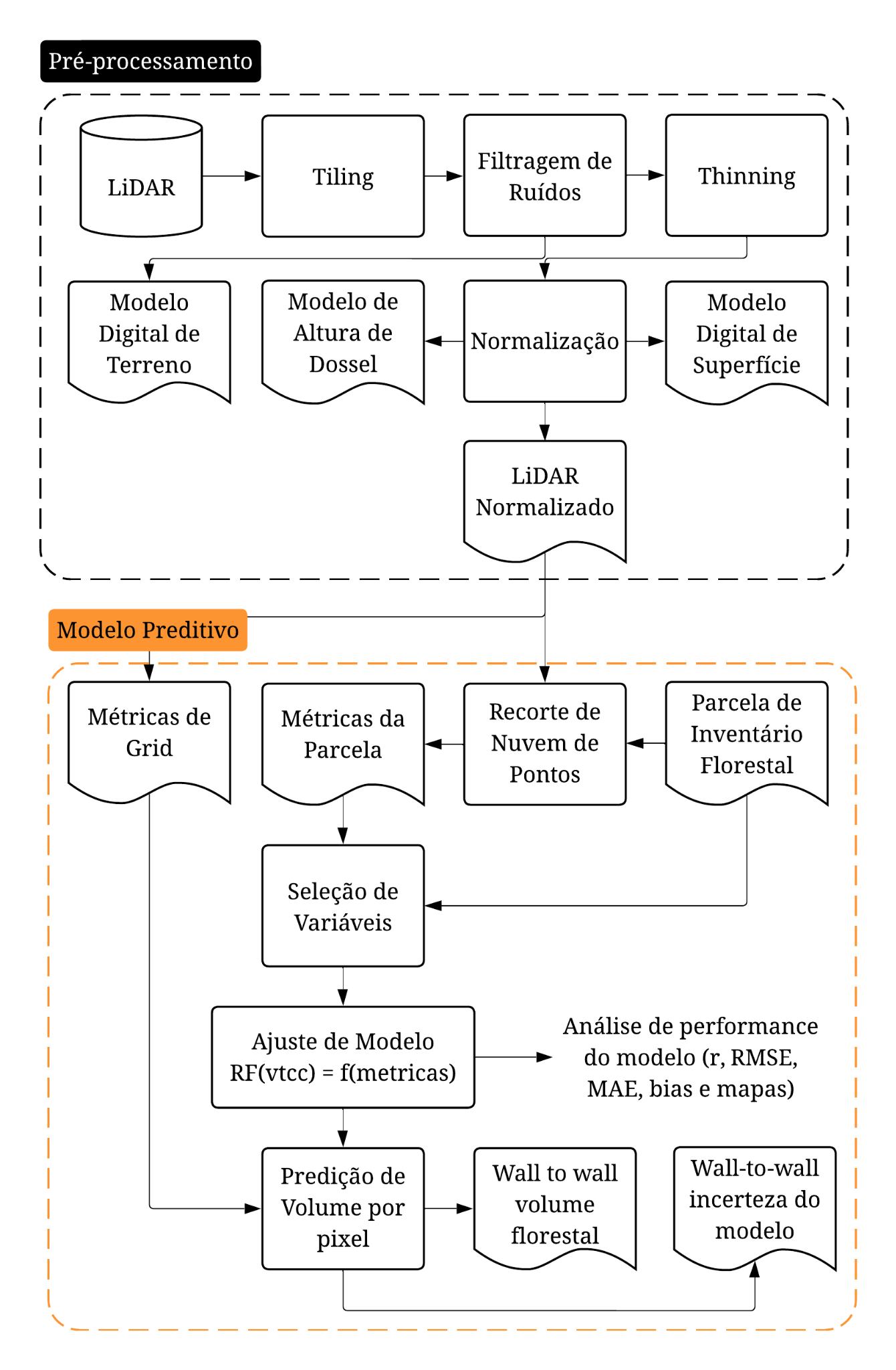


Figura 2. Fluxograma geral do processamento de dados LiDAR e desenvolvimento do modelo preditivo de volume florestal.

**2.3 Inventário florestal**

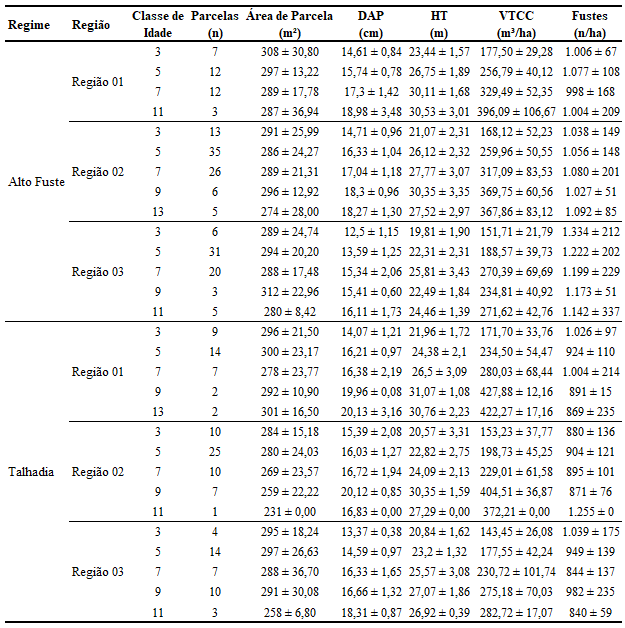
Foram utilizados dados de inventário florestal provenientes de 309 parcelas permanentes, de formato retangular com dimensões de 15x20 metros. As medições de campo foram realizadas em período concomitante ao sobrevoo LiDAR, garantindo compatibilidade temporal entre as fontes de dados.

Em cada parcela, o diâmetro a 1,3 m de altura (DAP) de todas as árvores foi mensurado em duas direções perpendiculares utilizando suta milimetrada, registrando-se a média aritmética como o valor final para cada árvore. A altura total (HT) foi obtida para todas as árvores utilizando clinômetro digital Haglöf Vertex (Haglöf Sweden AB, Långsele, Suécia).

O volume total individual com casca (VTCC), considerando um diâmetro mínimo comercial de 3 cm no topo, foi calculado para cada árvore empregando-se equações volumétricas do tipo Schumacher e Hall (1933), ajustadas localmente por macrorregião e regime de manejo, utilizando um histórico de dados de cubagem rigorosa pelo método de Smalian em seções de 1 metro (e.g., Husch et al., 2003).

Os vértices de cada parcela foram georreferenciados utilizando um sistema GNSS (Global Navigation Satellite System) geodésico [Marca e Modelo, e.g., Trimble R10, Leica GS18 T], operando em modo RTK (Real-Time Kinematic) com base local própria.

