UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS RURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA FLORESTAL

Mateus Sabadi Schuh

MODELOS PREDITIVOS DE BIOMASSA EM FLORESTA AMAZÔNICA A PARTIR DE DADOS LIDAR

Santa Maria, RS
2019

Mateus Sabadi Schuh

MODELOS PREDITIVOS DE BIOMASSA EM FLORESTA AMAZÔNICA A PARTIR DE DADOS LIDAR

Dissertação apresentada ao curso de Mestrado do Programa de Pós-Graduação em Engenharia Florestal, área de Concentração em Manejo Florestal, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do grau de **Mestre em Engenharia Florestal**.

Orientador: Prof. Dr. Rudiney Soares Pereira

Santa Maria, RS
2019

Schuh, Mateus
Modelos preditivos de biomassa em floresta amazônica a partir de dados LiDAR / Mateus Schuh.- 2019.
81 p.; 30 cm

Orientador: Rudiney Soares Pereira
Dissertação (mestrado) - Universidade Federal de Santa Maria, Centro de Ciências Rurais, Programa de Pós Graduação em Engenharia Florestal, RS, 2019

1. Mensuração Florestal 2. Inteligência Artificial 3. Amazônia 4. Sensoriamento Remoto I. Soares Pereira, Rudiney II. Título.

Sistema de geração automática de ficha catalográfica da UFSM. Dados fornecidos pelo autor(a). Sob supervisão da Direção da Divisão de Processos Técnicos da Biblioteca Central. Bibliotecária responsável Paula Schoenfeldt Patta CRB 10/1728.

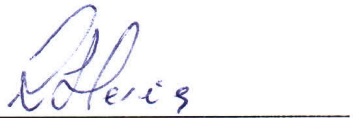
E-mail: mateuschuh@gmail.com

Mateus Sabadi Schuh

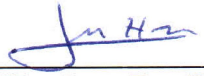
MODELOS PREDITIVOS DE BIOMASSA NA FLORESTA AMAZÔNICA A PARTIR DE DADOS LIDAR

Dissertação apresentada ao curso de Mestrado do Programa de Pós-Graduação em Engenharia Florestal, área de Concentração em Manejo Florestal, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do grau de **Mestre em Engenharia Florestal**.

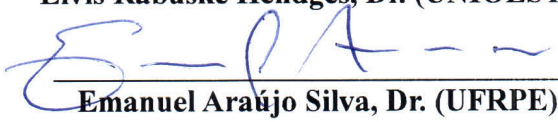
Aprovado em 30 de janeiro de 2019:



Rudiney Soares Pereira, Dr. (UFSM)
(Presidente/Orientador)



Elvis Rabuske Hendges, Dr. (UNIOESTE)



Emanuel Araújo Silva, Dr. (UFRPE)

Santa Maria, RS
2019

AGRADECIMENTOS

À minha família, em especial meus pais Luiz e Marlei, pelo amor e dedicação incondicional de sempre. Agradeço pelos ensinamentos e incentivo na busca de aperfeiçoamento pessoal e intelectual;

Ao meu orientador, Prof. Dr. Rudiney Soares Pereira, a quem dirijo respeito e admiração como profissional e ser humano. Agradeço pela oportunidade e confiança ao reabrir as portas do Laboratório de Sensoriamento Remoto, viabilizando meu ingresso na Pós-Graduação;

Aos meus amigos e colegas de LABSERE: Alessandra Marasciulo, Bruna Simões, Dionatas Honnef, Elisiane Alba, Eliziane Mello, Fábio Batista, Helena Oliveira, José Augusto Spiazzi Favarin, Juliana Marchesan, Matheus Frigo, Matheus Ziembowicz, Robson Righi, Rodrigo Carvalho e Tiago Luis Badin. Mesmo para os quais o tempo de convivência foi menor nesses últimos dois anos, agradeço pela participação direta na pesquisa ou indireta ao propiciarem um ambiente agradável e construtivo, tornando mais leve a rotina do trabalho;

Aos membros da banca examinadora, Elvis Rabuske Hendges e Emanuel Araújo Silva, pela disponibilidade e tempo dispendido para enriquecimento do trabalho;

À Universidade Federal de Santa Maria, ao Programa de Pós-Graduação em Engenharia Florestal, pela estrutura física e de corpo docente, componente basilar para minha formação acadêmica;

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, pela bolsa de estudos disponibilizada;

Enfim, à todas as pessoas que de alguma forma contribuíram para a realização do trabalho.

“Essentially, all models are wrong,
but some are useful”

(George E. P. Box)

RESUMO

MODELOS PREDITIVOS DE BIOMASSA EM FLORESTA AMAZÔNICA A PARTIR DE DADOS LIDAR

AUTOR: Mateus Sabadi Schuh
ORIENTADOR: Rudiney Soares Pereira

Dados de sensores remotos LiDAR (*Light Detection and Ranging*), combinados com técnicas de aprendizado de máquina tem apresentado grande potencial para a modelagem de atributos florestais em larga escala. Nesse contexto, o trabalho tem como objetivo avaliar a aplicação de técnicas de aprendizado de máquina na construção de modelos que relacionam métricas LiDAR e dados de inventário florestal na predição de biomassa em floresta tropical. Inicialmente, foi computada a biomassa acima do solo por meio de uma equação alométrica ajustada, fazendo uso de variáveis biométricas inventariadas em 85 unidades amostrais na Fazenda Cauaxi, município de Paragominas/PA. A biomassa das parcelas (variável de interesse) foi relacionada com 87 métricas LiDAR (variáveis explicativas), obtidas via processamento das nuvens de pontos LiDAR. Essa base de dados foi dividida aleatoriamente em 70% para ajuste dos modelos e 30% destinados à validação. Comparou-se o desempenho preditivo de três diferentes técnicas de aprendizado de máquina (*Random Forest* - RF, *Support Vector Machine* - SVM e *Artificial Neural Network* - ANN) frente a técnica *Generalized Linear Model* - GLM, tradicionalmente empregada em estimativas não paramétricas. Os resultados indicaram que as informações derivadas do levantamento LiDAR aerotransportado mostraram-se eficientes e perfeitamente aplicáveis ao processo de modelagem da biomassa em ambiente tropical. À exceção do modelo RF, com R^2 de 0,60, os modelos de aprendizado de máquina obtiveram melhor desempenho na etapa de treinamento. O valor de 0,99 para o R^2 e o desempenho superior nos demais indicadores da qualidade de ajuste (RMSE, Syx, BIAS e DM), conferiram ao modelo ANN a condição de melhor adequação aos dados de treino. Já na etapa de validação, os modelos GLM e RF que haviam apresentado os piores indicadores em relação ao ajuste, mostraram desempenho superior, enquanto que as estimativas ANN apresentaram a maior distorção. De modo geral, a correlação de *Spearman* entre os valores estimados e observados apresentou comportamento inversamente proporcional ao grau de ajuste dos modelos na etapa de treinamento, variando de 0,57 a 0,87 para os modelos ANN e GLM respectivamente. A despeito do ajuste inferior do modelo RF e da menor capacidade de generalização dos modelos ANN e SVM, a estatística *Wilcoxon Rank Sum Test* não detectou diferença significativa entre os valores de biomassa observados e preditos pelos diferentes modelos. Dessa forma, foi possível observar que os algoritmos de aprendizado de máquina conseguiram detectar e reproduzir bem a estrutura não paramétrica dos dados e fazer frente a regressão generalizada, sem a necessidade da aplicação de técnicas de redução da dimensionalidade dos dados, o que conferiu mais agilidade ao processo de modelagem.

Palavras-chave: Mensuração Florestal. Inteligência Artificial. Amazônia. Sensoriamento Remoto.

ABSTRACT

PREDICTIVE BIOMASS MODELS IN AMAZON RAINFOREST BY LIDAR DATA

AUTHOR: Mateus Sabadi Schuh
ADVISOR: Rudiney Soares Pereira

LiDAR (Light Detection and Ranging) remote sensing data combined with machine learning techniques has presented great potential for modeling large-scale forest atributes. In this context, the work aims to evaluate the application of machine learning techniques in the construction of models that relate LiDAR metrics and forest inventory data in the prediction of biomass in tropical forest. Initially, above-ground biomass was computed by an adjusted allometric equation, using biometric variables inventoried in 85 sample units at Fazenda Cauaxi, in the municipality of Paragominas / PA. The biomass of the plots (variable of interest) was related to 87 LiDAR metrics (explanatory variables), obtained by processing the LiDAR points clouds. This database was randomly divided into 70% for model adjustment and 30% for validation. The predictive performance of three different machine learning techniques (Random Forest - RF, Support Vector Machine - SVM and Artificial Neural Networks - ANN) was compared to a Generalized Linear Model (GLM) technique, traditionally used in non-parametric estimations. The results indicated that the information derived from the airborne LiDAR survey proved to be efficient and perfectly applicable to the modeling process of biomass in a tropical environment. With the exception of the RF model, with R^2 of 0.60, the machine learning models obtained better performance in the training stage. The value of 0.99 for the R^2 and the superior performance in the other adjustment quality indicators (RMSE, Syx, BIAS and DM), gave the ANN model the condition of better adaptation to the training data. In the validation stage, the GLM and RF models that presented the worst indicators in relation to the adjustment, showed superior performance, while the ANN estimates showed the greatest distortion. In general, the Spearman correlation between the estimated and observed values presented a behavior inversely proportional to the degree of adjustment of the models in the training stage, varying from 0.57 to 0.87 for the ANN and GLM models respectively. In spite of the lower adjustment of the RF model and the lower generalization capacity of the ANN and SVM models, the Wilcoxon Rank Sum Test did not detect a significant difference between the biomass values observed and predicted by the different models. In this way, it was possible to observe that the machine learning algorithms were able to detect and reproduce well the non-parametric data structure and to cope with generalized regression, without the need for data dimensionality reduction techniques, which gave more agility to the modeling process.

Keywords: Forest Measurement. Artificial Intelligence. Amazon. Remote Sensing.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE EQUAÇÕES

LISTA DE ABREVIATURAS E SIGLAS

ACP	Análise de Componentes Principais
AGB	*Above-Ground Biomass*
ALS	*Airbourne Laser Scanner*
ANN	*Artificial Neural Network*
BIAS	Viés
COP21	*21st Conference of the Parties*
CP	Componente Principal
CSV	*Comma-separated values*
DAP	Diâmetro à altura do peito
DM	Desvio médio
EIR	Exploração de Impacto Reduzido
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
GLM	*General Linear Model*
GPS	*Global Position System*
IMU	*Inertial Measurement Unit*
LASER	*Light Amplification by Stimulated Emission of Radiation*
LiDAR	*Light Detection and Ranging*
MDT	Modelo Digital do Terreno
Mg	Megagrama (10^6 gramas, equivalente a uma tonelada)
ML	*Machine Learning*
R^2	Coeficiente de determinação
REDD	*Reducing Emissions from Deforestation and Forest Degradation*
RF	*Random Forest*
RMSE	Raiz do erro médio quadrático
SVM	*Support Vector Machine*
Syx	Erro padrão da estimativa
TSL	*Terrestrial Laser Scanning*
UA	Unidade amostral

SUMÁRIO

1 INTRODUÇÃO

O bioma amazônico representa um dos maiores remanescentes de floresta tropical do planeta. Presente em nove países da América do Sul, engloba uma cobertura florestal remanescente de aproximadamente 6 milhões de km², destes, 67% em território brasileiro (SANTOS et al., 2007). No Brasil, a floresta amazônica perdeu 752 mil km^2, cerca de 20% dos 5 milhões de km² da cobertura original (INPE, 2016). Ainda que no acumulado anual as taxas de desmatamento para esse bioma tenham apresentado uma redução de 72% entre 2004 e 2018, no último ano registrou-se um aumento de 14% (INPE, 2018).

Além do desmatamento, a extração seletiva de madeira tem crescido significativamente como uma forma de uso da terra na Amazônia (ASNER et al., 2004). A área sob exploração seletiva foi semelhante à área desmatada no início deste século no Brasil, o que leva a entender que a exploração seletiva contribui significativamente para a dinâmica dos fluxos de carbono na Amazônia brasileira, bem como em outras regiões tropicais (ASNER et al., 2005).

O mecanismo REDD + (*Reduce Emissions from Deforestation and Forest Degradation*), parte fundamental do acordo assinado na COP21 em Paris, é um conjunto de incentivos econômicos com o fim de reduzir as emissões de gases de efeito estufa resultantes do desmatamento e da degradação florestal (UN, 2015). Bustamante et al. (2016) e Morton (2016), relatam a existência de uma necessidade de medir o impacto da degradação florestal nos estoques de carbono nas florestas tropicais para apoiar o REDD + na melhoria da precisão dos orçamentos globais de carbono.

Florestas tropicais são importantes reservatórios de carbono e biodiversidade. No entanto, o carbono liberado da degradação florestal é altamente incerto, devido tanto a área afetada quanto a perda de carbono por degradação serem mal quantificadas (ANDERSEN et al., 2014). Caracterizar a distribuição espacial da biomassa acima do solo, do inglês *Above-Ground Biomass-AGB*, é um pré-requisito para a compreensão da dinâmica do ciclo do carbono nas florestas tropicais ao longo do tempo (LEITOLD et al., 2015).

A implicação da redução das florestas no ciclo do carbono, um elemento importante na modelagem dos ciclos biogeoquímicos, criou uma demanda para o desenvolvimento de métodos não destrutivos para a determinação da biomassa florestal (WATZLAWICK et al., 2009). Para Mutanga et al. (2012), os métodos destrutivos tradicionalmente empregados na determinação da biomassa podem ser trabalhosos, demorados e onerosos, com o deslocamento de equipes em áreas de difícil acesso, tornando a prática executável em áreas relativamente pequenas.

Desde que existam informações prévias tomadas a campo, métodos indiretos, segundo Watzlawick et al. (2006), permitem estimativas em menor tempo, menores custos, bem como mapeamento das variáveis para áreas da mesma tipologia florestal. Nesse sentido, métodos baseados em sensoriamento remoto na estimativa de biomassa acima do solo em ecossistemas florestais ganharam atenção crescente, e pesquisas substanciais foram conduzidas nas últimas três décadas (LU et al., 2014).

O sistema LiDAR aerotransportado tem sido usado com sucesso para medir a estrutura da floresta e estimar a biomassa acima do solo em uma série de ecossistemas florestais (NÆSSET, 1997; ASNER et al., 2008), e pode ser uma ferramenta valiosa para estimar o estoque de carbono florestal, bem como monitorar a degradação da floresta por meio da exploração madeireira (ANDERSEN et al., 2014). Leitold et al. (2015) relatam que a abordagem típica para predição de AGB com dados de *laser* aerotransportado é a construção de modelos de regressão que associam métricas LiDAR às estimativas de biomassa obtidas via inventário florestal, de modo que os modelos ajustados são usados para estimativas de grandes áreas.

Um aspecto relevante na elaboração dessas estimativas se refere ao método empregado para a construção dos modelos preditivos. A regressão é a ferramenta padrão de ajuste de modelos para as várias tarefas da mensuração florestal (MONTAÑO, 2016). Um importante desenvolvimento estatístico dos últimos trinta anos foi o avanço na análise de regressão proporcionada pelos modelos GLMs (*Generalized Linear Models*), o que ocorreu devido a sua capacidade de tratamento de informações, podendo ser aplicado a dados que não apresentam normalidade na sua distribuição, ao contrário da regressão linear clássica (GUISAN et al., 2002).

Entretanto, especialistas na área buscam alternativas que possam aproveitar o poder computacional atual, bem como fazer uso das várias informações disponíveis em suas bases de dados, o que não é possível com a maioria dos modelos baseados em regressão tradicional, em razão da sua baixa flexibilidade e rigidez (MONTAÑO, 2016). Nesse sentido, as técnicas de aprendizado de máquina, uma subdivisão da inteligência artificial, que utiliza algoritmos de arquitetura sofisticada como *Random Forest* – RF, *Support Vector Machine* – SVM e *Artificial Neural Network* - ANN, têm demonstrado grande capacidade para construção de modelos mais complexos, como a regressão não linear multivariada e regressão não paramétrica (LARY et al., 2016). O uso desses métodos vem ganhando papel de destaque na modelagem e classificação dos elementos da paisagem, especialmente em razão da capacidade de tratar

informações e gerar resultados com grande eficiência computacional, a partir de uma base de dados complexa e volumosa.

Nesse contexto, o trabalho tem como objetivo geral a aplicação de técnicas de aprendizado de máquina na construção de modelos que relacionam métricas LiDAR e dados de inventário florestal na predição de biomassa acima do solo em floresta tropical. Assim, para o desenvolvimento da pesquisa, foram estabelecidos os seguintes objetivos específicos:

1. Realizar estimativas da biomassa acima do solo por meio de equação alométrica ajustada fazendo uso de variáveis biométricas previamente inventariadas a campo;
2. Aplicar três diferentes técnicas de aprendizado de máquina para realização das estimativas via dados LiDAR: *Random Forest* – RF, *Support Vector Machine* – SVM e *Artificial Neural Network* – ANN;
3. Avaliar o desempenho dos modelos quanto ao ajuste e realizar a validação das estimativas;
4. Avaliar a capacidade preditiva dos algoritmos de aprendizado de máquina ao trabalhar com a totalidade das variáveis preditoras, sem aplicação de técnicas para redução da dimensionalidade dos dados;
5. Verificar o desempenho dos modelos de aprendizado de máquina frente a um modelo do tipo *Generalized Linear Model* – GLM, tradicionalmente empregado em estimativas não paramétricas.