O volume total com casca por parcela (VTCC, m³ha⁻¹), obtido pela soma dos volumes individuais das árvores dentro da parcela e posterior expansão para hectare, foi considerado como o valor observado de referência para comparação e validação das estimativas geradas a partir dos dados LiDAR.



**2.4 Aquisição e pré-processamento de dados LiDAR**

Os dados LiDAR aerotransportados (ALS) foram adquiridos no período entre abril e outubro de 2022, utilizando o sistema LaserScan Optech ALTM Gemini embarcado na aeronave [Modelo da Aeronave]. O planejamento de voo estabeleceu uma altitude nominal de 600 metros acima do nível do solo e uma velocidade média da aeronave de 70 nós. A cobertura integral da área de estudo foi assegurada mediante múltiplas linhas de voo paralelas, com um espaçamento lateral de 305 metros e uma sobreposição de 30% entre faixas adjacentes. As configurações operacionais do sistema LiDAR durante a aquisição incluíram uma taxa de repetição de pulsos de 142 kHz e uma frequência de varredura do scanner de 30 Hz. O ângulo de varredura foi fixado em ±20° em relação ao nadir (campo de visão total de 40°), o que resultou em uma densidade nominal mínima de 5 pontos por metro quadrado (pts/m²) na área sobrevoada. Os dados brutos foram entregues georreferenciados no sistema de coordenadas SIRGAS 2000, projeção UTM Zona 23S.

O processamento subsequente da nuvem de pontos LiDAR foi conduzido utilizando o software FUSION/LDV (McGaughey, 2022). Inicialmente, para otimizar a manipulação e análise dos dados, a nuvem de pontos de cada área de voo (projeto) foi segmentada em um grid regular de até 64 partes (tiles), dividindo-se a extensão da área em até 8 seções nos eixos latitudinal e longitudinal, para otimizar o processamento. Posteriormente, procedeu-se à filtragem de ruído para remoção de pontos discrepantes. Esta etapa foi executada com o comando FilterData, configurado para identificar e eliminar pontos cuja elevação se desviasse em mais de 5 desvios padrão da média local, calculada dentro de uma janela de análise móvel de 100x100 metros.

A classificação dos pontos da nuvem em solo e não-solo foi realizada utilizando o algoritmo descrito por Axelsson (2000), baseado em modelos adaptativos de TIN. Com base nos pontos classificados como solo, foi interpolado um Modelo Digital de Terreno (MDT) com resolução espacial de 1 metro, representando a superfície topográfica da área de estudo. A normalização da nuvem de pontos foi então efetuada, consistindo na subtração das elevações do MDT das coordenadas Z de cada ponto da nuvem, resultando em alturas relativas ao nível do terreno.

Posteriormente, para homogeneizar a densidade de pontos em toda a área de estudo, particularmente reduzindo a densidade em áreas com contagens de retorno excessivas e garantindo uma representação mais uniforme, a nuvem normalizada foi submetida a um processo de redução de densidade de pontos (thinning), utilizando o comando ThinData do software FUSION/LDV. Este processo visou uma densidade alvo de 5 pontos por metro quadrado, aplicando-se o critério de seleção de [especificar o critério do ThinData, ex: ponto mais alto, aleatório] dentro de células de agregação de 50x50 metros.

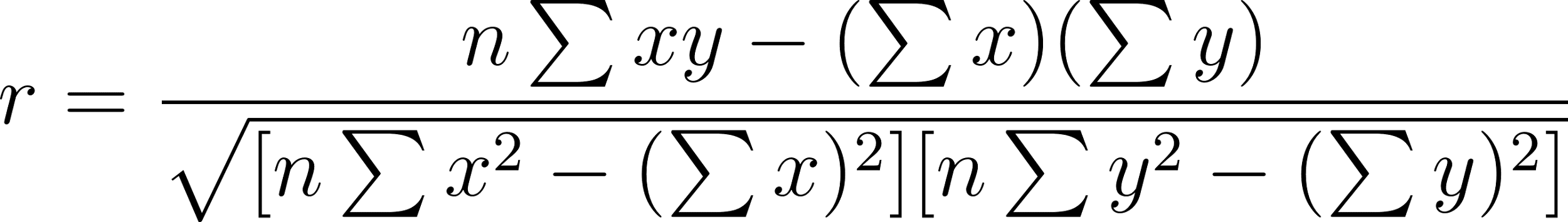
**2.5 Extração e seleção de variáveis preditivas LiDAR**

A partir da nuvem de pontos LiDAR normalizada e processada (Seção 2.4), foram extraídas diversas métricas descritivas da estrutura vertical da vegetação utilizando o software FUSION/LDV (McGaughey, 2022). As métricas geradas (detalhadas em [Apêndice X ou referenciar o manual do FUSION para GridMetrics]) abrangem estatísticas baseadas na distribuição das alturas dos retornos LiDAR (e.g., percentis, média, desvio padrão, curtose, assimetria), densidade de retornos em diferentes estratos verticais e, quando aplicável, estatísticas de intensidade. Para o cálculo das métricas derivadas do GridMetrics, foi empregado um limiar de altura (heightbreak) de 5 metros. Embora este estudo não tenha implementado uma otimização individual de limiares para cada métrica, reconhece-se que tal abordagem pode refinar o potencial preditivo das variáveis LiDAR, conforme demonstrado por Gorgens et al. (2017)

Para fins de predição espacial (wall-to-wall), as métricas LiDAR foram calculadas e rasterizadas com uma resolução espacial de 17x17 metros. Esta resolução foi escolhida para aproximar-se da área nominal das parcelas de inventário (~300 m²), permitindo uma compatibilização espacial entre os dados de campo e os preditores derivados do LiDAR, ao mesmo tempo em que considera possíveis variações na área efetiva da parcela devido ao ajuste de inclinação do terreno.

No entanto, para o desenvolvimento do modelo preditivo, as métricas foram associadas a cada parcela de inventário através do recorte direto da nuvem de pontos normalizada pelos polígonos exatos das parcelas. Desta forma, todos os retornos LiDAR contidos nos limites de cada unidade amostral foram utilizados para o cômputo das métricas específicas daquela parcela.

A etapa subsequente consistiu na seleção das variáveis preditivas mais relevantes para a modelagem do VTCC. Inicialmente, avaliou-se a força e a direção da associação linear entre cada métrica LiDAR candidata e o VTCC observado nas parcelas de inventário por meio do cálculo do coeficiente de correlação de Pearson (Equação X.X) (Pearson, 1895; Rodgers & Nicewander, 1988). Este coeficiente, denotado por r, varia de -1 a +1, onde valores próximos a +1 indicam uma forte correlação linear positiva, valores próximos a -1 indicam uma forte correlação linear negativa, e valores próximos a zero sugerem ausência de correlação linear.

[](https://www.codecogs.com/eqnedit.php?latex=r%3D%5Cfrac%7Bn%5Csum%20xy-(%5Csum%20x)(%5Csum%20y)%7D%7B%5Csqrt%7B%5Bn%5Csum%20x%5E2-(%5Csum%20x)%5E2%5D%5Bn%5Csum%20y%5E2-(%5Csum%20y)%5E2%5D%7D%7D#0)

Foram pré-selecionadas para a etapa seguinte as métricas LiDAR que apresentaram um coeficiente de correlação com o VTCC superior a um valor absoluto de 0,6 (|r| > 0,6), resultando em um subconjunto inicial de 14 variáveis candidatas.

Para estas variáveis, aplicou-se o método de Eliminação Recursiva de Atributos (Recursive Feature Elimination - RFE) combinado com validação cruzada para determinar o número ótimo de preditores e identificar as variáveis mais influentes. O RFE é um processo wrapper que ajusta um modelo de forma iterativa, classificando os atributos por importância e removendo recursivamente os menos importantes em cada passo (Guyon et al., 2002).

Para implementar o RFE e otimizar o número de atributos a serem selecionados, utilizou-se a biblioteca Scikit-learn (Pedregosa et al., 2011), tendo como estimador base um Random Forest Regressor. A otimização do número de atributos a serem mantidos pelo RFE foi realizada por meio de uma busca em grade (Grid Search CV), testando-se um intervalo de 1 até o número total de atributos pré-selecionados. Este processo empregou validação cruzada de 10 folds (KFold com n\_splits=10, shuffle=True e random\_state definido para reprodutibilidade) e utilizou a métrica de raiz do erro quadrático médio negativo (scoring='neg\_root\_mean\_squared\_error') para avaliar e selecionar o subconjunto de atributos que minimiza o erro de predição durante a busca em grade.

O GridSearchCV aplicado ao RFE identificou que o conjunto ótimo, o qual apresentou a melhor performance (menor RMSE) na validação cruzada, consistia em seis métricas LiDAR. Estas variáveis, que demonstraram maior poder preditivo combinado após o processo de eliminação recursiva, foram: Elev P90; Elev P95; Elev P80; Elev P99; Elev CURT mean CUBE; Elev P70 (Tabela XX). Consequentemente, estas seis métricas LiDAR foram selecionadas como o conjunto principal de preditores derivados da nuvem de pontos a serem considerados para as etapas subsequentes de desenvolvimento do modelo de estimação de VTCC.

**2.6 Inclusão de variáveis auxiliares de cadastro**

Para complementar as métricas estruturais derivadas do LiDAR, foram incorporadas ao conjunto de potenciais preditores três variáveis auxiliares, obtidas a partir dos registros do cadastro florestal da empresa. A idade do povoamento (em meses), o regime de manejo, classificado como "Alto Fuste" ou "Talhadia", e a macrorregião geográfica (Região 01, Região 02, Região 03).

Com a adição dessas variáveis de cadastro, o conjunto final de preditores candidatos para a modelagem do VTCC foi formado pelas seis métricas LiDAR selecionadas, pela idade do povoamento, por uma variável dummy representando o regime de manejo (onde 1=Talhadia, 0=Alto Fuste) e por duas variáveis dummy representando as macrorregiões (Região 01, Região 02 e Região 03), totalizando nove potenciais preditores.

**2.7 Desenvolvimento e avaliação de modelo preditivo**

Para a predição do VTCC a partir das variáveis derivadas do LiDAR e variáveis auxiliares, optou-se pelo algoritmo Random Forest Regressor (Breiman, 2001). O Random Forest é um método de aprendizado de máquina do tipo ensemble, que constroi múltiplas árvores de decisão durante o treinamento e combina suas predições, o que tende a reduzir o overfitting e melhorar a generalização do modelo.

Para determinar o subconjunto ótimo de variáveis que proporcionasse o modelo Random Forest mais acurado e parcimonioso, e para otimizar os hiperparâmetros do algoritmo simultaneamente, foi implementado um processo de avaliação exaustiva de todas as 511 combinações possíveis dessas 9 variáveis candidatas.

**2.7.1 Otimização de Hiperparâmetros**

Para cada combinação de variáveis, um modelo Random Forest Regressor foi ajustado para encontrar a combinação ótima de hiperparâmetros. A grade de hiperparâmetros explorada incluiu n\_estimators (número de árvores, variando de 10 a 50), max\_features (número de atributos a considerar em cada divisão), max\_depth (profundidade máxima das árvores), min\_samples\_split (número mínimo de amostras para dividir um nó), min\_samples\_leaf (número mínimo de amostras por folha) e bootstrap (amostragem com ou sem reposição) (Tabela Z.Z).

Quadro 1. Parâmetros de otimização do modelo random forest em grid.

|  |  |  |
| --- | --- | --- |
| **Hiperparâmetro** | **Descrição** | **Valores Testados / Intervalo** |
| n\_estimators | Número de árvores na floresta. | Inteiros linearmente espaçados de 10 a 50 (total de 20 valores). |
| max\_features | Número de atributos a considerar ao procurar a melhor divisão em um nó. | 'sqrt', 'log2', 1 (um atributo), 0.2, 0.3, 0.4, 0.5, 0.8 (proporção do total de atributos). |
| max\_depth | Profundidade máxima de cada árvore de decisão. | Inteiros linearmente espaçados de 10 a 150 (total de 10 valores), e None (nós expandidos até folhas puras ou min\_samples\_split). |
| min\_samples\_split | Número mínimo de amostras necessárias para dividir um nó interno. | 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 |
| min\_samples\_leaf | Número mínimo de amostras necessárias para formar um nó folha (terminal). | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 |
| bootstrap | Método de amostragem para construir as árvores. | True, False |

A busca aleatória foi configurada com 1000 iterações (n\_iter=1000) e utilizou validação cruzada de 10 folds (KFold com n\_splits=10, shuffle=True, random\_state=42), tendo como métrica de minimização do RMSE (scoring='neg\_root\_mean\_squared\_error’').