2 REVISÃO DE LITERATURA

2.1 FLORESTA AMAZÔNICA

A Floresta Amazônica pode ser caracterizada como a formação vegetal fundamental do Bioma Amazônico, o qual segundo ISA (2009), pode ser denominado como Domínio Ecológico Amazônico ou Domínio Biogeográfico Amazônico e corresponde ao conjunto de ecossistemas florestais existentes na Bacia Amazônica. O bioma amazônico abrange uma área de 4,2 milhões de km^2, o que corresponde aproximadamente a 49,3% do território nacional, além de abrigar a maior rede hidrográfica do mundo que concentra 15% da água doce superficial não congelada do planeta (SNIF, 2018).

Esse ambiente é formado principalmente por florestas densas e abertas, porém abriga uma diversidade de outros ecossistemas, como florestas estacionais, florestas de igapó, campos alagados, várzeas, savanas, refúgios montanhosos, campinaranas e formações pioneiras, que abrigam vastos estoques de madeira comercial e de carbono e possuem uma grande variedade de produtos florestais não madeireiros que permite a manutenção de diversas comunidades locais (SNIF, 2018). Na floresta amazônica existem cerca de 2.500 espécies arbóreas (ou um-terço de toda a madeira tropical do mundo) e 30 mil espécies de plantas (das 100 mil da América do Sul), essas estimativas situam a região como a maior reserva de madeira tropical do mundo, além de recursos não madeireiros como enormes estoques de borracha, castanha, peixe e minérios (MMA, 2019).

A floresta amazônica é o maior reservatório natural da diversidade vegetal do planeta, onde cada um de seus diferentes ambientes florestais possui um contingente florístico rico e variado, muitas vezes exclusivo de determinado ambiente (OLIVEIRA e AMARAL, 2004). Em termos quantitativos, estimativas do Serviço Florestal Brasileiro apontam que para o ano de 2015 a amazônia brasileira continha cerca de 91.691 milhões de metros cúbicos de madeira e 104.735 milhões de toneladas de biomassa, que correspondem a 68.571 milhões de toneladas de carbono estocado (SNIF, 2019). No entanto, segundo MMA (2019), essa riqueza natural se contrapõe dramaticamente aos baixos índices socioeconômicos da região, de baixa densidade demográfica e crescente urbanização, de modo que o uso dos recursos florestais é estratégico para o desenvolvimento da região.

2.2 SENSORIAMENTO REMOTO

De acordo com Jensen (2009), a definição máxima globalizante do sensoriamento remoto é a de que o mesmo se refere à "aquisição de dados sobre um objeto sem tocá-lo". Ainda segundo o autor, de forma mais aprofundada pode-se defini-lo como o registro da informação nas regiões do ultravioleta, visível, infravermelho e micro-ondas, sem contato, por meio de instrumentos tais como câmeras, escâneres, lasers, dispositivos lineares e/ou matriciais localizados em plataformas tais como aeronaves ou satélites, e a análise da informação adquirida por meio visual ou processamento digital de imagens. Já Silva (2007) define o sensoriamento remoto como uma técnica de aquisição e análise das informações sobre as propriedades físico-químicas de objetos de interesse, ou fenômenos dinâmicos da superfície terrestre, com base nas interações da radiação eletromagnética com os alvos e o meio ambiente.

De acordo com Di Maio et al. (2008), os sensores remotos operam em determinadas faixas do espectro eletromagnético e são dispositivos capazes de detectar a energia eletromagnética proveniente de determinados alvos, convertê-la em sinais elétricos, e registrá-las, de modo que esse registro possa ser armazenado ou transmitido em tempo real para posteriormente ser convertido em informações que descrevem as feições dos objetos que compõem a superfície terrestre. Ainda segundo os autores, a energia eletromagnética proveniente dos alvos pode ser adquirida por sistemas sensores do tipo imageadores ou não-imageadores.

Segundo Moreira (2005), quanto a fonte de radiação, os sensores remotos podem ser classificados em passivos e ativos. De acordo com o autor, quando o sistema sensor emite a radiação e, após ter interagido com o alvo, capta a parte que voltou, o sistema é denominado ativo, isto é, possui sua própria fonte de radiação, podendo ser operado durante o dia ou a noite, como ocorre com os sistemas radares, o *laser*, e os radiômetros de micro-ondas. Já os sistemas sensores que não possuem própria fonte de radiação são denominados passivos (JENSEN, 2009).

Para Novo (2010), a evolução dessa variedade de produtos dos sistemas de sensoriamento remoto, tem representado inúmeras opções de uso com maior agilidade e precisão como por exemplo na agricultura, cartografia, ecologia, florestal, hidrologia, militar, extração de minerais, oceanografia, planejamento urbano, entre outras áreas. Sousa e Ponzoni (1998), relatam que na área florestal, essas técnicas são de grande utilidade, contribuindo como método indireto nas estimativas de parâmetros como biomassa e estoque de carbono, entre outras variáveis.

2.2.1 Sistema LiDAR

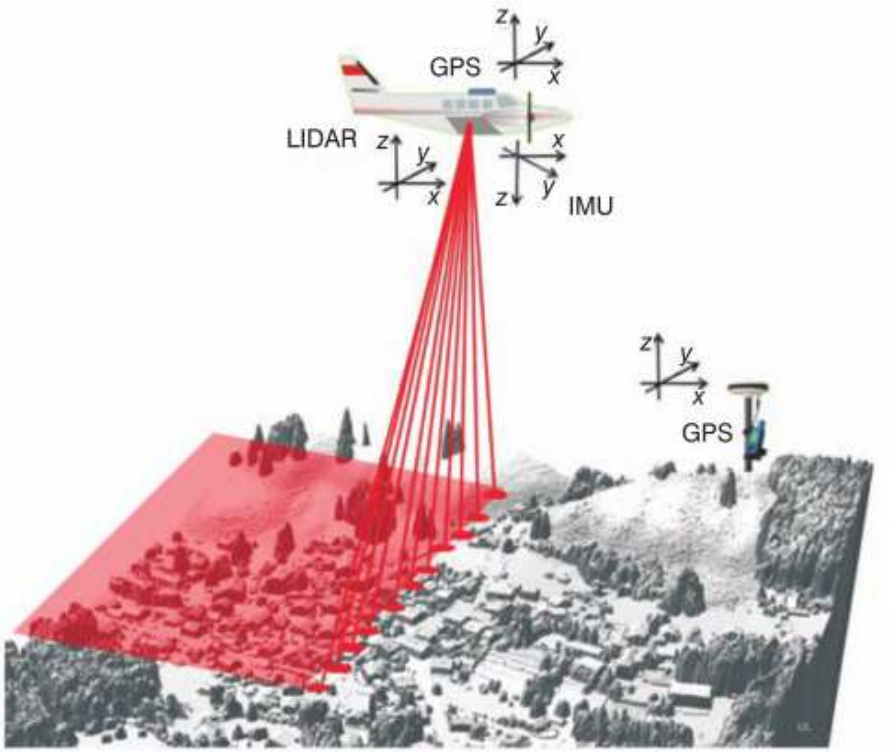
A varredura a laser aerotransportada (*Airborne laser scanning - ALS*), também conhecida como LiDAR (*Light Detection and Ranging*) consiste numa técnica de sensoriamento remoto baseada na transmissão de pulsos laser para a superfície terrestre, os quais interagem com objetos (árvores, edifícios, superfície do solo, entre outros) e são refletidos para o sistema sensor (SMREČEK e DANIHELOVÁ, 2013). Os sensores LiDAR medem o período de tempo entre a emissão e retorno dos pulsos, para calcular o espaço entre o dispositivo e os alvos (YADAV, 2016).

Schawlow e Townes (1958) estabeleceram a base para obtenção de luz amplificada com emissão estimulada de radiação, especialmente para as regiões ópticas e infravermelho do espectro eletromagnético, por meio do uso de uma cavidade ressonante de dimensões centimétricas. Apesar do *laser* não ser uma tecnologia nova, sua utilização na aquisição de dados geográficos é relativamente recente, seu uso em sistemas LiDAR vem demonstrando uma excelente capacidade para a aquisição de uma grande quantidade de informações, em pequeno intervalo de tempo (GIONGO et al., 2010)

Os instrumentos LiDAR podem ser operados a partir do solo (*Terrestrial Laser Scanning - TLS*) de plataformas aerotransportadas (*Airborne Laser Scanning - ALS*) ou de satélites (ANDERSON, 2016). De acordo com Vosselman e Mass (2010), a técnica de varredura a laser aerotransportado é baseada em dois componentes principais: Um sistema de *scanner* a *laser* que mede a distância até um ponto no solo iluminado pelo pulso e uma combinação GPS/IMU (*Global Position System*/*Inertial Measurement Unit*) para medir exatamente a posição e a orientação da plataforma que comporta o sistema, conforme mostra a Figura 1. Ainda segundo os autores, sistemas ativos baseados em varredura a laser são independentes da luz solar, de modo que podem ser operados durante o dia ou à noite, o que confere uma vantagem considerável ao *laser* aerotransportado em relação aos outros métodos de levantamentos de paisagem. A combinação desses elementos, permite ao sistema LiDAR a geração de nuvens tridimensionais (3D) densas e georreferenciadas que retratam a parcela reflexiva da superfície terrestre (YADAV, 2016).

Outra característica importante desses sistemas é a área de cobertura *laser* instantânea, conhecida na literatura no *footprint (instantaneous laser footprint).* Segundo Jensen (2009), dependendo, principalmente, da altitude do instrumento LiDAR e do ângulo em que é enviado, cada pulso irá iluminar no terreno uma área com o formato aproximado de um círculo, com cerca de 30cm (*small-footprint*) em sistemas aerotransportados.

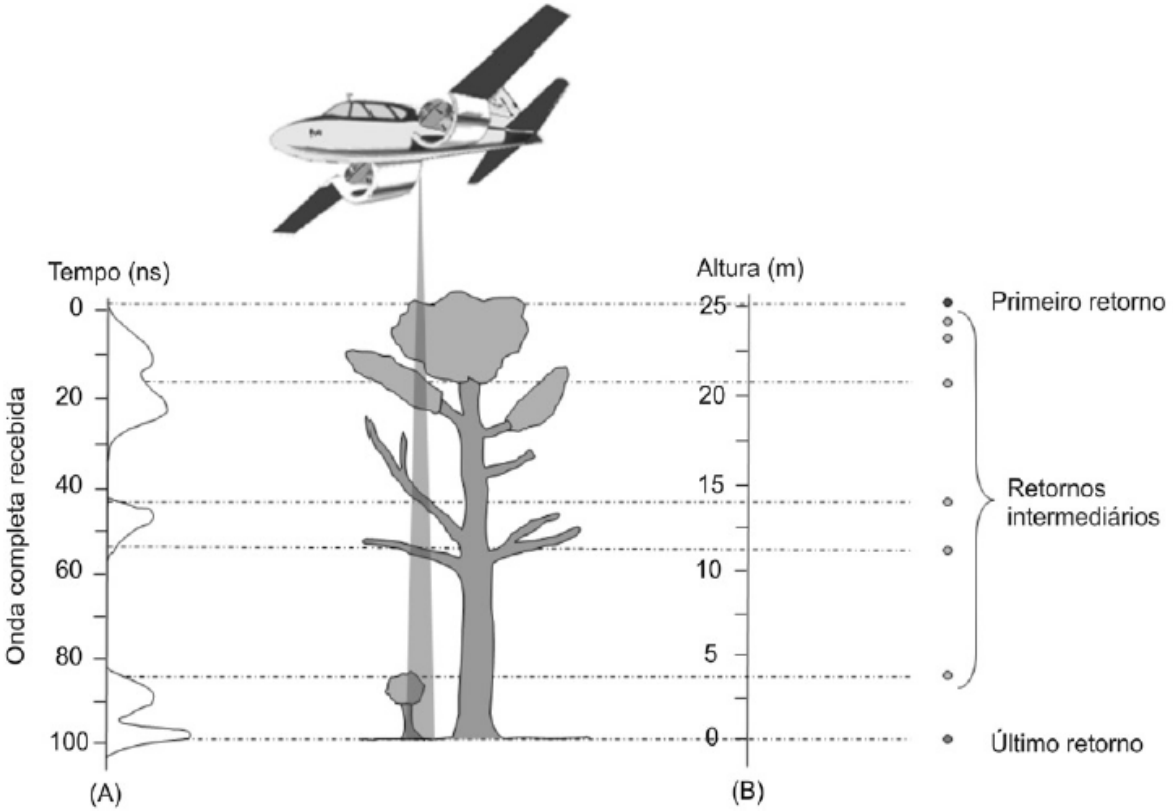
Figura 1 – Princípio de funcionamento do sistema aerotransportado de varredura laser



Fonte: Vosselman e Mass (2010).

Outra característica da tecnologia LiDAR que merece destaque está relacionada aos diferentes tipos de retorno dos pulsos emitidos pelo sistema sensor. Os sistemas LiDAR podem ser classificados em sistemas de amostragem de retorno discreto (*discrete-return)* ou de forma de onda (*waveform*) (COOPS et al., 2007), os dois sistemas são caracterizados na Figura 2. Segundo Giongo et al. (2010), os primeiros sistemas LiDAR registravam apenas o primeiro ou o último retorno dos pulsos, já com a evolução dos sistemas, as duas informações são registradas de forma simultânea, bem como alguns retornos intermediários. Nos últimos anos, foram desenvolvidos os sistemas de ondas contínuas (*full-waveform*), capazes de registrar integralmente todo sinal do pulso (WAGNER et al., 2006).

Figura 2 – Diferentes sistemas de varredura a *laser* aerotransportada, (A) sistema de ondas contínuas (*full-waveform*) e (B) sistema de pulsos (*discrete-return*)



Fonte: Adaptado de Giongo et al. (2010).

2.2.1.1 Uso do LiDAR no levantamento da cobertura florestal

Os sistemas LIDAR de uso tipicamente florestal são geralmente do tipo *discrete-return e small-footprint*, esses sistemas permitem o posicionamento preciso em três dimensões de superfícies reflexivas no solo e nos elementos constituintes da vegetação como folhas, ramos e troncos (RODRIGUEZ, et al., 2010). Segundo Giongo et al. (2010), a tecnologia LiDAR foi inicialmente prevista para a utilização em levantamento de dados para a elaboração de Modelos Digitais de Elevação (MDE), em que os métodos tradicionais não eram suficientes, principalmente em áreas de difícil acesso, no entanto, o sensor de varredura a *laser* aerotransportado tem grande potencial para aplicações florestais, em termos de sua capacidade de mobilização de uma grande quantidade de pontos com alta precisão, baixo custo e alta velocidade de aquisição de dados da estrutura vertical e horizontal das áreas florestais.

Do ponto de vista do sensoriamento remoto, a tecnologia LiDAR tem demonstrado essencial valor para obtenção de dados florestais. LiDAR aerotransportado tem sido amplamente utilizado nos últimos anos para estimar parâmetros biofísicos de árvores em ecossistemas florestais (SHRESTHA e WYNNE, 2012; ZANDONÁ et al., 2008).

Nilsson (1996) destaca que o sistema LiDAR tem grande potencial tanto para realizar medidas diretas sobre a cobertura florestal, bem como estimativas de variáveis dendrométricas. De acordo com Giongo et al. (2010), a determinação da altura do dossel pelo perfilhamento *laser*, pode ser descrita como medida direta, já outras características estruturais importantes, como a biomassa da parte aérea, área basal e diâmetro podem ser obtidas por técnicas de modelagem e/ou estimadas a partir de medições diretas.

Estudos recentes têm demostrado o uso crescente de dados LiDAR em estimativas de biomassa acima do solo em regiões de floresta tropical (HUNTER et al., LEITOLD et al., 2015; CHEN et al., LONGO et al., SATO et al., 2016; ANTHOM et al., FENG et al., SILVA et al., 2017a; LEITOLD et al., MEYER et al., RAPPAPORT et al., 2018). A Tabela 1 traz exemplos de estudos realizados nas duas últimas décadas que utilizaram bases de dados LiDAR aplicados ao levantamento florestal em diversas finalidades.

Tabela 1 - Exemplos de uso do LiDAR em diferentes estudos florestais

(continua)

Variáveis	**Forma de obtenção das variáveis**	**Literatura**
Altura de árvores individuais	Medida direta	Chen et al. (2006); Figueiredo (2014); Lingnau et al. (2008); Næsset; Okrand (2002); Roberts et al. (2005).
Altura do dossel	Medida direta	Andersen et al. (2003); Coops (2007); Holmgren et al. (2003); Popescu et al. (2002); Zimble et al. (2003).
Área basal	Modelagem biométrica	Andersen et al. (2003); Gobakken; Hudak et al.(2006); Næsset (2004).
Área de projeção de copa	Medida direta	Figueiredo (2014).
Área de preservação permanente Área de acesso restrito Hidrografia	Modelagem	d'Oliveira; Papa (2013)
Biomassa	Modelagem biométrica	Andersen et al. (2011,2014); Danilin; Medvedev (2004); d'Oliveira et al. (2012); Lim et al. (2004); Næsset; Gobakken (2008).

Tabela 1 - Exemplos de uso do LiDAR em diferentes estudos florestais

(conclusão)

Variáveis	Forma de obtenção das variáveis	Literatura
Biomassa de árvores individuais	Modelagem biométrica	Figueiredo (2014); Popescu et al. (2002).
Diâmetro à altura do peito (DAP)	Modelagem biométrica	Gobakken; Næsset (2004); Lingnau et al.(2008)
Diâmetro de copa	Medida direta	Figueiredo (2014); Næsset; Okrand (2002); Popescu et al. (2003); Roberts et al. (2005).
Impacto de exploração	Modelagem	d'Oliveira et al. (2012); Figueiredo (2014).
Manto de copa	Medida direta	Figueiredo (2014); Roberts et al. (2005); Sasaki et al. (2008).
Número de indivíduos	Medida direta	Bottai et al. (2013); Chen et al. (2006); Gonçalves et al. (2011); Næsset (2004); Tiede et al. (2005); Zonete (2009).
Volume de copa	Medida direta	Coops (2007); Figueiredo (2014).
Volume do fuste de árvores individuais	Modelagem biométrica	Bottai et al. (2013); Figueiredo (2014); Popescu et al. (2003).
Volume	Modelagem biométrica	Andersen et al. (2003); Holmgren et al. (2003); Loki et al. (2010).