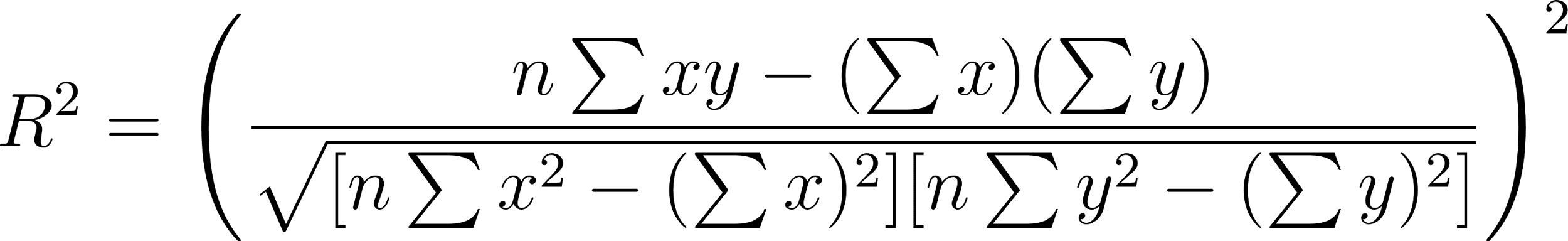
**2.7.2 Avaliação do modelo otimizado com validação cruzad**a

Após identificar os melhores hiperparâmetros para a combinação de variáveis em teste, um novo modelo Random Forest foi instanciado com esses parâmetros ótimos. A performance deste modelo otimizado foi então avaliada utilizando a técnica de predição por validação cruzada com 10 folds (KFold com n\_splits=10, shuffle=True, random\_state=42), aplicada a todo o conjunto de dados de 309 parcelas.

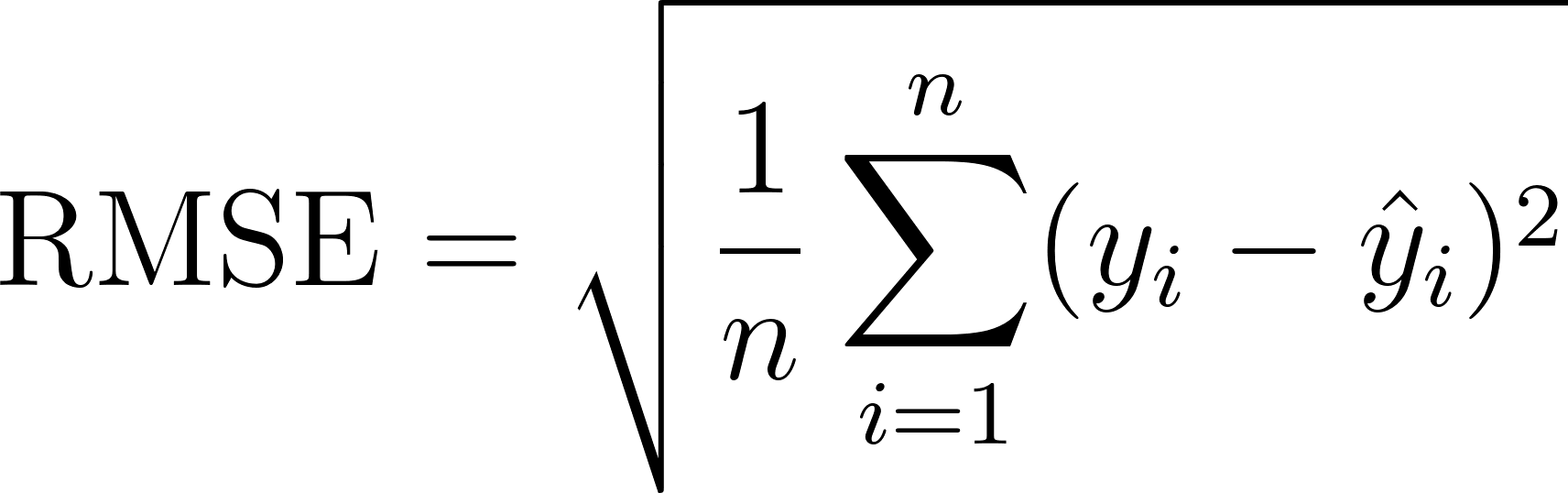
**2.7.3 Cálculo de Métricas de Performance**

Com base nas predições obtidas pela validação cruzada, foram calculadas as seguintes métricas estatísticas para cada combinação de variáveis:

1. **Coeficiente de Determinação (R²):** Esta métrica indica a proporção da variância total da variável dependente que é explicada pelo modelo preditivo. Valores de R² variam de 0 a 1, com valores mais próximos de 1 indicando um melhor ajuste do modelo aos dados (Draper & Smith, 1998).

[](https://www.codecogs.com/eqnedit.php?latex=R%5E2%3D%5Cleft(%5Cfrac%7Bn%5Csum%20xy-(%5Csum%20x)(%5Csum%20y)%7D%7B%5Csqrt%7B%5Bn%5Csum%20x%5E2-(%5Csum%20x)%5E2%5D%5Bn%5Csum%20y%5E2-(%5Csum%20y)%5E2%5D%7D%7D%5Cright)%5E2#0)

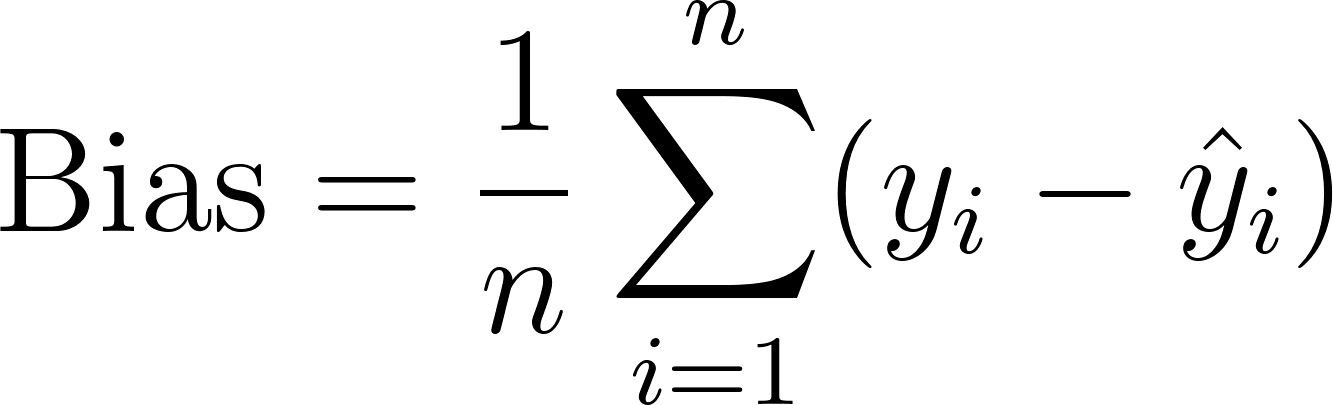
1. **Raiz do Erro Quadrático Médio (RMSE)**: O RMSE é uma medida da magnitude média dos erros de predição, expressa na mesma unidade da variável resposta. Ele penaliza erros maiores mais fortemente devido ao termo quadrático. Valores menores de RMSE indicam melhor performance do modelo (Chai & Draxler, 2014).