Fonte: Adaptado de D'Oliveira et al. (2014).

2.3 MODELAGEM PREDITIVA POR REGRESSÃO

A análise de regressão é uma técnica estatística utilizada para investigar e modelar, com base em um banco de dados, a relação entre uma variável de interesse e um conjunto de variáveis explicativas (BAYER, 2011). Batista (2014), relata que a inferência de determinada característica por meio de variáveis explicativas correlacionadas, é resultado muitas vezes da impossibilidade de obtenção de atributos da variável de interesse por ordem prática ou operacional, como é o caso do volume e biomassa arbórea.

Fisher (1922), discutiu três aspectos do problema geral da inferência válida: (1) especificação do modelo, (2) estimação dos parâmetros do modelo e (3) estimativa da precisão. Por sua vez, Burnham e Anderson (2002), preferem particionar a especificação do modelo em dois componentes: formulação de um conjunto de modelos candidatos e seleção de um modelo (ou pequeno número de modelos) a ser usado na realização de inferências.

De acordo com Schneider et al. (2009), a seleção do melhor modelo ajustado pode ser feita mediante certos critérios estatísticos como o coeficiente de determinação (R^2), erro padrão da estimativa (Syx), distribuição dos resíduos, entre outros fatores. A cada item avaliado, os modelos são ordenados, de modo que a soma dos escores permite o ranqueamento dos modelos segundo sua qualidade de ajuste aos dados.

Outro aspecto considerado na avaliação da qualidade das inferências diz respeito a base de dados para ajuste e validação. De acordo com Montaño (2016) qualquer análise de desempenho será mais otimista caso a validação seja aplicada nas estimativas computadas sobre a mesma base de dados utilizada para treinamento dos modelos. Theodoridis e Koutroumbas (2009), relatam que para resultados mais realistas, a base de dados deve ser dividida de modo aleatório em um subconjunto de treinamento para ajuste dos modelos e outro de teste, utilizado na validação. De acordo com Witten e Frank (2005), não existe um limiar padrão para o tamanho das bases, no entanto, são observados normalmente valores entre 70-30% ou 80-20%, como proporção entre o número de amostras de treinamento e teste.

2.3.1 Modelos lineares generalizados

Os Modelos Lineares Generalizados, do inglês *Linear Generalized Model – GLM,* foram propostos por Nelder e Wedderburn (1972) como uma maneira de unificar vários modelos estatísticos em um modelo geral de estimativa por máxima verossimilhança. O modelo GLM pode ser definido em termos de um conjunto de variáveis aleatórias independentes, pertencentes a uma distribuição da família exponencial (DOBSON e BARNETT, 2008). O pressuposto do uso de observações independentes condiciona o uso dos modelos GLM à uma avaliação da estrutura residual em termos da auto-correlação ou correlação em série, condicionante dos modelos clássicos de regressão que precisa ser observada (MCCULLAGH e NELDER, 1989)

Para Cordeiro e Demétrio (2008), a composição dos modelos lineares generalizados envolve uma variável resposta univariada (denominada componente aleatória do modelo) que possui uma distribuição pertencente à família exponencial, variáveis explanatórias (componente sistemática do modelo) na forma de uma estrutura linear, e uma função de ligação

adequada, responsável pela associação das duas componentes do modelo. A Tabela 2 traz alguns exemplos de distribuições e suas funções de ligação.

Tabela 2 – Principais distribuições pertencentes à família exponencial e suas respectivas funções de ligação utilizadas nos Modelos Lineares Generalizados

Distribuição	**Funções de ligação**
Binomial	*logit, probit e cauchit*
Gaussian	*identity, log e inverse*
Gamma	*inverse, identity e log*
Inverse.gaussian	*$1/\mu^2$, inverse, identity e log*
Poisson	*log, identity e sqrt*
Quasi	*logit, probit, cloglog, identity, inverse, log, $1/\mu^2$ e sqrt*
Quasibinomial	*log*
Quasipoisson	*logit*

Fonte: Modificado de R Documentation (2018).

Os modelos GLMs frequentemente fazem uso de técnicas de redução de variáveis, como o procedimento *Stepwise* e Análise de Componentes Principais – ACP, previamente ao ajuste. Para Cordeiro e Demétrio (2008), a parcimônia é uma exigência na regressão GLM, de modo que o número de parâmetros seja tão pequeno quanto possível, seguindo conceito de que a explicação mais simples é a melhor.

2.3.2 Aprendizado de máquina

O aprendizado de máquina, do inglês *Machine Learning (ML)*, incorpora, segundo Lary et al. (2016), uma ampla gama de procedimentos complexos, sendo uma subdivisão da inteligência artificial baseada no processo de aprendizagem biológica que abrange diferentes domínios, como mineração de dados, e engloba uma variedade de algoritmos (por exemplo, redes neurais, máquinas de vetor de suporte, mapa de auto-organização, árvores de decisão, florestas aleatórias, raciocínio baseado em casos, programação genética, entre outros) que podem fornecer regressão não paramétrica, não linear multivariada ou classificação.

Segundo Ali et al. (2015) , dentre as vantagem do uso de algoritmos de aprendizado de máquinas pode-se elencar: Por diversas vezes, mais preciso do que regras criadas por humanos,

já que as mesmas são baseadas nos dados; Método automático para procurar hipóteses explicando dados; Flexível e pode ser aplicado a qualquer tarefa de aprendizado, e; Interação rica entre teoria e prática, com melhores resultados à medida que os conjuntos de dados aumentam.

As técnicas de aprendizado de máquina empregam um princípio de inferência denominado indução, no qual obtém-se conclusões genéricas a partir de um conjunto particular de exemplos, de modo que esse aprendizado indutivo pode ser dividido em dois tipos principais: supervisionado e não supervisionado (LORENA e CARVALHO, 2007). De acordo com Melo (2010), o aprendizado supervisionado se dá com a utilização do atributo classe para a correta construção do modelo de aprendizado baseado no conjunto de treinamento, tal como ocorre com algoritmos de Redes Neurais, Árvores de Decisão e Redes Bayesianas. Ainda segundo o autor, o aprendizado não supervisionado é processado sem a necessidade do atributo de classe, de modo que esses algoritmos são mais utilizados para agrupamento de dados, atribuição de classes e detecção de *outliers*.

Como ocorre nos modelos tradicionais, os Modelos de Aprendizado de Máquina são compostos por parâmetros que constituem seu arranjo matemático e são estimados a partir do ajuste aos dados de treinamento (MURPHY, 2012). Segundo Júnior (2018), nestes modelos, existem ainda parâmetros denominados hiperparâmetros, que se diferenciam dos primeiros por não serem estimados do mesmo modo e por necessitarem de uma definição de valores definitiva, antes mesmo que o treinamento se inicie.

Os hiperparâmetros de um algoritmo de aprendizado de máquina são propriedades de alto nível do modelo estatístico de algoritmo, que podem influenciar fortemente sua complexidade, sua velocidade na aprendizagem e seus resultados de aplicação (CHICCO, 2017). Dessa forma, para Júnior (2018), os hiperparâmetros estão diretamente ligados ao desempenho do modelo treinado e ao número de operações computacionais necessárias no aprendizado, assim como seu respectivo tempo de duração.

2.3.2.1 Florestas aleatórias

A floresta aleatória, ou *Random Forest* - RF é um algoritmo de aprendizado baseado em conjuntos (do inglês *ensemble learning*) (BREIMAN, 2001). De acordo com Shao e Zhang (2016), o algoritmo RF foi projetado para melhorar o método de árvores de classificação e regressão linear tradicional, de modo a integrar um grande conjunto de árvores de decisão baseado em uma técnica determinística, selecionando um conjunto aleatório de variáveis e uma

amostra aleatória do treinamento. Montaño (2016), relata que o algoritmo utiliza processos diferentes dependendo da natureza do problema, de modo que para classificação, o resultado final é alcançado pela votação de cada árvore da floresta, enquanto na regressão, a saída da floresta é a média entre os resultados das árvores.

Segundo Lucas (2011), o RF faz uso, basicamente, de três técnicas combinadas ao realizar o tratamento dos dados: *Bagging, boosting e randomizing*. No processo de formação de cada uma das árvores, o RF utiliza o método de *bagging,* um processo de reamostragem (*bootstraping*), de um conjunto de amostra da base de dados original através de seleção aleatória com reposição (IBAÑEZ, 2016). A ideia essencial do *bagging* é calcular a média de muitas árvores enviesadas, mas aproximadamente imparciais, e assim reduzir a variação (HASTIE et al., 2017). Para Lopes et al. (2017), a utilização da técnica *bagging* no processo de treinamento além de reduzir a variância ajuda a evitar o *overfitting,* ou sobreajuste, que segundo Chicco (2017), é um efeito presente nos casos em que um algoritmo se adapta excessivamente ao conjunto de treinamento e se mostra ineficaz na predição com novos dados de entrada, executando mal o conjunto de teste/validação.

No processo de *boosting*, as árvores sucessivas dão um peso extra aos pontos incorretamente previstos pelos preditores anteriores Liaw e Wiener (2002). Já o processo de aleatorização ou *randomizing,* de acordo com Lucas (2011), é responsável por desenvolver, a cada interação, uma árvore de decisão utilizando um subconjunto de variáveis preditoras selecionadas aleatoriamente para executar a divisão dos nós (*split*), de modo que para uma mesma árvore, os subconjuntos sejam diferentes em cada nó.

Prasad et al. (2006), relatam que o algoritmo RF não é limitado pela distribuição de covariáveis e sensível a *outliers* e ruídos. O *Random Forest* tornou-se popular no sensoriamento remoto como uma alternativa não-linear e não-paramétrica, com capacidades preditivas promissoras para conjuntos de dados de alta dimensão (PAL, 2005).

2.3.2.2 Máquina de vetor de suporte

A Máquina de Vetor de Suporte – SVM, consiste em um método de aprendizado supervisionado desenvolvido por Cortes et al. (1995), baseado no conceito de planos de decisão que definem limites de decisão. A abordagem de modelos SVM está fundamentada no aprendizado estatístico, combinando controle, generalização com uma técnica para tratar o problema da alta dimensionalidade dos dados (JUNIOR, 2007).

Segundo Shao e Zhang (2016), a ideia principal por trás dos algoritmos SVM é estimar a dependência linear entre pares de vetores de entrada n-dimensionais e variáveis-alvo unidimensionais, ajustando um hiperplano de aproximação ótimo a um conjunto de amostras de treinamento. As instâncias (observações) que possuem distância mínima até o hiperplano são denominadas vetores de suporte. Treinado o modelo, a predição de uma instância desconhecida é feita determinando-se a posição da mesma em relação ao hiperplano de separação (MERSCHMANN, 2007).

Neste método, hiperparâmetros devem ser configurados, dentre estes, tem-se o custo (C) que permite equilíbrio entre a precisão e a complexidade do modelo e ainda a função *kernel* que permite projetar valores para um plano maior, onde os dados apresentam mais probabilidade de serem linearmente separáveis (MONTAÑO, 2016). Outro elemento a ser configurado no método SVM é o parâmetro *gamma.* Segundo Pedregosa et al. (2011), o parâmetro *gamma* define intuitivamente até onde chega a influência de um único exemplo de treinamento, com valores baixos significando "longe" e valores altos significando "perto", de forma que os parâmetros *gamma* podem ser vistos como o inverso do raio de influência das amostras selecionadas pelo modelo como vetores de suporte.

De acordo com Lorena e Carvalho (2007), a técnica SVM vem recebendo crescente atenção nos últimos anos, sendo utilizadas em diversas tarefas de reconhecimento de padrões, obtendo resultados superiores aos alcançados por outras técnicas de aprendizado em várias aplicações. Os resultados obtidos apontam as características promissoras desse método, como a boa capacidade de generalização intrínseca e a robustez ao ruído no caso de disponibilidade limitada das amostras de referência (ALI et al., 2015).

2.3.2.3 Redes neurais artificiais

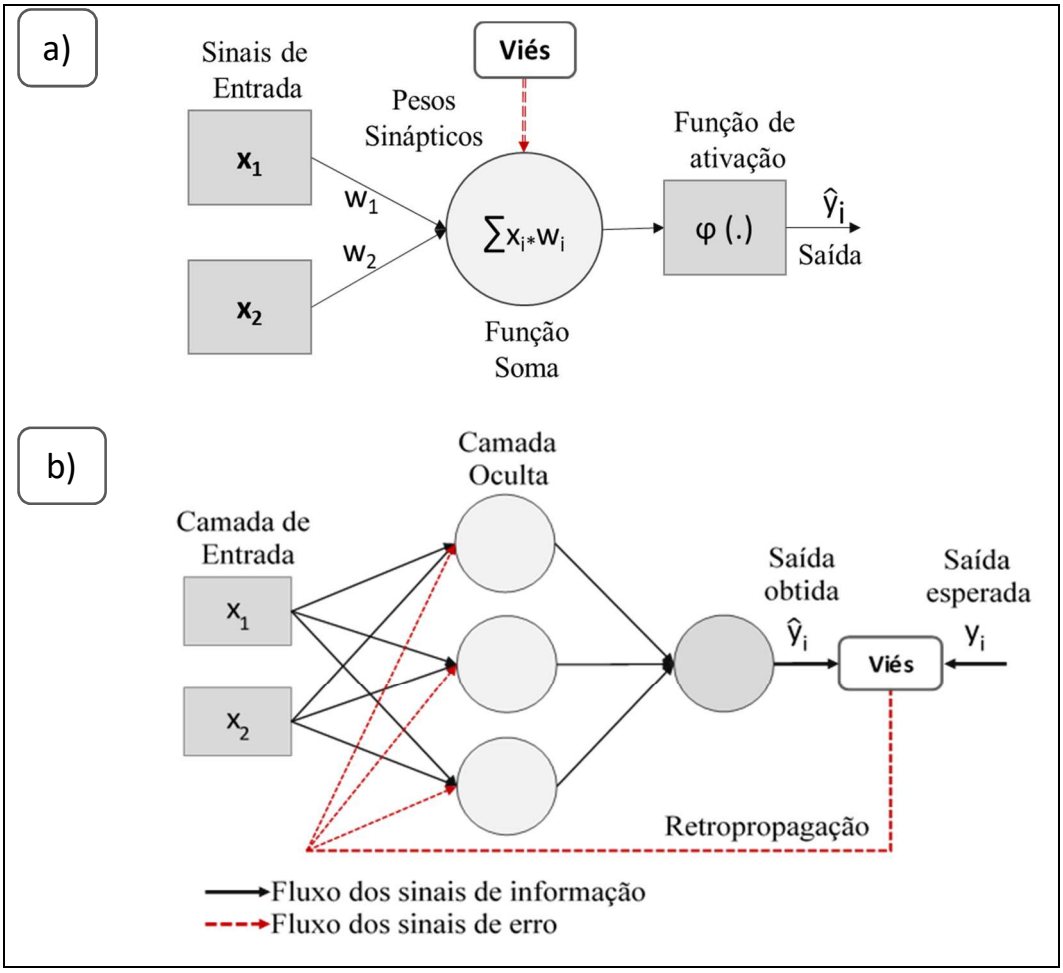
As Redes Neurais Artificiais (*Artificial Neural Network - ANN*), são módulos matemáticos semelhantes ao nosso sistema nervoso, compostas por unidades de processamento análogas aos em uma versão matemática de um neurônio (FERNEDA, 2006). As Redes Neurais Artificiais (*Artificial Neural Network - ANN*) têm o poder de recuperar os padrões complexos, dinâmicos e não lineares dos dados, sendo um dos mais antigos métodos de aprendizado de máquina, eles são bem estudados e são fáceis de implementar, já que muitas bibliotecas e ferramentas de *software* estão disponíveis (ALI et al., 2015).

O algoritmo *backpropagation* é um dos mais utilizados dentre as RNAs, baseado no método *Multilayer Perceptron* (MLP), o qual cria uma arquitetura constituída por uma camada

de entrada, uma ou mais camadas ocultas e uma camada de saída (SOARES et al., 2011). De acordo com Ye (2014), o algoritmo de aprendizado *backpropagation* utiliza o conceito de retropropagação para procurar uma função de erro mínimo (erro quadrático médio) no espaço de ponderação, de forma que a combinação de pesos que minimiza a função de erro, é considerada uma solução do problema de aprendizagem.

A Figura 3 traz uma ilustração simplificada do mecanismo de funcionamento de um neurônio isolado, bem como a sua combinação dentro de uma rede MLP. Os valores recebidos em cada terminal de entrada do neurônio são ponderados e combinados por uma função de soma, cujo resultado é submetido a uma função de ativação que define a passagem ou não do sinal, segundo um limiar de aceitação (FACELI et al., 2011). Ainda segundo os autores, a saída do neurônio é confrontada com o resultado esperado e o resíduo é retropropagado compondo as interações subsequentes.

Figura 3 – Representação da estrutura de um neurônio artificial (a) e sua combinação dentro de uma rede *Multilayer Perceptron* utilizando o algoritmo *backpropagation* (b)



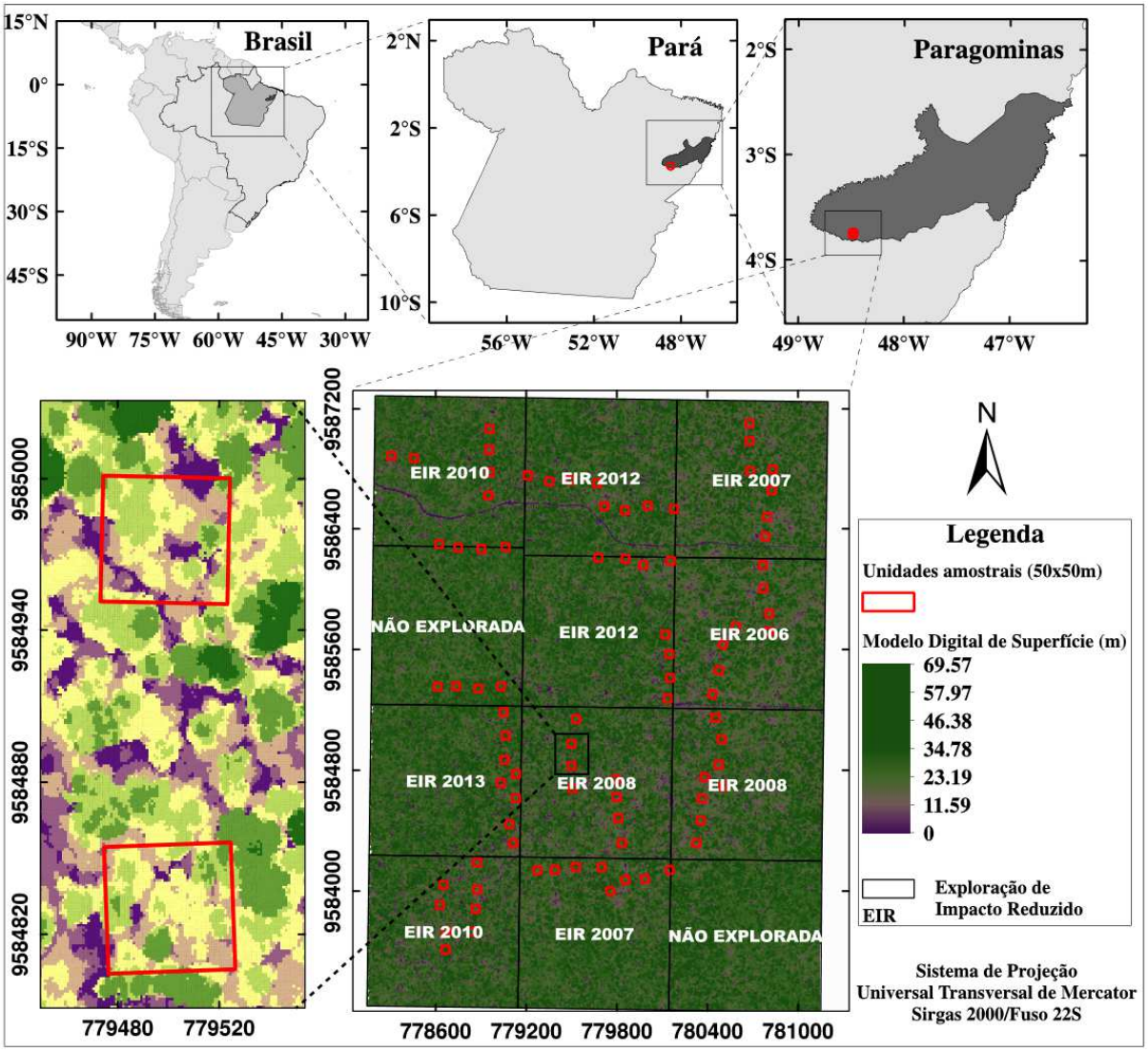
Fonte: Adaptado de Faceli et al. (2011).

3 MATERIAIS E MÉTODOS

3.1 CARACTERIZAÇÃO DA ÁREA DE ESTUDO

A área de estudo compreende uma extensão de aproximadamente 1.200 ha, localizada na Fazenda Cauaxi, no município de Paragominas, nordeste do estado do Pará, Brasil, conforme apresentado na Figura 4. A área, de propriedade privada, é dividida em 12 unidades, das quais duas estão intactas e em dez houve intervenção desde 2006, por meio da exploração de impacto reduzido – EIR (SILVA et al., 2017a).

Figura 4 – Arranjo das unidades amostrais distribuídas ao longo da área de estudo na fazenda Cauaxi, município de Paragominas/PA



Fonte: Autor.

3.1.1 Clima

Segundo Bastos et al. (2006), o município de Paragominas apresenta temperaturas entre 21 e 34º C, umidade relativa anual em torno de 81% e precipitação em torno de 2000 mm, sendo abundante de janeiro a maio. Ainda segundo os autores, o município pode ser enquadrado como de clima quente e úmido, do tipo Awi pela classificação de Köppen (Clima tropical chuvoso, com expressivo período de estiagem) e B1wA'a', da classificação de Thornthwaite (Clima tropical úmido, com expressivo déficit hídrico).

3.1.2 Geologia, Geomorfologia e Pedologia

De acordo com IBGE (2003a), existem duas formações geológicas presentes no município, a formação Itapecuru e a Cobertura Detrito-Laterítica Paleogênica. A primeira, para Duarte e Carneiro (2017), é originada do período Cretácio e constituída predominantemente de arenitos finos, caulínicos, argilitos laminados e calcários margosos fossilíferos, já a segunda é composta por uma crosta laterítica-bauxítica, argilas cauliníticas e arenitos argilosos às vezes conglomeráticos,

Á área do município encontra-se sob o domínio geomorfológico de Bacias sedimentares e Coberturas Inconsolidadas, cuja unidade geomorfológica é o Planalto Dissecado do Gurupi – Grajaú (IBGE, 2003b). As feições geomorfológicas dominantes são vastas chapadas interligadas, mas raramente isoladas, separadas por zonas mais baixas topograficamente que apresentam maior diversidade geomorfológica (KOTSCHOUBEY et al., 2005).

Latossolos Amarelos de texturas média e muito argilosa são dominantes, abrangendo mais de 81% da área do município (RODRIGUES et al., 2003). Ao analisar o mapa pedológico da Amazônia Legal elaborado pelo Instituto Brasileiro de Geografia e Estatística – IBGE (IBGE, 2003c) , percebe-se que a totalidade da área de estudo encontra-se dentro dessa classe. De acordo com Duarte e Carneiro (2017), Latossolos Amarelos são caracterizados como solos profundos, dissaturados e bem drenados, apresentam textura média a muito argilosa e podem ser encontrados em áreas com relevo plano a ondulado.

3.1.3 Vegetação

A tipologia florestal predominante no município é a Floresta Ombrófila Densa, formação Submontana (MMA, 2002). Segundo Veloso et al. (1991), a Floresta Ombrófila Densa (conhecida também por floresta pluvial tropical, Floresta Amazônica e Floresta

Atlântica) caracteriza-se por fanerófitos, lianas e epífitas em abundância e está condicionada a ocorrência de temperaturas elevadas, em média 25ºC, altas precipitações, bem distribuídas durante o ano, cujo período seco varia de 0 a 60 dias.

De acordo com IBGE (2012), a formação Submontana abrange áreas dissecadas do relevo montanhoso e dos planaltos com solos medianamente profundos que são ocupadas por uma formação florestal que apresenta fanerófitos com altura aproximadamente uniforme, de alto porte, alguns ultrapassando 50 m na Amazônia, integrada por plântulas de regeneração natural, poucos nanofanerófitos e caméfitos, além da presença de palmeiras de pequeno porte e lianas herbáceas em maior quantidade.

O Anexo A, traz um resumo do inventário florestal realizado no local de estudo no ano de 2014[1]. No total, foram levantadas 1649 árvores vivas, cuja identificação englobou 40 famílias botânicas e 151 espécies (com exceção de 23 indivíduos não identificados). A família *Lecythidaceae* foi a mais abundante com 421 indivíduos levantados, estes distribuídos em 8 espécies. A família *Fabaceae* também obteve destaque, com 261 exemplares, no entanto em 40 espécies distintas. Juntas, as famílias *Lecythidaceae, Sapotaceae e Fabaceae,* responderam por aproximadamente 60% do total de indivíduos identificados a campo. A nível de espécie, *Lecythis idatimon*, *Eschweilera coriácea* e *Rinorea guianensis* foram as mais abundantes, com 157, 141 e 109 indivíduos respectivamente, cuja soma equivaleu a aproximadamente 25% dos registros.

3.1.4 Histórico de ocupação

De acordo com IBGE (2017), o município de Paragominas foi formado por colonizadores goianos, mineiros, baianos e paulistas. Por volta da década de 30, esses agricultores passaram a desenvolver, principalmente, pequenas atividades agrícolas de arroz, mandioca, e feijão, expandindo-se à medida que a população se reproduzia ou novas famílias se instalavam na área (SILVA et al, 2011).

Para Almeida e Uhl (1998), o município contém nas suas fronteiras, áreas dedicadas à exploração madeireira, à pecuária e à agricultura de corte e queima, atividades que chegaram ao município em momentos diferentes e causaram impactos econômicos, sociais e ecológicos de grandeza e intensidade distintas. Ainda segundo os autores, a agricultura de corte e queima, iniciada na década de 30, foi a primeira atividade econômica realizada no município, com a

[1] Levantamento descrito na sequência (item 3.2.1 Base de dados).

formação de colônias agrícolas de pequenos produtores. Já na década de 60, a implantação da rodovia BR-010, que liga Belém a Brasília, impulsionou o desenvolvimento da atividade pecuarista, que em pouco tempo, tornou-se a base econômica municipal (IBGE, 2017).

Vale et al (2017), mencionam que na década de 80, em função da redução dos incentivos destinados a pecuária, a atividade madeireira ganhou maior expressão, firmando-se como uma alternativa bastante lucrativa, tendo em vista a redução dos rebanhos e o esgotamento das áreas de pastagens. De acordo com Pinto et al. (2009), em razão das práticas de desmatamento provocadas pela atividade agropecuária, grandes áreas de floresta original foram substituídas por florestas secundárias, em diversos estágios de desenvolvimento. Os autores mencionam ainda que o grande número de áreas degradadas que caracterizam o processo histórico de ocupação antrópica, condicionam a adoção de medidas como o plantio de espécies florestais, tida como alternativa à recuperação do potencial produtivo dessas áreas, bem como uma atividade econômica de base florestal.

3.2 MATERIAIS

3.2.1 Base de dados

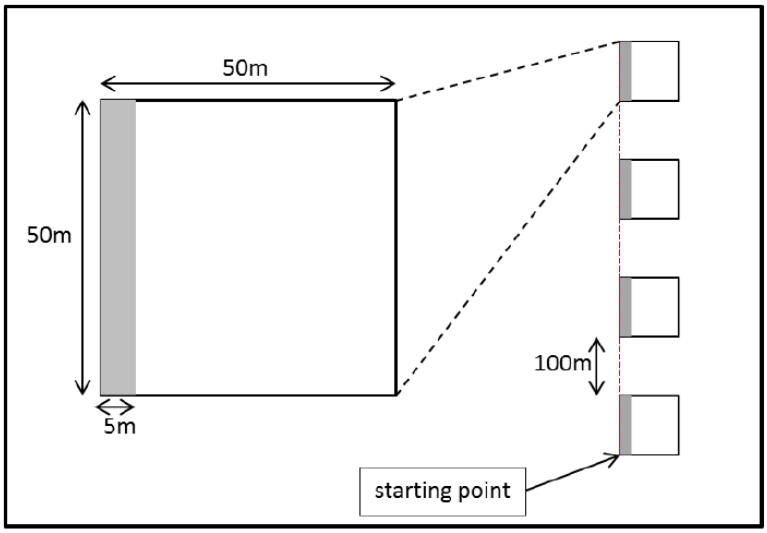
A base de dados utilizada nesse estudo foi a que se encontra disponível no ambiente web do projeto Paisagens Sustentáveis Brasil, produto da cooperação entre a Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA e o Serviço Florestal dos Estados Unidos (EMBRAPA , 2016). No portal, são disponibilizados dados de levantamento LiDAR da cobertura florestal em diferentes áreas, e em alguns casos o inventário florestal de campo correspondente.

A área inventariada na fazenda Cauaxi, no ano de 2014, compreende 88 unidades amostrais (UAs) de 50x50 m (2500 m²), dispostas em 22 linhas de referência (Figura 4). Em cada uma das parcelas, foram estabelecidas subparcelas com tamanho de 5x50 m (250 m²), conforme o detalhamento apresentado na Figura 5. Do total de parcelas inventariadas a campo, três não possuem cobertura de dados LiDAR, de modo que 85 UAs foram utilizadas no processo de modelagem.

Dentre as informações tomadas estão a identificação dos indivíduos, diâmetro a altura do peito (DAP), altura total e comercial, raio de copa (4 quadrantes), posição sociológica, luz de copa, densidade da madeira, coordenadas UTM, entre outras. O diâmetro de inclusão das árvores nas parcelas foi de 35 cm, enquanto que nas subparcelas foi de 10 cm.

Em relação ao levantamento dos dados LiDAR, o sobrevoo das unidades amostrais inventariadas representadas na Figura 4 foi realizado em dezembro de 2014. As características de voo e os atributos do sistema de sensor LiDAR utilizados estão resumidos na Tabela 3.

Figura 5 – Detalhamento das unidades amostrais inventariadas na Fazenda Cauaxi



Fonte: Paisagens Sustentáveis Brasil (2016).

Tabela 3 – Dados de voo e do sensor LiDAR utilizado no levantamento da área de estudo

Densidade média de retornos	61.38 ppm²
Densidade média do primeiro retorno	37.5 ppm²
Altitude média de voo	850 m
Ângulo de visada	12°
Sensor	OPTECH, ORION M300
Frequência do scanner	83 Hz
GNSS	APPLANIX, 09SEN243
Frequência GNSS	5 Hz
UMI (Unidade de Mensuração Inercial)	LITTON, 413996
Frequência UMI	100 kHz

Fonte: Adaptado de Paisagens Sustentáveis Brasil (2016).

3.2.2 Aplicativos utilizados

Os aplicativos utilizados nas diferentes etapas do trabalho encontram-se elencados abaixo, cuja aplicação individual é detalhada na descrição dos métodos (item 3.3).

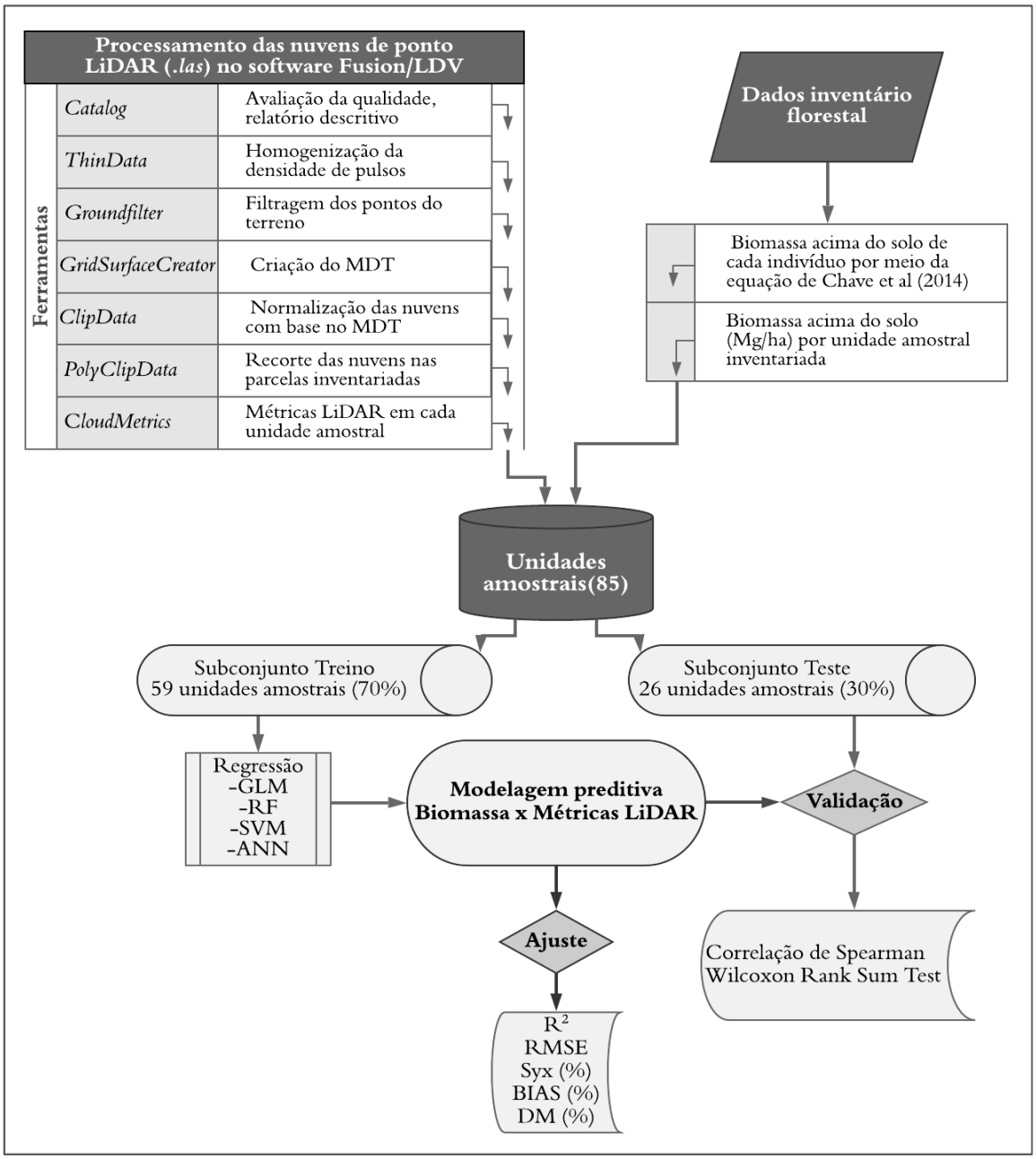
a) Fusion/LDV 3.8 (MCGAUGHEY, 2018);
b) EasyFit 5.6® - Licença de teste (EASYFIT, 2018);
c) R 3.5.0 (R CORE TEAM, 2018) no ambiente da interface gráfica RStudio® 1.1.463 (RSTUDIO TEAM, 2016), bem como o suporte dos seguintes pacotes:

- *FactoMineR* (LE et al., 2008);
- *randomForest* (LIAW e WIENER, 2002);
- *e1071* (MEYER et al., 2018);
- *neuralnet* (FRITSCH e GUENTHER, 2016);
- *NeuralNetTools* (BECK, 2018).