[](https://www.codecogs.com/eqnedit.php?latex=%5Ctext%7BRMSE%7D%3D%5Csqrt%7B%5Cfrac%7B1%7D%7Bn%7D%5Csum_%7Bi%3D1%7D%5En(y_i-%5Chat%7By%7D_i)%5E2%7D#0)

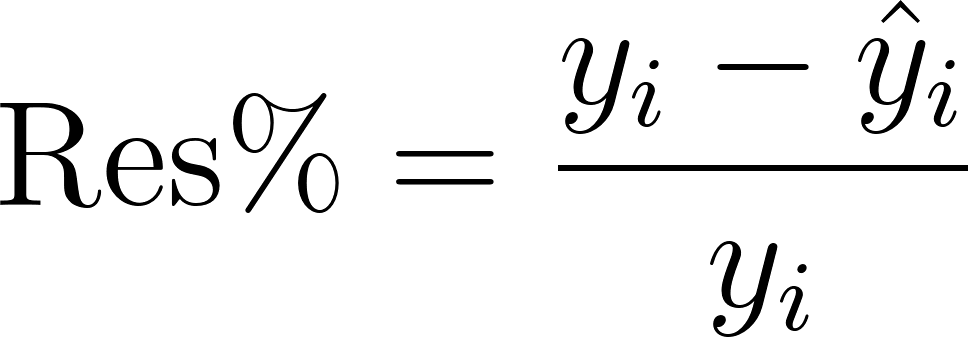
1. **Erro Absoluto Médio (MAE)**: O MAE também mede a magnitude média dos erros, mas sem dar peso desproporcional a erros grandes, pois utiliza o valor absoluto da diferença entre observado e predito. Assim como o RMSE, é expresso na unidade da variável resposta, e valores menores são preferíveis (Willmott & Matsuura, 2005).

[](https://www.codecogs.com/eqnedit.php?latex=%5Ctext%7BMAE%7D%3D%5Cfrac%7B1%7D%7Bn%7D%5Csum_%7Bi%3D1%7D%5En%7Cy_i-%5Chat%7By%7D_i%7C#0)

1. **Viés (Bias)**: O viés indica a tendência sistemática do modelo em superestimar ou subestimar os valores da variável resposta. Um modelo ideal apresenta um viés próximo de zero (Walther & Moore, 2005). Foi calculado como a média das diferenças entre os valores preditos e observados:

[](https://www.codecogs.com/eqnedit.php?latex=%5Ctext%7BBias%7D%3D%5Cfrac%7B1%7D%7Bn%7D%5Csum_%7Bi%3D1%7D%5En(y_i-%5Chat%7By%7D_i)#0)

1. **Resíduo Percentual Médio:** Para avaliar o erro relativo médio, calculou-se o resíduo percentual para cada observação, e então a média desses resíduos. Isso fornece uma indicação da magnitude do erro em relação ao valor observado, sendo útil para comparar a performance em diferentes escalas de VTCC.

[](https://www.codecogs.com/eqnedit.php?latex=%5Ctext%7BRes%5C%25%7D%3D%5Cfrac%7By_i-%5Chat%7By%7D_i%7D%7By_i%7D#0)

Este processo iterativo permitiu a comparação da performance de modelos Random Forest otimizados para múltiplos subconjuntos de variáveis. A seleção do conjunto de variáveis e dos hiperparâmetros que constituíram o modelo final, baseada na otimização do menor e RMSE e maior R² na validação cruzada, é detalhada na Seção 3 (Resultados).

**2.8 Aplicação e validação do modelo em áreas independentes**

Para avaliar a capacidade de generalização e a performance operacional do modelo Random Forest final (selecionado na Seção 2.7), este foi aplicado a um conjunto de validação independente composto por [Número] talhões. Estes talhões não foram utilizados em nenhuma etapa do treinamento ou otimização do modelo. Os dados de referência para estes talhões consistiram no volume de inventário florestal pré corte (IFPC, m³.ha⁻¹), obtido a partir dos registros de inventário da empresa. Notavelmente, o intervalo de tempo entre a medição de campo para obtenção do IFPC e o sobrevoo LiDAR variou de 0 a 12 meses entre os talhões, configurando um teste robusto para a aplicabilidade temporal do modelo.

Adicionalmente aos mapas de predição de volume, foram gerados mapas de incerteza para cada talhão. Esta incerteza foi quantificada, em nível de pixel, calculando-se o desvio padrão das predições individuais das árvores que compõem o ensemble do Random Forest (Breiman, 2001). Estes mapas de incerteza visam fornecer uma medida da confiança nas predições espaciais do modelo.

A validação final da performance do modelo nos talhões independentes foi realizada comparando-se o volume médio estimado por talhão com o volume de referência do talhão base IFPC. Para esta avaliação, foram empregadas as mesmas métricas estatísticas detalhadas na Seção 2.7.3: Coeficiente de Determinação (R²), Raiz do Erro Quadrático Médio (RMSE), Erro Absoluto Médio (MAE), Viés (Bias) e Resíduo Percentual Médio. A análise gráfica dos resíduos também foi conduzida. Os valores resultantes destas métricas e as análises gráficas são apresentados na Seção 3 (Resultados).

**3 RESULTADOS**

**3.1 Seleção do Modelo Preditivo Final e Desempenho na Validação Cruzada**

O processo iterativo de avaliação de todas as 511 combinações possíveis das nove variáveis candidatas, detalhado na Seção 2.7, permitiu identificar o subconjunto de preditores e a configuração de hiperparâmetros do modelo Random Forest que resultaram na melhor performance para a estimativa do Volume Total com Casca (VTCC) nas 309 parcelas de inventário. A seleção do modelo final baseou-se na otimização do coeficiente de determinação (R²) e da raiz do erro quadrático médio (RMSE), ambos obtidos a partir de validação cruzada de 10 folds.

A análise dos resultados indicou que o modelo mais acurado e parcimonioso foi composto por cinco variáveis preditivas: 'Elev P95' (percentil 95 das alturas LiDAR), 'Elev CURT mean CUBE' (curtose da média cúbica das alturas LiDAR), 'Idade (meses)' (idade do povoamento), 'regime de manejo' (alto fuste ou talhadia) e 'macroregião’' (região 01, região 02 e região 03).

A seleção dessas variáveis é ecologicamente e estatisticamente coerente. A métrica Elev P95 é um forte preditor, pois está diretamente relacionada à altura das árvores dominantes e codominantes, que é um dos principais determinantes do volume de um povoamento (Næsset, 2004; White et al., 2013). A Elev CURT mean CUBE descreve a forma da distribuição vertical da vegetação; valores elevados de curtose, especialmente após a elevação ao cubo que dá maior peso às copas superiores, podem indicar uma estrutura de dossel mais homogênea e concentrada verticalmente, o que influencia o volume. As variáveis auxiliares Idade, Regime de Manejo e Macroregião foram fundamentais para capturar a variabilidade na produtividade e na estrutura do povoamento que não é totalmente explicada apenas pelas métricas LiDAR, como diferenças de sítio, idade e práticas silviculturais (Fassnacht et al., 2014).

Este modelo final, após otimização dos hiperparâmetros (Tabela 3.1), alcançou um desempenho robusto na validação cruzada (Tabela 3.2). O coeficiente de determinação (R² CV) foi de 0,76, indicando que aproximadamente 76% da variabilidade do VTCC nas parcelas foi explicada pelo modelo. Este valor é consistente com os encontrados na literatura para estimativa de volume em plantios florestais (e.g., Næsset, 2002; Maltamo et al., 2004). A raiz do erro quadrático médio (RMSE CV) foi de 42,06 m³.ha⁻¹, enquanto o erro absoluto médio (MAE CV) registrou 32,16 m³.ha⁻¹. O modelo apresentou um leve viés (Bias CV) positivo de 0,73 m³.ha⁻¹ e um resíduo percentual médio de -3,21%, sugerindo uma ligeira tendência geral a subestimar o volume, o que pode ser considerado desprezível em um contexto operacional.

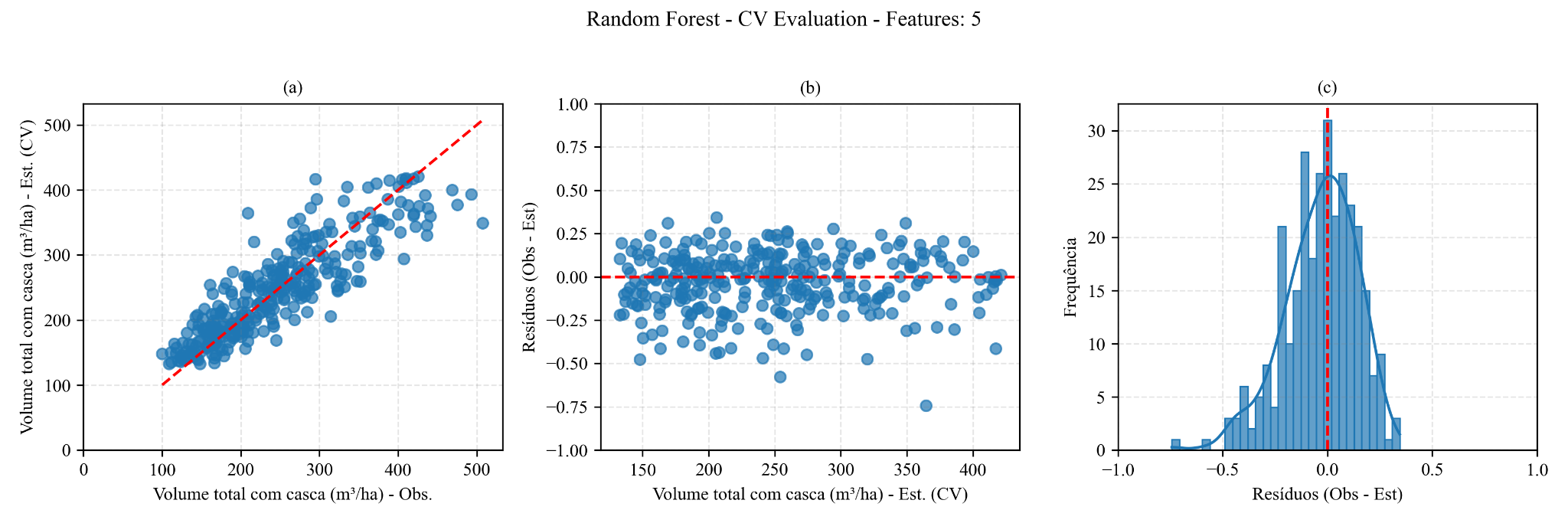


Figura 3. Gráficos de análise estatistica para o random forest em validação cruzada. (a) Gráfico de dispersão entre valores observados e estimados de VTCC (m³.ha¹); (b) Gráfico de dispersão resídual sobre os valores estimados; (c) Gráfico de frequência de resíduos do modelo ajustado.

**3.2 Aplicação do Modelo e Validação em Áreas Independentes**

* SUBITEM EM DESENVOLVIMENTO

**4 DISCUSSÃO**

O presente estudo demonstrou a eficácia da combinação de dados LiDAR e variáveis de cadastro para estimar o volume de madeira em plantios de eucalipto, utilizando a abordagem baseada em área (ABA) e o algoritmo de aprendizado de máquina Random Forest. O modelo final, composto por cinco preditores, alcançou um desempenho robusto com um coeficiente de determinação (R²) de 0,76 e um RMSE de 42,06 m³.ha⁻¹ na validação cruzada, resultados que se alinham e fortalecem os achados de estudos anteriores em ambientes florestais similares (e.g., Næsset, 2002; Maltamo et al., 2004).

A escolha do algoritmo Random Forest (RF) foi um fator chave para o sucesso da modelagem. O RF é conhecido por sua capacidade de lidar com relações não-lineares complexas entre preditores e a variável resposta, sua robustez a outliers e a não necessidade de premissas sobre a distribuição dos dados (Breiman, 2001). A escolha do método de predição pode ter um impacto substancial na acurácia, e o RF frequentemente se destaca como um dos algoritmos de melhor desempenho para estimativas de biomassa e volume florestal (Autores, ano). A principal desvantagem do RF, seu caráter de "caixa-preta", foi mitigada neste estudo por um rigoroso processo de seleção de variáveis (RFE), que garantiu a escolha de um conjunto parcimonioso e interpretável de preditores.

A importância da métrica de altura LiDAR (Elev P95) como principal preditor confirma o princípio fundamental de que a altura do dossel é o principal indicador do volume e da biomassa florestal (Means et al., 2000; Zhao et al., 2018). A inclusão da curtose (Elev CURT mean CUBE) como segunda variável estrutural mais importante sugere que, além da altura, a distribuição vertical da vegetação é crucial para refinar as estimativas de volume. No entanto, a contribuição significativa das variáveis auxiliares (Idade, Manejo e Macroregião) destaca uma limitação inerente aos dados LiDAR: eles capturam a estrutura física da floresta em um dado momento, mas não toda a variabilidade temporal ou espacial relacionada à produtividade do sítio, genética ou histórico de manejo. A integração dessas variáveis de cadastro foi, portanto, essencial para explicar parte da variância residual e melhorar a generalização do modelo, uma prática recomendada em inventários operacionais (White et al., 2013).