3.3 MÉTODOS

As etapas do método utilizado no estudo estão sintetizadas na Figura 6. O procedimento inicial correspondeu a obtenção das métricas LiDAR no *software* Fusion/LDV, bem como a estimativa de biomassa a partir dos dados de inventário. A base de dados derivada foi então subdividida em amostras de treinamento, para o ajuste, e de teste, para a validação de quatro diferentes modelos utilizados nas estimativas.

Figura 6 - Fluxograma do processamento dos dados, modelagem e verificação da qualidade dos modelos preditivos de biomassa acima do solo



Fonte: Autor.

3.3.1 Processamento dos dados de Inventário Florestal

A biomassa individual das árvores vivas presentes nos levantamentos do inventário florestal foi estimada a partir da equação de Chave et al. (2014) (Equação 1).

$$AGB\ (kg) = exp\ [-1{,}803 - 0{,}976E + 0{,}976\ln(\rho) + 2{,}673\ln(dap) - 0{,}0299\ [\ln(dap)]^2)] \quad (1)$$

Onde AGB (kg) é a biomassa acima do solo da árvore viva em Kg; dap é o diâmetro na altura do peito (1,30 m); $\boldsymbol{\rho}$ é a densidade da madeira e E é uma medida do estresse ambiental, no local de estudo considera-se E = -0,103815.

3.3.2 Processamento dos dados LiDAR

O processamento do conjunto de dados LiDAR foi realizado no ambiente de programação do *software* Fusion/LDV (MCGAUGHEY, 2018), combinando os métodos utilizados por Silva et al. (2017a) e Silva et al. (2017b), compreendendo as etapas descritas abaixo.

a) Checagem da qualidade: Avaliação da consistência das informações contidas em cada arquivo das nuvens de ponto, por meio da ferramenta *Catalog* que gera um relatório descritivo com algumas características do conjunto de dados LIDAR, como extensão da cobertura, tipos de retorno e densidade de pulso;
b) Homogeneização da densidade de pulsos: O conjunto de dados LiDAR original teve sua densidade de pulso homogeneizada e reduzida de 37,5 ppm² para 12 ppm², utilizando o algoritmo *ThinData*, segundo o método proposto por Silva, et al. (2017a). *ThinData* é útil ao comparar os resultados da análise de aquisições Li DAR coletadas usando diferentes densidades de pulso, ou quando a densidade dentro de um único conjunto de dados LiDAR não é uniforme, como acontece nas faixas de sobreposição das linhas de voo (MCGAUGHEY, 2016).
c) Filtragem dos pontos do terreno: Uso da ferramenta *Groundfilter* para reclassificação da nuvem de pontos e diferenciação entre os pontos do terreno e vegetação;
d) Criação do MDT (Modelo Digital do Terreno): A partir dos pontos do terreno obtidos na etapa anterior, foi aplicada a função *GridSurfaceCreate,* de interpolação, para criação de um *raster* (1m de resolução espacial) correspondente a superfície do terreno sob o dossel vegetal;
e) Normalização da nuvem: Subtração da coordenada altimétrica de cada ponto da nuvem em relação ao MDT. O processo de normalização assegura que a altimetria de cada ponto corresponda à altura acima do solo e não à elevação elipsoidal. Esse processo foi realizado pela ferramenta *ClipData*.

f) Recorte das unidades amostrais: Recorte do conjunto de dados LiDAR pela ferramenta *PolyClipData* usando como máscara arquivo no formato *shapefile* com as parcelas de inventário espacializadas.

g) Obtenção das métricas LiDAR em cada uma das unidades amostrais: Processo realizado pela ferramenta *CloudMetrics,* que, de acordo com Mcgaughey (2016), calcula uma variedade de parâmetros estatísticos que descrevem determinado conjunto de dados LiDAR, usando elevação e valores de intensidade dos pontos. O Anexo B traz uma breve descrição das métricas computadas, de acordo com o manual do usuário do aplicativo FUSION/LDV, versão 3.8.

3.3.3 Modelagem da biomassa

No processo de modelagem, a biomassa acima do solo obtida em cada unidade amostral inventariada (item 3.3.1), foi estabelecida como variável dependente, em função das 87 métricas LiDAR (item 3.3.2), tomadas como variáveis explicativas. As variáveis obtidas foram tabuladas em arquivo na extensão *.csv*, cujas linhas correspondiam as 85 UAs (repetições), dando origem a base de dados para construção dos modelos.

A fim de evitar o fenômeno de sobreajuste, a base de dados foi dividida de modo aleatório em 70% (59 UAs) para treinamento/ajuste dos modelos e 30% (26 UAs) para teste/validação destes. Segundo Corrales et al. (2015), a simples validação cruzada, é uma boa prática estatística de avaliação, mas não uma garantia, pois os modelos podem ser supertreinados, de modo que o critério definitivo é medir o poder preditivo a partir de novos dados, isto é, dados que nunca foram "vistos" pelo modelo em nenhum estágio de sua construção.

A organização do banco de dados, bem como o processo de modelagem foram implementados em linguagem R (R CORE TEAM, 2018) no ambiente da interface gráfica RStudio® 1.1.463 (RSTUDIO TEAM, 2016). Os modelos contemplaram quatro métodos preditivos distintos a seguir listados:

- Modelo Linear Generalizado – *Generalized Linear Model* – ***GLM***[2] (NELDER e WEDDERBURN, 1972)*;*
- Floresta Aleatória - *Random Forest* – ***RF*** (BREIMAN, 2001);,
- Máquina de Vetor de Suporte - *Suport Vector Machine* – ***SVM*** (CORTES et al., 1995);

[2] Para o restante do documento convencionou-se o uso das formas abreviadas *GLM, RF, SVM* e *ANN*, por estarem em consonância com padrões observados em publicações nacionais e internacionais.

- Rede Neural Artificial - *Artificial Network Neural – **ANN** (MCCULLOCH e PITTS, 1943)*.

3.3.3.1 Regressão GLM

A primeira etapa para implementar o modelo GLM foi identificar o modelo teórico de distribuição de probabilidade da biomassa estimada, procedimento realizado no *software* EasyFit 5.6® (EASYFIT, 2018). O ajuste do modelo GLM também foi precedido pela redução da dimensionalidade dos dados, por meio da correlação de *Spearman* e Análise de Componentes Principais – ACP, detalhadas na sequência:

a) Seleção prévia das variáveis pela análise da correlação segundo método de *Spearman.* Como critério de seleção, filtrou-se as métricas LiDAR que apresentaram correlação significativa[3] superiores a 0,5 (proporção direta) e inferiores a -0,5 (proporção inversa) com a biomassa acima do solo.
b) Previamente ao procedimento de ACP, em razão da multidimensionalidade das variáveis selecionadas, os dados foram padronizados com média zero e variância unitária.
c) Após a padronização, as variáveis selecionadas foram então submetidas ao teste de esfericidade de *Bartlett*, que de acordo com IBM (2016) analisa a hipótese de que a matriz de correlação das variáveis explicativas é uma matriz identidade, o que indicaria que seus elementos não estão relacionados e, portanto, inadequados para a detecção de sua estrutura via ACP.
d) Realização da ACP por meio da função *PCA* do pacote *FactoMineR* (LE et al., 2008), da linguagem R.
e) Avaliação da importância de cada componente gerada, e das variáveis dentro destas.
f) Como critério de escolha para montagem do modelo GLM, foram selecionadas as métricas que apresentaram contribuição relativa acima da média, dentro da(s) componente(s) selecionada(s).

Com as métricas selecionadas na etapa anterior realizou-se o ajuste do modelo com a função *glm* do pacote *stats*, base do aplicativo R. De forma complementar ao ajuste do modelo

[3] A função *cor.test* no *R* testa a significância da hipótese de que não existe associação entre as variáveis pareadas pelo critério da correlação de Spearman, de modo que para valores de probabilidade *p-value* ≤ 0.05, conclui-se que a correlação não é atribuída ao acaso e possui valor significativo (PIEGORSCH, 2015).

procedeu-se uma Análise de variância ANOVA acompanhada da estatística Qui-quadrado como teste de inferência para detecção das variáveis preditoras que explicam porção significativa dos desvios da regressão GLM, conforme sugerido por Guisan et al. (2002). Seguindo o critério de maior importância na ACP, as variáveis foram individualmente inseridas no modelo GLM até que se obtivesse um ajuste que comportasse o maior número de variáveis significativas, de acordo com a estatística Qui-quadrado.

3.3.3.2 Regressão RF

A obtenção do modelo de estimativa da biomassa acima do solo a partir do algoritmo RF foi por meio do pacote *randonForest* (LIAW e WIENER, 2002), disponível no *software* R. De acordo com seus desenvolvedores, existem dois parâmetros básicos a serem definidos para execução do modelo RF por meio do pacote *randonForest*:

- *ntree*: Número de árvores a serem criadas a partir da seleção das amostras pelo método *bootstrap;*
- *mtry*: Número de preditores a serem amostrados aleatoriamente para montagem de cada um dos nós das árvores.

Após alguns testes, definiu-se o valor de 1000 para o número de árvores e manteve-se o valor *default* do parâmetro *mtry*, que para regressão, corresponde a um terço do número total de variáveis utilizado. Os demais parâmetros de entrada do modelo RF, passíveis de configuração, conforme descrição de Liaw e Wiener (2002), não foram alterados da sua forma padrão de implementação.

Por fim, obteve-se a importância das variáveis preditoras no modelo ajustado, pelo critério da redução média na impureza do nó. Segundo Liaw e Wiener (2002), o critério leva em conta a diminuição total nas impurezas do nó da divisão na variável, medida pela soma residual dos quadrados e calculada em média para todas as árvores.

3.3.3.3 Regressão SVM

Para implementação, utilizou-se a função *svm* do pacote *e1071* (MEYER et al., 2018), do aplicativo R. Adotou-se o valor de 1 para o Custo (C) e o hiperparâmetro *gamma,* definido como o inverso do número de amostras, conforme sugestão de Pedregosa et al. (2011). Entre os diferentes tipos de *kernel* testados (polinomial, gaussiano e sigmoidal), o polinomial foi o que permitiu melhor ajuste aos dados. A importância das variáveis foi obtida pela álgebra matricial

proposta por Chang e Lin (2008), multiplicando a matriz com os vetores de suporte pelo peso atribuído a cada variável de entrada.

3.3.3.4 Regressão ANN

A regressão ANN foi implementada por meio da função *neuralnet* do pacote *neuralnet* (FRITSCH e GUENTHER, 2016) do R. O único parâmetro de entrada configurado foi o número de neurônios internos na camada oculta do modelo, seguindo o método proposto por Ikeda e Shibata (2009), na forma da Equação 2:

$$N^h = \sqrt{N^i * N^0} \quad (2)$$

Onde N^h corresponde ao número de neurônios internos, N^i o número de *inputs* ou variáveis de entrada (variáveis explicativas), e N^o, o número de *outputs*, que corresponde a biomassa estimada.

Dessa forma, definiu-se como 9 o número de neurônios ocultos, considerando o resultado da equação acima, para 87 variáveis explicativas. Já a importância de cada variável foi obtida pela função *garson* do pacote *NeuralNetTools* (BECK, 2018), do R.

3.3.4 Verificação do ajuste e validação dos modelos

Ao final do processo de modelagem, para avaliação do grau de ajuste, os modelos de regressão foram submetidos aos indicadores apresentados na Tabela 4.

Tabela 4 – Indicadores estatísticos utilizados na avaliação do ajuste dos modelos

(continua)

Indicador	Expressão matemática
Coeficiente de determinação R^2	$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$
Raiz do erro médio quadrático RMSE	$RMSE\ (Mg/ha) = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$
Erro padrão da estimativa percentual Syx (%)	$Syx\ (\%) = \frac{\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}}{\bar{y}} * 100$

Tabela 4 – Indicadores estatísticos utilizados na avaliação do ajuste dos modelos

(conclusão)

Indicador	**Expressão matemática**
Tendência relativa (viés relativo) Bias (%)	$\text{Bias } (\%) = \frac{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)}{n}}{\frac{\sum_{i=1}^{n} \hat{y}_i}{n}} * 100$
Desvio médio relativo DM (%)	$\text{DM } (\%) = \frac{\sum_{i=1}^{n} \frac{\hat{y}_i - y_i}{y_i}}{n} * 100$

Onde: n é o número de unidades amostrais, y_i é o valor observado para a biomassa na unidade amostral i, $\bar{y}$ é a biomassa média para a unidade amostral i, e $\hat{y}_i$ é o valor predito para a unidade amostral i.

Fonte: Adaptado de Schneider et al. (2009).

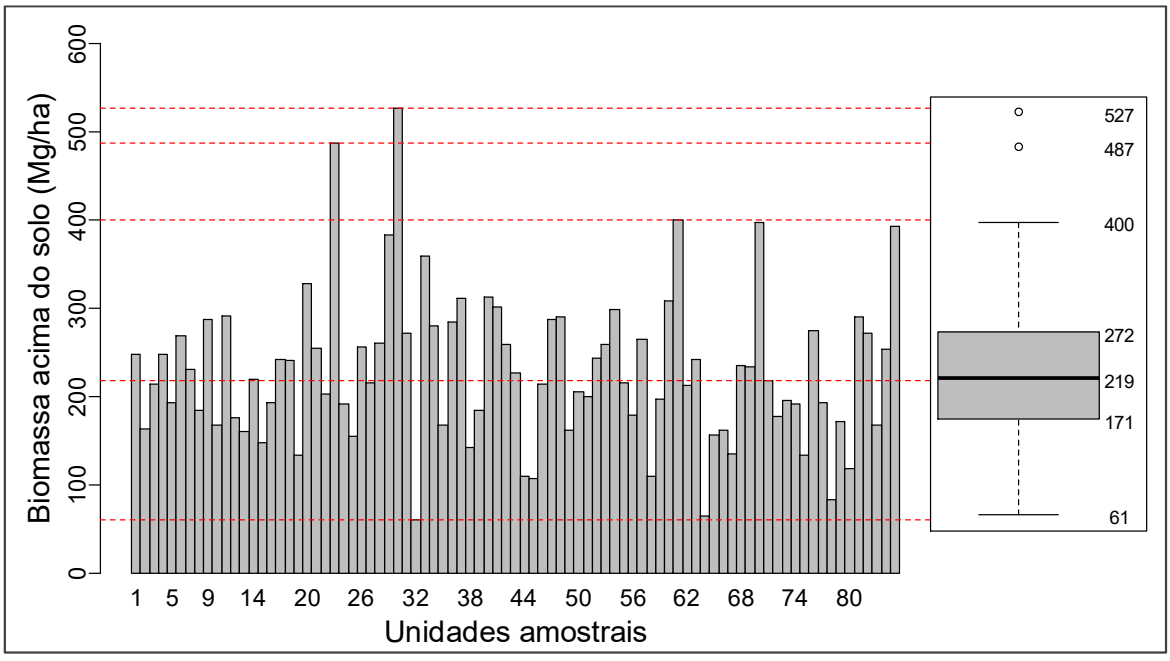
A validação dos modelos, realizada a partir do confronto entre a biomassa predita e observada para as amostras de teste, levou em conta dois elementos estatísticos. A correlação de *Spearman*, executada a partir da função *cor.test* no R, e o teste de *Wilcoxon* para detecção de diferença entre medianas. O *Wilcoxon Rank Sum Test* (equivalente ao *Teste U* de *Mann-Whitney*) testa a hipótese nula na qual as medianas das distribuições de A e B diferem pelo valor de μ, desde que o valor padrão de μ=0, signifique que a hipótese nula é a igualdade das medianas (BEASLEY, 2004).

4 RESULTADOS E DISCUSSÕES

4.1 BIOMASSA NAS PARCELAS INVENTARIADAS

Os dados de biomassa acima do solo estimada com base nas informações do inventário florestal, para todas as unidades amostrais, encontram-se representados na Figura 7. Ainda que de modo exploratório, percebeu-se uma grande variabilidade na sua distribuição, cujo valor médio de 229,57 Mg/ha apresentou um desvio padrão de 84,48 Mg/ha. O desvio elevado explica a grande amplitude observada nos gráficos de barras e *boxplot* (Figura 7). O gráfico boxplot permitiu ainda observar a distribuição das observações nos quartis de modo relativamente simétrico em relação a mediana, bem como a presença de dois possíveis valores discrepantes (*outliers*).

Figura 7 – Distribuição da biomassa acima do solo ao longo das 85 unidades amostrais inventariadas na fazenda Cauaxi, município de Paragominas, no ano de 2014



Fonte: Autor.

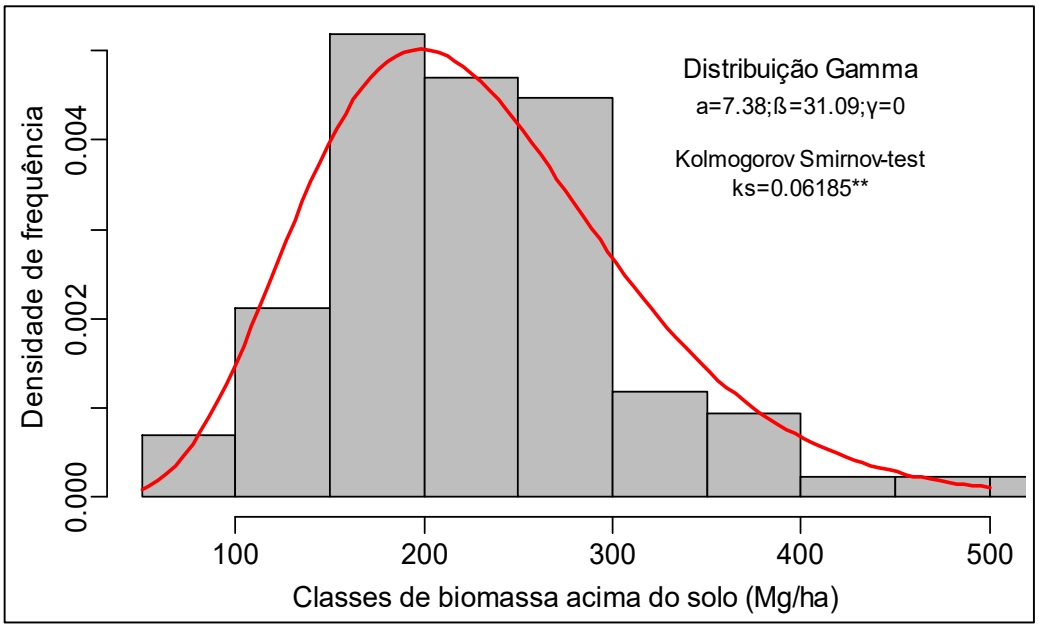
Em estudo anterior para a mesma tipologia florestal, no estado do Amazonas, Matos e Kirchner (2008) registraram valores de biomassa variando entre 272,38 e 402,92 Mg/ha, onde a média encontrada foi de 343,06 Mg/ha, com desvio de 34,45 Mg/ha. A despeito dos métodos de levantamento e dos fatores ambientais dos diferentes locais, o manejo florestal estabelecido por meio da Exploração de Impacto Reduzido – EIR, desde o ano de 2006 (conforme relato de

Silva et al., (2017a)), pode ser o fator que melhor explique a grande heterogeneidade nos valores observados de biomassa na Fazenda Cauaxi.

Tendo em vista a existência de exploração florestal na área de estudo, optou-se pela manutenção dos pontos extremos, já que provavelmente não tenham sido originados de erros sistemáticos, mas sim da característica heterogênea da área, em razão do regime de desbaste seletivo. Característica essa, pertinente ao avaliar os diferentes algoritmos em temos da capacidade de adaptação aos dados e generalização na predição.

A manutenção dos pontos extremos estabeleceu uma característica de dispersão com assimetria positiva ao conjunto de dados da biomassa. O ajuste do modelo teórico de distribuição de probabilidade, realizado no *software* EasyFit 5.6®, confirmou essa suposição. A distribuição *Gamma* foi a que melhor se ajustou aos dados, significativa pela estatística ks=0.062 do teste de *Kolmogorov Smirnov*, com os parâmetros ajustados de α=7,38 e β=31,09 (Figura 8).

Figura 8 – Histograma de dispersão e curva de distribuição ajustada aos dados de biomassa acima do solo nas UAs da Fazenda Cauaxi, município de Paragominas, no ano de 2014



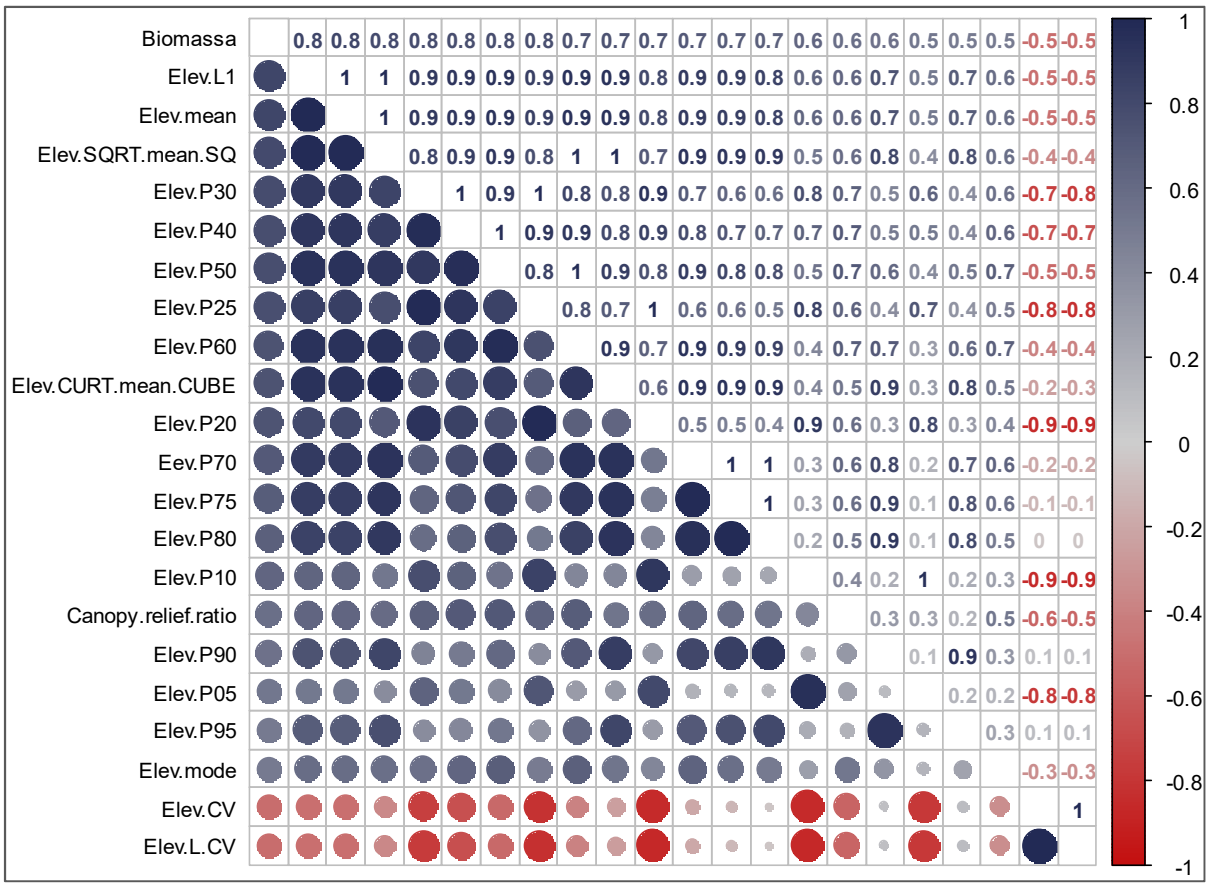
Fonte: Autor.

4.2 ANÁLISE DE COMPONENTES PRINCIPAIS – ACP

A seleção das variáveis pelo critério de correlação de *Spearman* permitiu uma redução expressiva no número de variáveis explicativas. Das 87 métricas LiDAR testadas, 21 apresentaram correlação significativa superior a 0.5 (em módulo), com a biomassa acima do

solo. Como é possível observar na Figura 9, as métricas derivadas da elevação dos pontos foram as que apresentaram maior relação com a biomassa, como é o caso da elevação média (*Elev.mean*), comumente utilizada em estudos que relacionam métricas LiDAR e dados biofísicos em processos de modelagem, como nos estudos apresentados por Popescu et al. (2003), Hawbaker et al. (2009), Sheridan et al. (2014) e Silva. et al. (2017a).

Figura 9 - Correlograma das métricas LiDAR[4] com a biomassa acima do solo



Fonte: Autor.

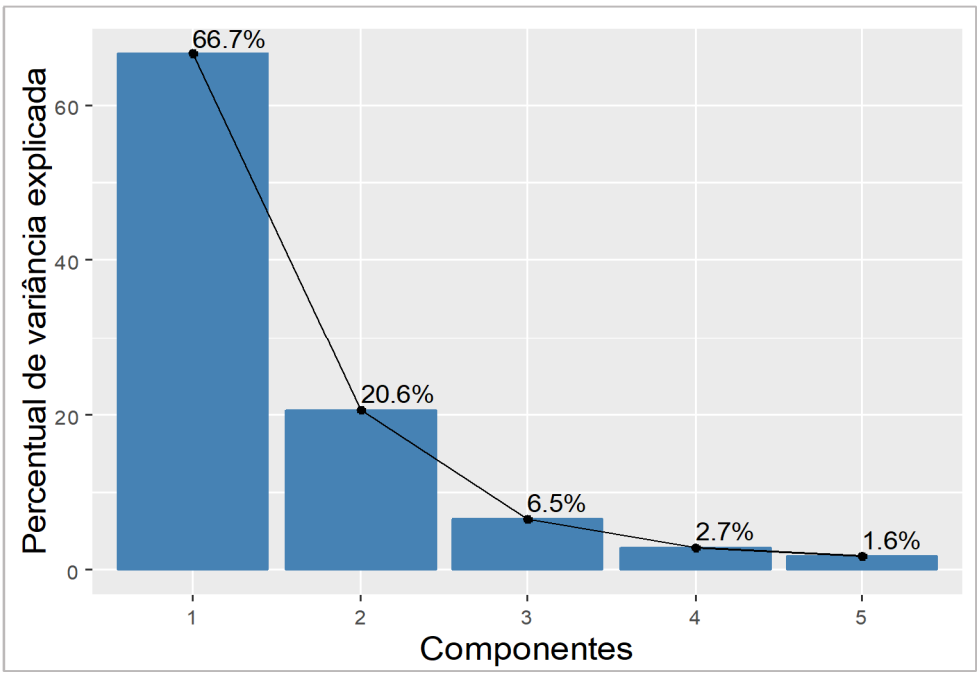
O teste de esfericidade de *Bartlett* para as 21 variáveis selecionadas pelo critério de correlação apresentou um valor de probabilidade *p-value* = 2,2 x 10^{-16}, muito abaixo do nível de significância (0,05), o que, de acordo com IBM (2016), indica que uma análise de componentes principais pode ser útil ao tratamento dos dados. Vicini (2005) relata que a técnica ACP é sensível a correlações pobres entre variáveis, pois, neste caso, as variáveis não apresentarão uma estrutura de ligação entre si, inviabilizando o uso da técnica, que tem como objetivo principal o estudo de conjuntos de variáveis correlacionadas. O resultado do teste de

[4] Uma breve descrição das métricas LiDAR utilizadas nas análises é apresentada no Anexo B.

Bartlett, de certa forma sintetizou em termos de probabilidade de significância, a alta correlação entre variáveis verificadas no correlograma apresentado na Figura 9**.**

Como primeiro produto da ACP obteve-se o percentual de variância retida pelas diferentes componentes. Conforme demonstrado na Figura 10, as componentes CP1 e CP2, somadas responderam por aproximadamente 87% da variabilidade total detectada nas 21 métricas LiDAR selecionadas para as análises.

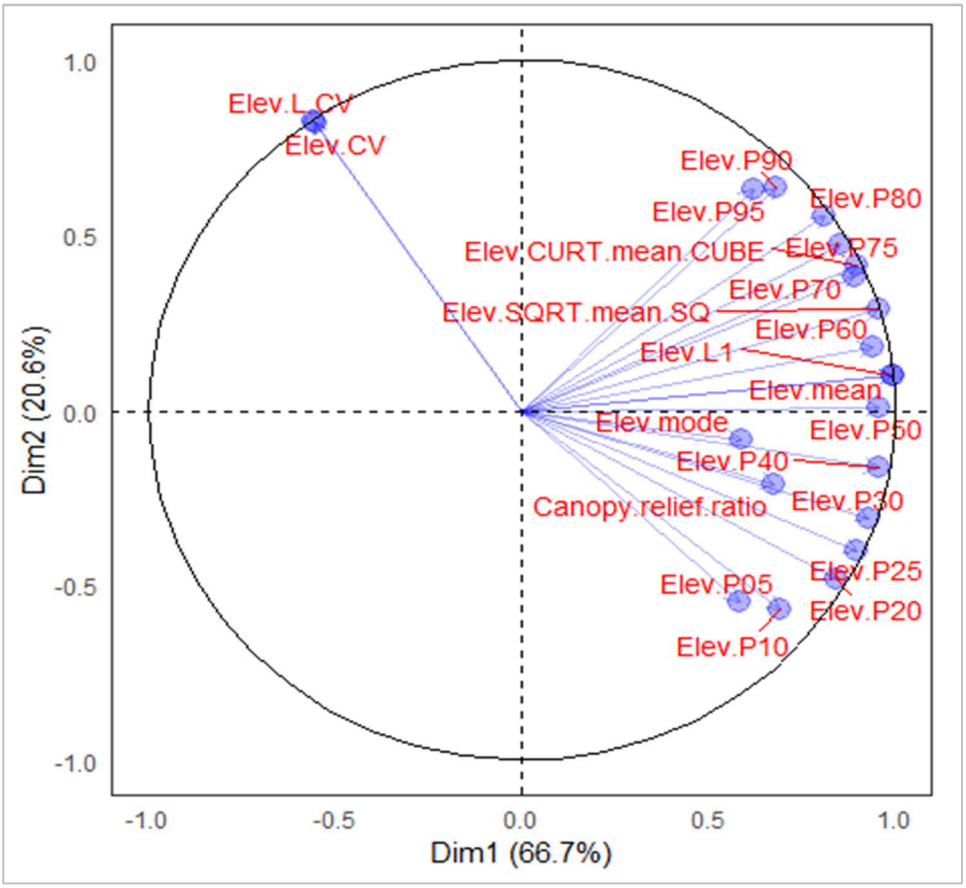
Figura 10 – Percentual da variância global das métricas LiDAR explicada pelas componentes principais



Fonte: Autor.

A Figura 11, traz o mapa fatorial com a correlação entre as métricas LiDAR e as duas primeiras componentes. No contexto da Análise de Componentes Principais, essa correlação é também denominada como carga ou *loading* e pode ser interpretada como a importância de cada variável dentro da componente principal (VICINI, 2005). As variáveis *Elev.mode* e *Elev.P05* apresentaram vetores com menor dimensão, já as variáveis *Elev.L.CV* e *Elev.CV*, embora possuem grande dimensão, sua correlação com a CP1 é negativa. De modo geral, as demais variáveis apresentaram grande correlação com as duas componentes, já que estão distribuídas próximas aos extremos da circunferência. São essas as variáveis que desempenham um papel mais relevante na análise, pois justificam a maior dispersão dos dados (LEONI et al., 2017).

Figura 11 – Mapa fatorial com a distribuição das métricas LiDAR nas duas primeiras componentes principais (CP1 x CP2)



Fonte: Autor.

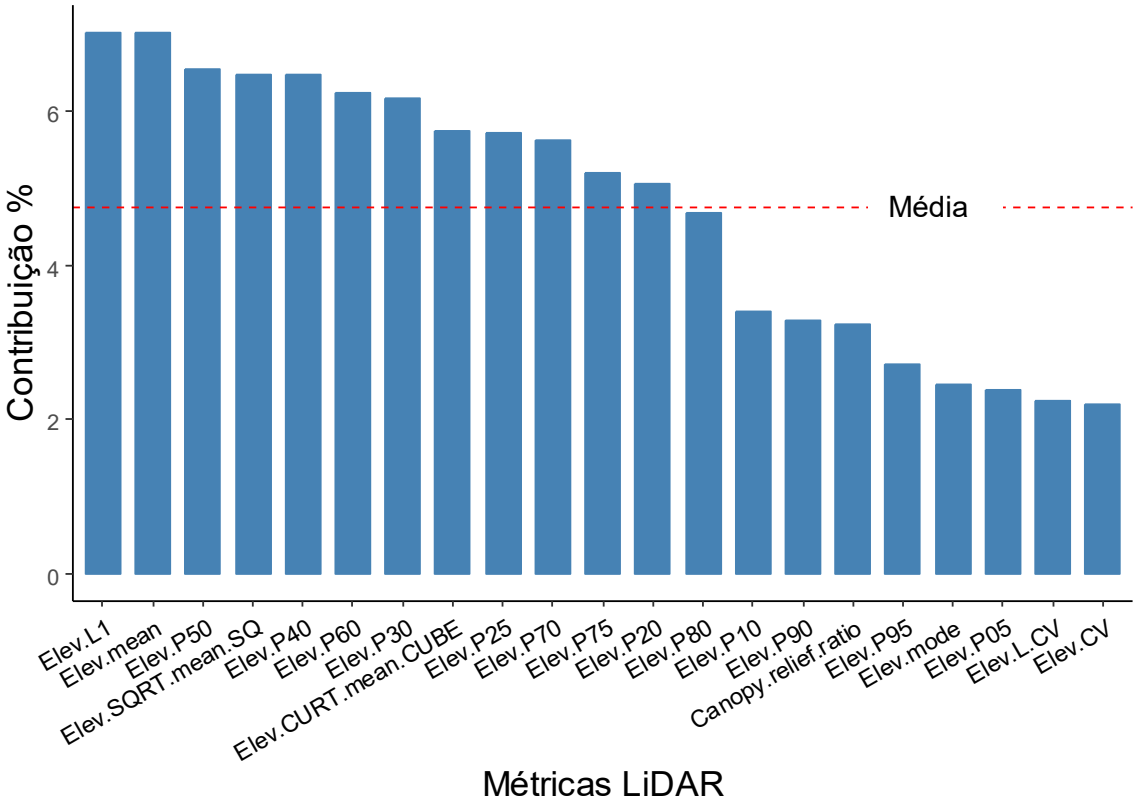
A determinação do número ótimo de componentes a serem retidas no modelo leva em conta a proporção de explicação da variância total, que o modelo de k componentes principais é responsável (HONGYU et al., 2015). Em muitos casos, segundo Johnson e Wichern (2007), adota-se modelos que expliquem pelo menos 80% da variação total. Quanto maior o percentual, melhor o ajuste matemático aos dados, no entanto, um grande número de componentes, pode limitar o processo de redução de variáveis.