A acurácia alcançada, embora robusta, ainda apresenta 24% de variabilidade não explicada. Essa variância residual não se deve apenas às limitações do modelo, mas também a fontes de incerteza inerentes ao processo. Estas incluem: (i) erros na medição em campo dos atributos das árvores (DAP, altura); (ii) incertezas associadas às equações alométricas usadas para calcular o volume individual das árvores; e (iii) erros de posicionamento das parcelas em campo com GNSS, que podem levar a um desalinhamento entre os dados de campo e as métricas LiDAR (Gobakken and Næsset, 2009; Frazer et al., 2011b). Embora o uso de equipamentos geodésicos RTK minimize esse último erro, uma perfeita co-localização é raramente alcançada. A magnitude do RMSE obtido (42,06 m³.ha⁻¹) é, portanto, um reflexo combinado de todas essas fontes de incerteza.

Como limitação, deve-se notar que o modelo desenvolvido é específico para as condições da área de estudo (espécies, idade, condições de sítio). A transferência direta do modelo para outras regiões exigiria uma validação cuidadosa, sendo provável a necessidade de uma recalibração local (Fassnacht et al., 2014; Zhao et al., 2018). Para trabalhos futuros, sugere-se a exploração de dados LiDAR multitemporais para modelar diretamente o crescimento e a dinâmica do volume, uma fronteira de pesquisa que pode reduzir a dependência de medições de campo contínuas (Zhao et al., 2018).

Em conclusão, este estudo reafirma o potencial da tecnologia LiDAR, quando integrada a dados de cadastro e modelada com algoritmos avançados, como uma ferramenta poderosa e eficiente para o inventário florestal. A metodologia aqui apresentada oferece uma alternativa viável, permitindo o mapeamento de estoques de volume em larga escala.

**5 CONSIDERAÇÕES FINAIS**

Este estudo demonstrou com sucesso o desenvolvimento e a validação de uma metodologia robusta para a estimativa de volume em plantios de Eucalyptus urograndis, baseada na integração de dados LiDAR e variáveis de cadastro por meio de um modelo Random Forest otimizado. O modelo final, utilizando um conjunto parcimonioso de cinco preditores, alcançou alta acurácia (R² = 0,76; RMSE = 42,06 m³.ha⁻¹) e provou sua aplicabilidade operacional. A combinação de métricas estruturais LiDAR com informações de idade, manejo e localização geográfica foi crucial para capturar a complexidade dos povoamentos e maximizar o poder preditivo.

O uso do algoritmo Random Forest, aliado a um rigoroso processo de seleção de variáveis e otimização, mostrou-se fundamental para modelar as relações não-lineares e gerar estimativas confiáveis. A metodologia validada oferece uma ferramenta poderosa para o mapeamento de estoques de volume em larga escala, subsidiando o planejamento tático e estratégico da gestão florestal. Embora o modelo seja específico para as condições estudadas e sua transferibilidade exija validação local, este trabalho reafirma o papel transformador da tecnologia LiDAR na transição para uma silvicultura de precisão. Para o futuro, recomenda-se a exploração de dados multitemporais para a modelagem direta do crescimento, consolidando ainda mais o avanço do inventário florestal moderno.

**REFERÊNCIAS**

ALVARES, C. A., STAPE, J. L., SENTELHAS, P. C., GONÇALVES, J. L. DE M., & SPAROVEK, G. (2013). KÖPPEN’S CLIMATE CLASSIFICATION MAP FOR BRAZIL. METEOROLOGISCHE ZEITSCHRIFT, 22(6), 711–728.

AXELSSON, P. DEM GENERATION FROM LASER SCANNER DATA USING ADAPTIVE TIN MODELS. *INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY AND REMOTE SENSING*, V. 33, N. B4/1, P. 110–117, 2000.

BELGIU, M., & DRĂGUŢ, L. (2016). RANDOM FOREST IN REMOTE SENSING: A REVIEW OF APPLICATIONS AND FUTURE RESEARCH DIRECTIONS. ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING, 114, 24-31.

BREIMAN, L. (2001). RANDOM FORESTS. MACHINE LEARNING, 45(1), 5-32.

CHAI, T., & DRAXLER, R. R. (2014). ROOT MEAN SQUARE ERROR (RMSE) OR MEAN ABSOLUTE ERROR (MAE)? – ARGUMENTS AGAINST AVOIDING RMSE IN THE LITERATURE. GEOSCIENTIFIC MODEL DEVELOPMENT, 7(3), 1247-1250.

DRAPER, N. R., & SMITH, H. (1998). APPLIED REGRESSION ANALYSIS (3RD ED.). WILEY.

FASSNACHT, F. E., HARTIG, F., LATIFI, H., BERGER, C., HERNÁNDEZ, J., CORVALÁN, P., & KOCH, B. (2014). IMPORTANCE OF SAMPLE SIZE, DATA TYPE AND PREDICTION METHOD FOR REMOTE SENSING-BASED ESTIMATIONS OF ABOVEGROUND FOREST BIOMASS. REMOTE SENSING OF ENVIRONMENT, 154, 102–114.

FRAZER, G. W., MAGNUSSEN, S., WULDER, M. A., & NIEMANN, K. O. (2011). SIMULATED IMPACT OF SAMPLE PLOT SIZE AND CO-REGISTRATION ERROR ON THE ACCURACY AND UNCERTAINTY OF LIDAR-DERIVED ESTIMATES OF FOREST STAND BIOMASS. REMOTE SENSING OF ENVIRONMENT, 115(2), 636–649.

GOBAKKEN, T., & NÆSSET, E. (2009). ASSESSING EFFECTS OF POSITIONING ERRORS AND SAMPLE PLOT SIZE ON BIOPHYSICAL STAND PROPERTIES DERIVED FROM AIRBORNE LASER SCANNER DATA. CANADIAN JOURNAL OF FOREST RESEARCH, 39(5), 1036–1052.

GORGENS, E. B., VALBUENA, R., & RODRIGUEZ, L. C. E. (2017). A METHOD FOR OPTIMIZING HEIGHT THRESHOLD WHEN COMPUTING AIRBORNE LASER SCANNING METRICS. PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING, 83(5), 343–350.

GUYON, I.; WESTON, J.; BARNHILL, S.; VAPNIK, V. GENE SELECTION FOR CANCER CLASSIFICATION USING SUPPORT VECTOR MACHINES. MACHINE LEARNING, V. 46, N. 1-3, P. 389–422, 2002.

HUSCH, B., BEERS, T. W., & KERSHAW, J. A. (2003). FOREST MENSURATION (4TH ED.). JOHN WILEY & SONS.

HYYPPÄ, J., HYYPPÄ, H., LECKIE, D., GOUGEON, F., YU, X., & MALTAMO, M. (2008). REVIEW OF METHODS OF SMALL-FOOTPRINT AIRBORNE LASER SCANNING FOR EXTRACTING FOREST INVENTORY DATA IN BOREAL FORESTS. INTERNATIONAL JOURNAL OF REMOTE SENSING, 29(5), 1339–1366.

MALTAMO, M., EERIKÄINEN, K., PITKÄNEN, J., HYYPPÄ, J., & VEHMAS, M. (2004). ESTIMATION OF TIMBER VOLUME AND STEM DENSITY BASED ON SCANNING LASER ALTIMETRY AND EXPECTED TREE SIZE DISTRIBUTION FUNCTIONS. REMOTE SENSING OF ENVIRONMENT, 90(3), 319–330.

MCGAUGHEY, R. J. (2021). FUSION/LDV: SOFTWARE FOR LIDAR DATA ANALYSIS AND VISUALIZATION. USDA FOREST SERVICE, PACIFIC NORTHWEST RESEARCH STATION.

MEANS, J. E., ACKER, S. A., HARDING, D. J., BLAIR, J. B., LEFSKY, M. A., COHEN, W. B., ... & DUNNE, T. A. (2000). USE OF LARGE-FOOTPRINT SCANNING AIRBORNE LIDAR TO ESTIMATE FOREST STAND CHARACTERISTICS IN THE WESTERN CASCADES OF OREGON. REMOTE SENSING OF ENVIRONMENT, 71(1), 298-308.

MEANS, J. E., ACKER, S. A., HARDING, D. J., BLAIR, J. B., LEFSKY, M. A., COHEN, W. B., ... & DUNNE, T. A. (2000). USE OF LARGE-FOOTPRINT SCANNING AIRBORNE LIDAR TO ESTIMATE FOREST STAND CHARACTERISTICS IN THE WESTERN CASCADES OF OREGON. REMOTE SENSING OF ENVIRONMENT, 71(1), 298–308.

NÆSSET, E. (2002). PREDICTING FOREST STAND CHARACTERISTICS WITH AIRBORNE SCANNING LASER USING A PRACTICAL TWO-STAGE PROCEDURE AND FIELD DATA. REMOTE SENSING OF ENVIRONMENT, 80(1), 88-99.

NÆSSET, E. (2004). PRACTICAL LARGE-SCALE FOREST STAND INVENTORY USING A SMALL-FOOTPRINT AIRBORNE SCANNING LASER. SCANDINAVIAN JOURNAL OF FOREST RESEARCH, 19(2), 164–179.

PEARSON, K. (1895). NOTE ON REGRESSION AND INHERITANCE IN THE CASE OF TWO PARENTS. PROCEEDINGS OF THE ROYAL SOCIETY OF LONDON, 58, 240–242.

PEDREGOSA, F. ET AL. SCIKIT-LEARN: MACHINE LEARNING IN PYTHON. JOURNAL OF MACHINE LEARNING RESEARCH, V. 12, P. 2825–2830, OUT. 2011.

POSO, S. (1983). BASIC FEATURES OF FOREST INVENTORY BY COMPARTMENTS. SILVA FENNICA, 17(3), 313–349.

RODGERS, J. L., & NICEWANDER, W. A. (1988). THIRTEEN WAYS TO LOOK AT THE CORRELATION COEFFICIENT. THE AMERICAN STATISTICIAN, 42(1), 59–66.

SCHUMACHER, F. X., & HALL, F. D. S. (1933). LOGARITHMIC EXPRESSION OF TIMBER-TREE VOLUME. JOURNAL OF AGRICULTURAL RESEARCH, 47(9), 719–734.

SCHUMACHER, F. X., & HALL, F. D. S. (1933). LOGARITHMIC EXPRESSION OF TIMBER-TREE VOLUME. JOURNAL OF AGRICULTURAL RESEARCH, 47(9), 719–734.

WALTHER, B. A., & MOORE, J. L. (2005). THE CONCEPTS OF BIAS, PRECISION AND ACCURACY, AND THEIR USE IN TESTING THE PERFORMANCE OF SPECIES RICHNESS ESTIMATORS, WITH A LITERATURE REVIEW OF ESTIMATOR PERFORMANCE. ECOGRAPHY, 28(6), 815-829.

WHITE, J. C., WULDER, M. A., VARHOLA, A., VASTARANTA, M., COOPS, N. C., COOK, B. D., ... & WOODS, M. (2013). A BEST PRACTICES GUIDE FOR GENERATING FOREST INVENTORY ATTRIBUTES FROM AIRBORNE LASER SCANNING DATA USING AN AREA-BASED APPROACH. CANADIAN WOOD FIBRE CENTRE, INFORMATION REPORT FI-X-010. NATURAL RESOURCES CANADA.

WILLMOTT, C. J., & MATSUURA, K. (2005). ADVANTAGES OF THE MEAN ABSOLUTE ERROR (MAE) OVER THE ROOT MEAN SQUARE ERROR (RMSE) IN MODEL PERFORMANCE EVALUATION. CLIMATE RESEARCH, 30(1), 79-82.

WULDER, M. A., BATER, C. W., COOPS, N. C., HILKER, T., & WHITE, J. C. (2008). THE ROLE OF LIDAR IN SUSTAINABLE FOREST MANAGEMENT. THE FORESTRY CHRONICLE, 84(6), 807–826.

ZHAO, K., SUAREZ, J. C., GARCÍA, M., HU, T., WANG, C., & LONDO, A. (2018). UTILITY OF MULTITEMPORAL LIDAR FOR FOREST AND CARBON MONITORING: TREE GROWTH, BIOMASS DYNAMICS, AND CARBON FLUX. REMOTE SENSING OF ENVIRONMENT, 204, 883–897.

ZIMBLE, D. A., EVANS, D. L., CARLSON, G. C., PARKER, R. C., GRADO, S. C., & GERARD, P. D. (2003). CHARACTERIZING VERTICAL FOREST STRUCTURE USING SMALL-FOOTPRINT AIRBORNE LIDAR. REMOTE SENSING OF ENVIRONMENT, 87(2-3), 171-182.