Na literatura observou-se a não existência de um limiar fixo em relação ao percentual mínimo de variância explicada. Na aplicação demonstrada em seu livro, Vicini (2005) considerou suficiente 66% distribuída em quatro componentes, já Hongyu et al. (2015) em seu trabalho consideraram que os 68,13% da variação retida nas duas primeiras componentes, resumem efetivamente a variância amostral total e podem ser utilizadas para o estudo do conjunto de dados. Tendo em vista esses exemplos, como critério final para seleção das métricas, optou-se por analisar a importância destas na primeira componente gerada, já que

essa, conforme apresentado na Figura 10, retém aproximadamente 67% da variância total dos dados.

A Figura 12 apresenta a contribuição relativa de cada métrica dentro da primeira componente gerada utilizando a técnica de ACP. Observou-se de modo geral, as variáveis que mais contribuem dentro da componente principal são também as que possuem maior correlação com a biomassa, especialmente *Elev. L1* e *Elev.mean*, conforme demonstrado na Figura 9. Das 21 métricas utilizadas na ACP, 12 apresentaram contribuição relativa acima da média, a citar: *Elev.L1, Elev.mean, Elev.P50, Elev.SQRT.mean.SQ, Elev.P40, Elev.P60, Elev.P30, Elev.CURT.mean.CUBE, Elev.P25, Elev.P70, Elev.P75* e *Elev.P20*.

Figura 12 – Importância de cada métrica LiDAR dentro da componente principal PC1



Fonte: Autor.

4.3 MODELOS AJUSTADOS

O resultado da análise de variância juntamente com o teste Qui-quadrado de verificação da significância das variáveis preditivas do modelo GLM final, é apresentado na Tabela 5. Um indicativo da qualidade da adequação do modelo GLM a distribuição dos dados segundo VIEIRA (2004), é a análise comparativa entre o parâmetro de dispersão da família utilizada e a

dispersão dos resíduos. Segundo esse método, a razão entre o desvio residual e os graus de liberdade do modelo ajustado para as 4 variáveis (2,116413/55 = 0,038480), retrata a dispersão residual, valor este semelhante ao da distribuição teórica ajustada aos dados (0,038656), indicando que o modelo apresentou bom ajuste aos dados de treinamento.

Tabela 5 – Estatísticas do modelo de regressão GLM ajustado com as amostras de treinamento

Variável	**Coeficiente**	**GL**	**Desvio Residual**	**Pr(>Chi)**
Elev.P40	-0,034004	58	3,122387	$2{,}22 \times 10^{-16}$ ***
Elev.P50	0,002439	57	2,973906	0,050011 .
Elev.SQRT.mean.SQ	-0,089217	56	2,294169	0,000027 ***
Elev.mean	0,234592	55	2,116413	0,032001 *
Total	-	59	-	-
Probabilidade da significância: 0 ‘***’ 0,001 ‘**’ 0,01 ‘*’ 0,05 ‘.’				
Parâmetro de dispersão considerado para a família Gamma 0,0386556				

Fonte: Autor.

Para os modelos de aprendizado de máquina, em função da complexidade de sua arquitetura, é apresentado na Tabela 6 apenas uma descrição sucinta dos modelos RF e SVM. O modelo de rede neural ajustado tem parte da sua estrutura representada na Figura 13, na qual estão representadas as variáveis de entrada, os 9 neurônios internos que combinam as variáveis explicativas e convergem até a predição da biomassa. Em razão da grande quantidade de informações não foi possível visualizar os pesos de cada variável preditora, de modo que somente o peso e o viés dos nós internos são visivelvente representados.

A Figura 14, por sua vez, traz a representação da importância atribuída as métricas LiDAR pelos algoritmos de aprendizado de máquina durante o processo de ajuste dos modelos de regressão. Percebe-se que no método RF, a importância é condensada em um menor número de preditores.

Essa espécie de seleção interna de atributos pode ser explicada pelo mecanismo de funcionamento do método RF proposto por Breiman (2001). A cada nó da árvore, um subconjunto de atributos é selecionado aleatoriamente e avaliado, de modo que o melhor atributo é escolhido para dividir o nó. Processo análogo ocorre no método SVM. Segundo Basak et al. (2007), o modelo produzido pelo SVM depende apenas de parte dos dados de treinamento,

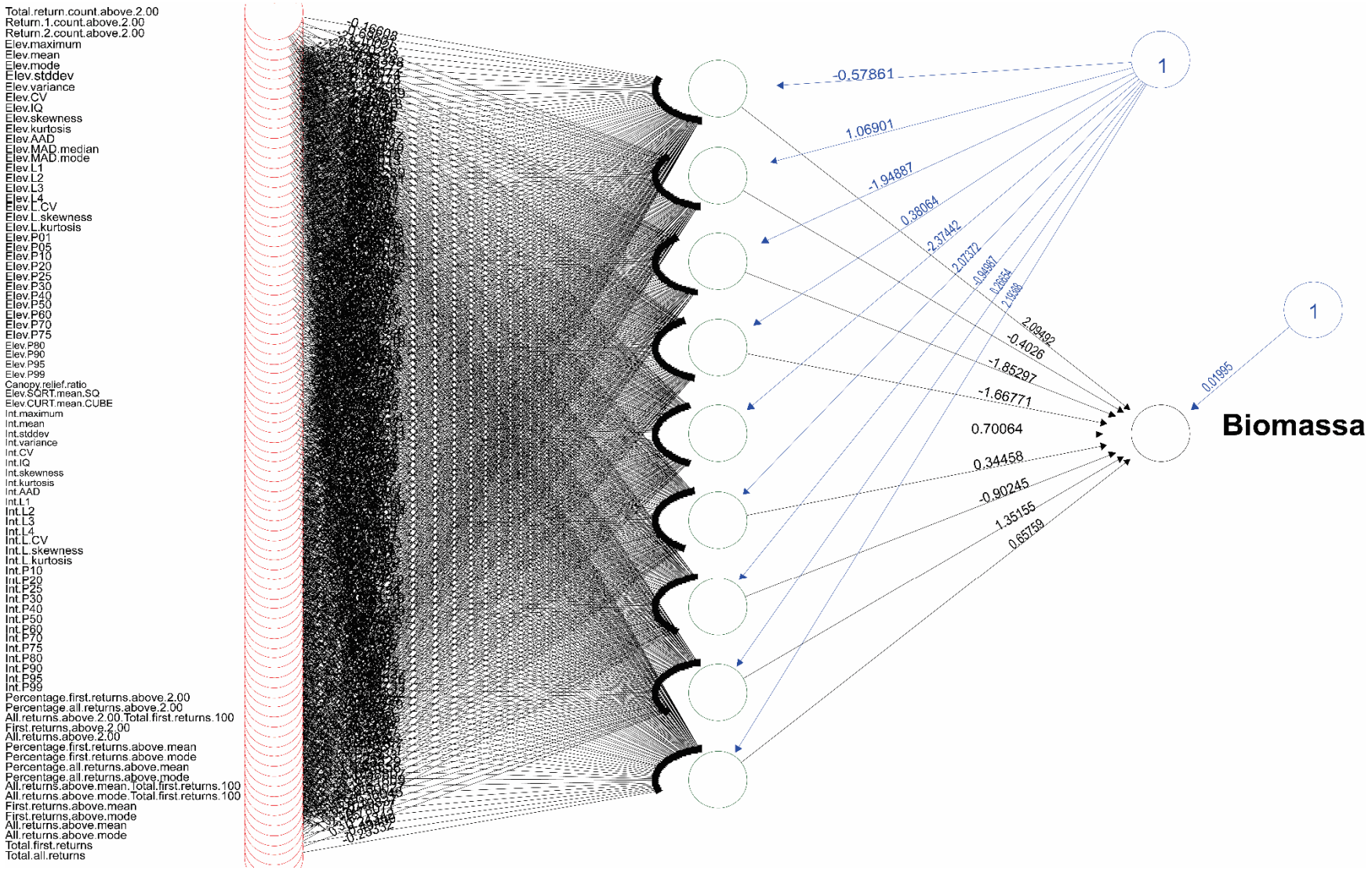
porque a função de custo utilizada para construir o modelo ignora quaisquer dados que estejam próximos (dentro de um limiar de desvio) à previsão do modelo, reduzindo o número de vetores de suporte. Já para as rede neurais *Perceptron*, o erro da estimativa é reduzido pela redistribuição dos pesos das variáveis de entrada, de modo que a importância atribuída aos preditores é mais uniforme.

Tabela 6 – Resumo descritivo dos modelos RF e SVM ajustados no aplicativo R

RF	**SVM**
Número de árvores:1000 Número de variáveis em cada divisão: 29 Resíduo médio quadrático: 2655.89 % Variância explicada: 61.76	*SVM-Kernel*: Polinomial Custo: 1 Grau do polinômio: 3 *gamma*: 0.01694915254 *coef.0*: 0 *epsilon*: 0.1 Número de vetores de suporte: 53

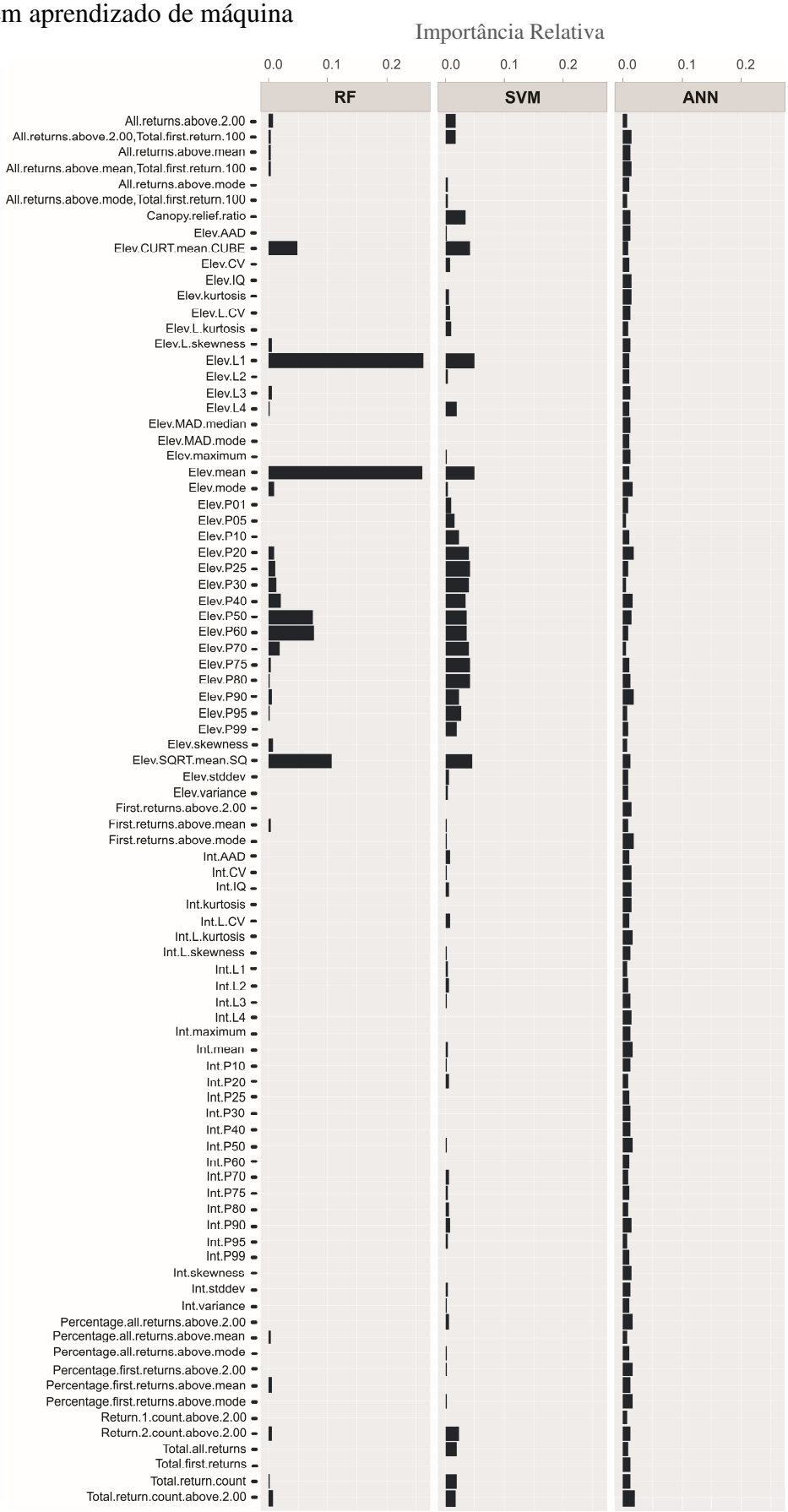
Fonte: Autor.

Figura 13 - Estrutura da rede neural ajustada aos dados de treinamento



Fonte: Autor.

Figura 14 – Importância relativa das variáveis explicativas no ajuste dos modelos de regressão baseados em aprendizado de máquina



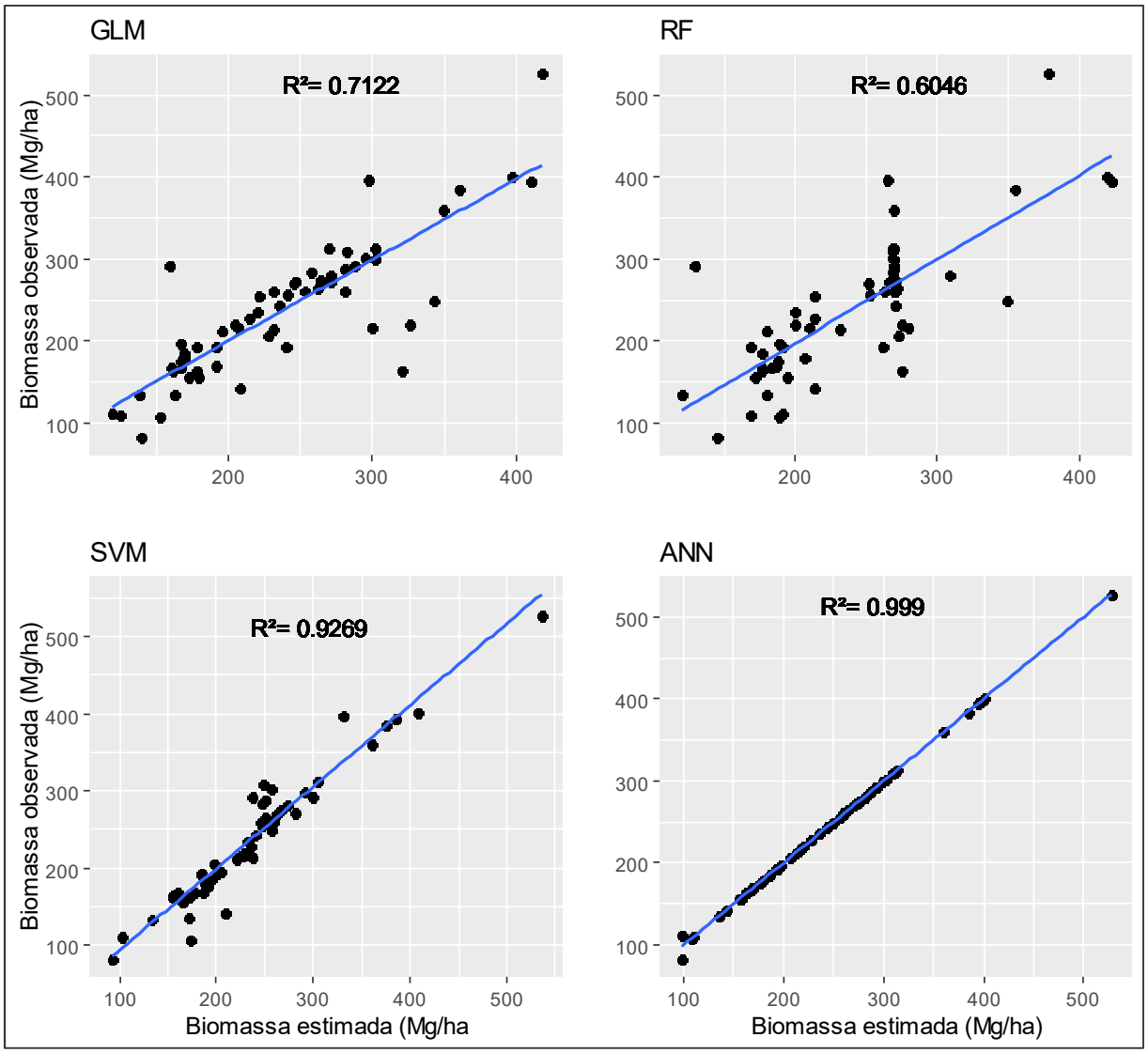
Fonte: Autor.

4.4 ANÁLISE COMPARATIVA DOS MODELOS

4.4.1 Avaliação do ajuste

O primeiro elemento comparativo entre os diferentes modelos de regressão utilizados encontra-se representado na Figura 15. Ao analisar o grau de associação entre as estimativas de biomassa das parcelas de inventário florestal e as estimativas obtidas pelo processamento dos dados LiDAR, percebeu-se que o modelo ANN apresentou melhor performance em termos de ajuste aos dados de treinamento. O percentual de biomassa observada explicada pela regressão ANN foi de 99,9%, seguido pelo modelo SVM (92,3%), GLM (71,2%) e RF (60,5%).

Figura 15 – Relação entre a biomassa acima do solo observada e as estimativas dos diferentes modelos para as 56 unidades amostrais de treinamento

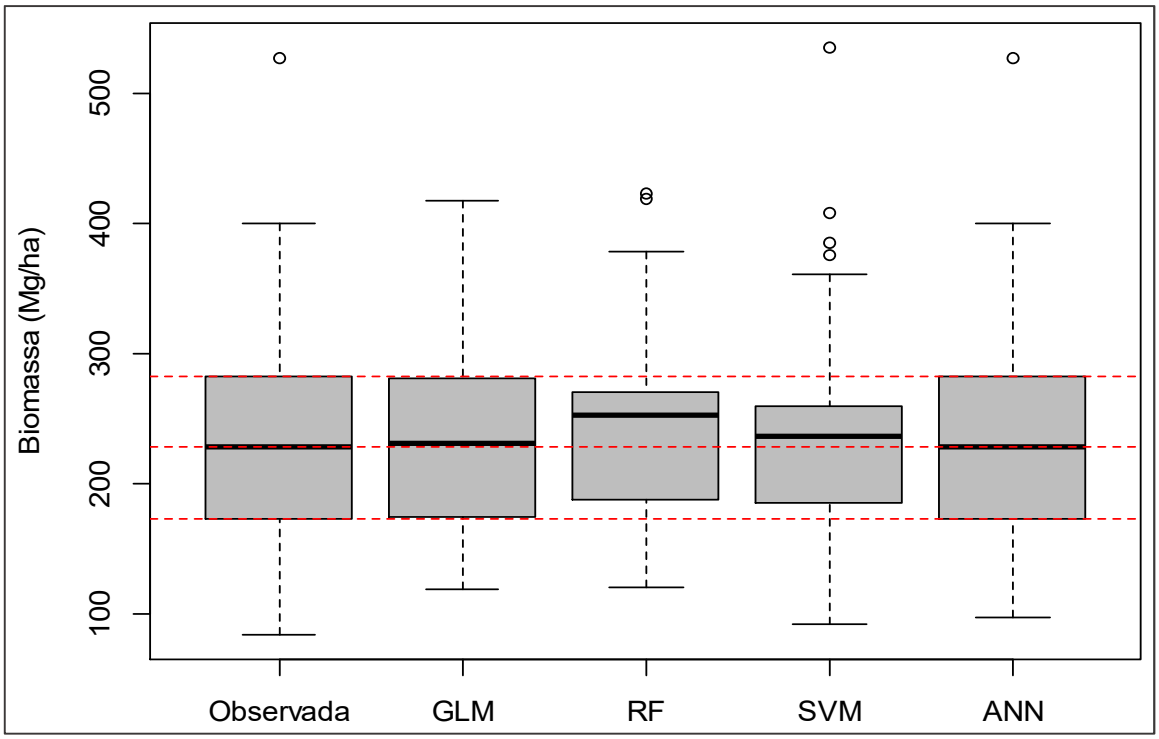


Fonte: Autor.

Alguns estudos preditivos de biomassa em áreas da floresta amazônica utilizando LiDAR, apresentam resultados semelhantes aos modelos RF e GLM. Andersen et al. (2014), obtiveram ajuste de 0,70, em termos de R^2, utilizando regressão linear múltipla nas estimativas em áreas de exploração seletiva no município de Rio Branco/AC. Chen et al. (2016), alcançou um R^2 de 0,94 combinando o uso do método *stepwise* para seleção de métricas e na sequência modelos não lineares com estratificação dos dados, em áreas de sistema agroflorestal no município de Tomé-Açu/PA. Para o mesmo local em Tomé-Açu, incluindo parcelas em áreas de floresta secundária, Feng et al. (2017), encontraram ajustes na ordem de 0,97 utilizando o algoritmo RF e 0,89 para o modelo de regressão SVM.

Os algoritmos ANN e SVM caracterizaram melhor a distribuição dos dados, ao ponto de conseguirem reproduzir um ponto discrepante nos valores de biomassa que compunha a amostra de treinamento, conforme observa-se na Figura 16. Em termos de tendência central, ligeiramente inferior ao modelo ANN, a regressão GLM apresentou mediana mais próxima em relação à observada. Ainda que a distribuição das estimativas SVM tenha apresentado menor amplitude interquartílica, sua assimetria positiva englobou toda a amplitude dos dados observados, o que explica o melhor ajuste em relação ao modelo GLM.

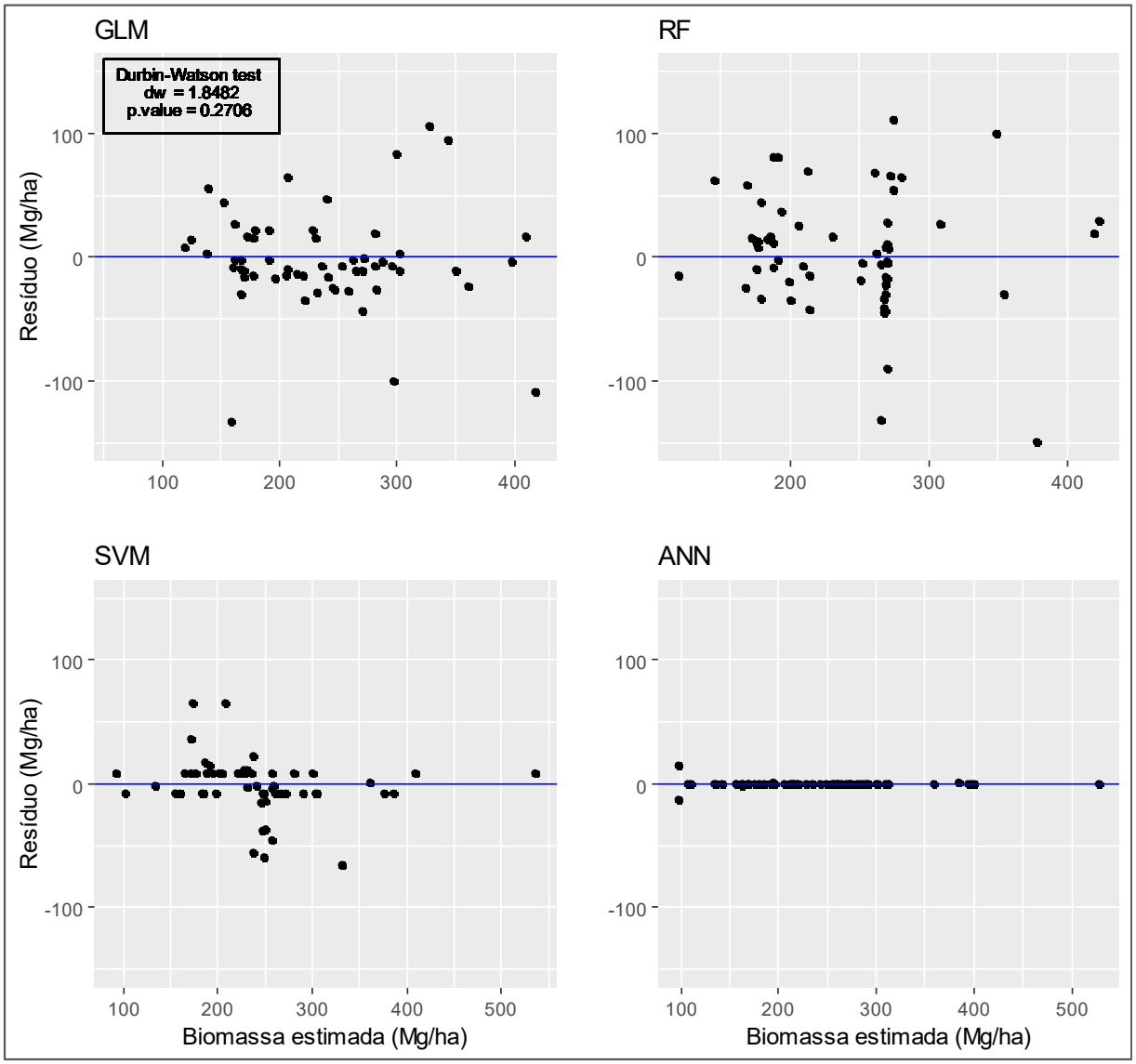
Figura 16 – Representação *boxplot* da distribuição de biomassa observada e estimada pelos diferentes modelos de regressão nas unidades amostrais de treinamento



Fonte: Autor.

O mesmo padrão de qualidade de ajuste é observado ao avaliar a distribuição residual de cada modelo (Figura 17). O modelo ANN foi o que apresentou distribuição mais uniforme, com baixa amplitude e sem tendenciosidade. Já para os modelos RF e GLM observa-se uma amplitude absoluta que extrapola valores entre -100 e 100 Mg/ha para biomassa, o que ratifica a condição de menor ajuste desses modelos. Embora, visualmente observa-se grande dispersão residual para o modelo GLM, o resultado do teste de Durbin-Watson não indicou a presença correlação em série. O valor da estatística DW de 1.84 e p-valor de 0.2706, permitem aceitar a hipótese de independência residual dos dados, condição necessária ao uso das predições GLM com segurança estatística (MCCULLAGH e NELDER, 1989).

Figura 17 – Distribuição dos resíduos em função da biomassa acima do solo estimada pelos diferentes modelos de regressão ajustados



Fonte: Autor.

As observações realizadas em relação ao ajuste dos modelos a partir da análise gráfica são confirmadas pelas informações da Tabela 7. O desempenho superior em todos os indicadores da qualidade de ajuste, conferem ao modelo ANN a condição de melhor adequação aos dados de treino. Em contrapartida, com o RMSE na ordem de 52 Mg/ha e Syx de 22%, o modelo RF, apresentou menor capacidade de "aprendizagem" com as informações fornecidas na etapa de treinamento.

Tabela 7 – Ranqueamento e estatísticas de ajuste dos diferentes modelos de regressão utilizados na predição de biomassa acima do solo a partir das métricas LiDAR

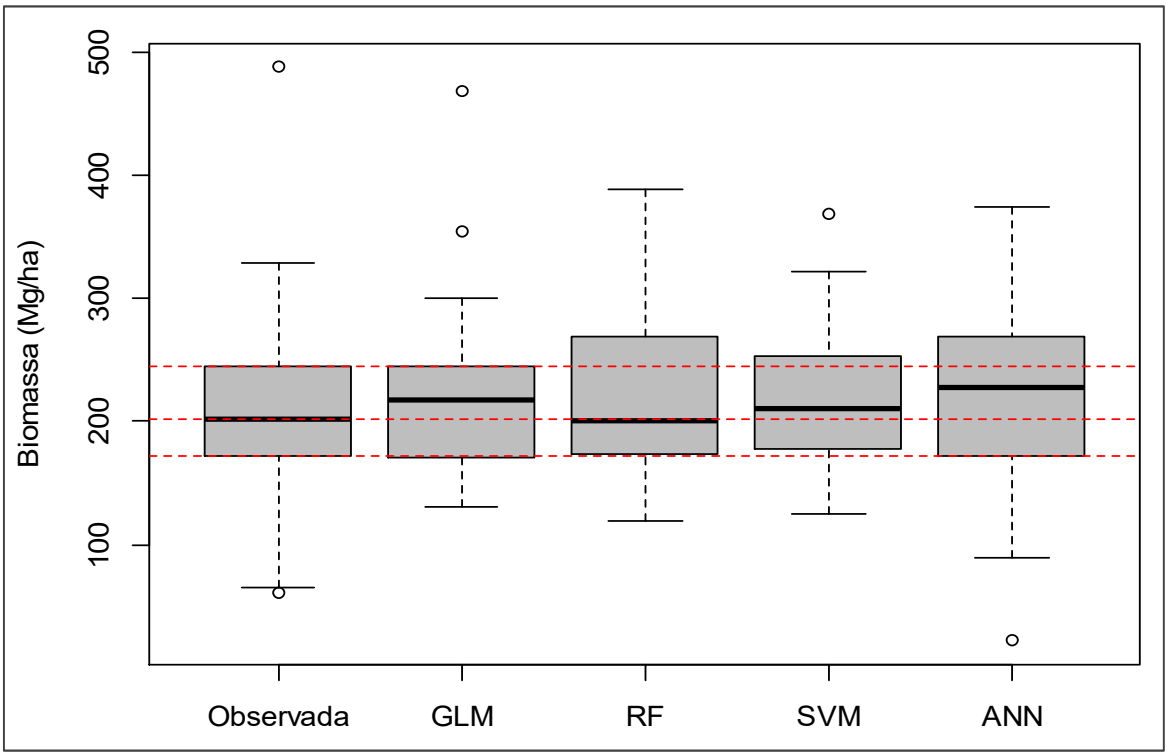
Método	R^2	RMSE (Mg/ha)	Syx (%)	BIAS (%)	DM (%)	Ranking
ANN	0.9995	1.9268	0.8125	-0.0059	0.0007	1
SVM	0.9269	22.5264	9.4985	-0.7589	1.0707	2
GLM	0.7122	44.7075	18.8515	0.0780	3.8210	3
RF	0.6046	52.3954	22.0932	0.5553	6.3194	4

Fonte: Autor.

4.4.2 Validação dos modelos

A distribuição da biomassa estimada para as 26 unidades amostrais destinadas a validação dos modelos encontra-se representada pelos gráficos *boxplot* da Figura 18. Os modelos GLM e RF que haviam apresentado os piores indicadores em relação ao ajuste, mostraram desempenho superior em termos de validação. A amplitude entre o primeiro e terceiro quartil da distribuição GLM coincide com a biomassa observada, o mesmo se aplica as estimativas RF em relação à mediana. Já o modelo ANN apresentou piora no desempenho ao ser aplicado ao conjunto de dados teste, a distribuição de suas estimativas é a que mais distoa dos valores observados.

Figura 18 – Gráficos *boxplot* da biomassa observada e estimada pelos diferentes modelos de regressão nas unidades amostrais de validação



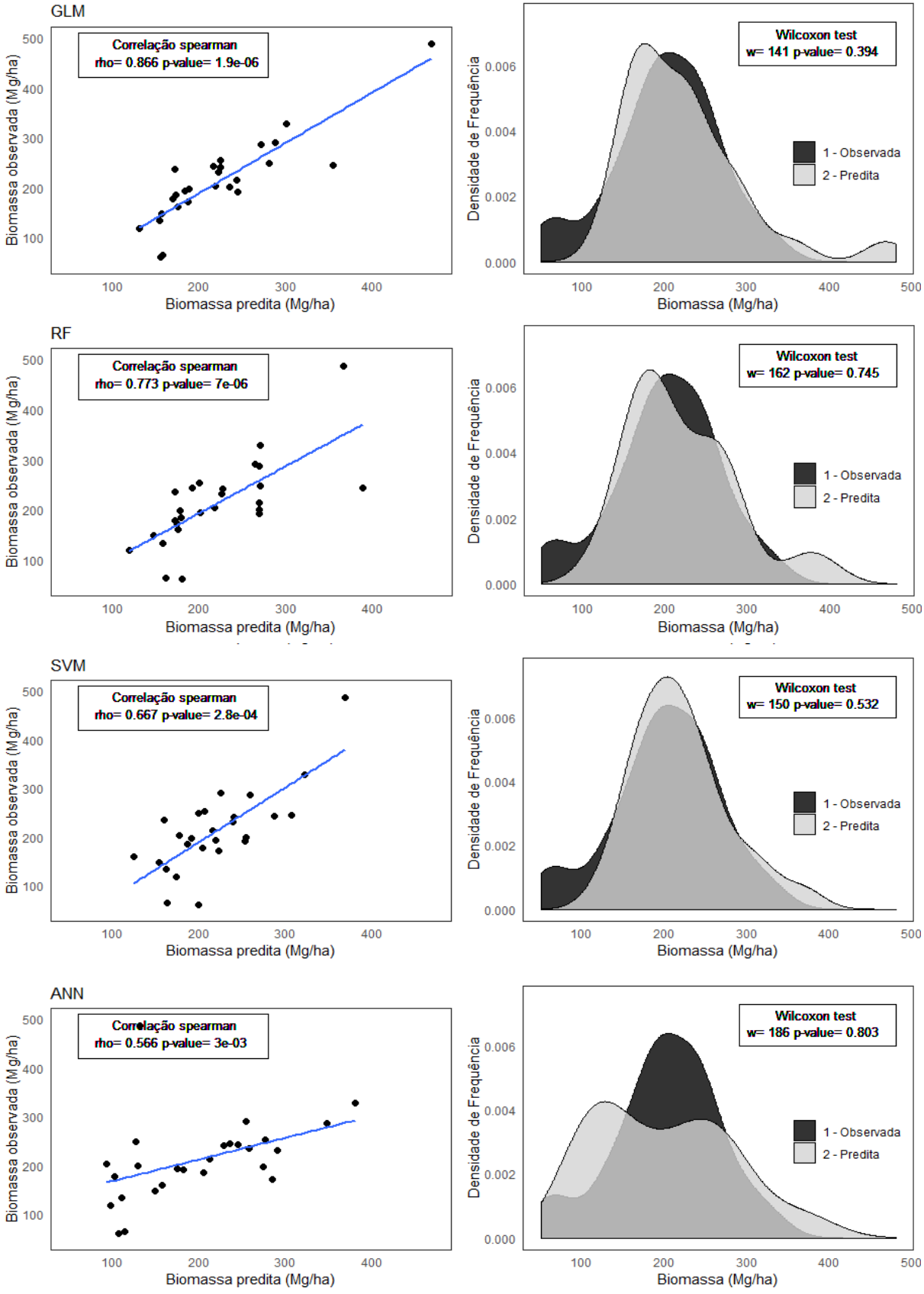
Fonte: Autor.

De modo geral, os modelos apresentaram estimativas com boa aproximação em relação à mediana dos valores observados, um indicativo sobre a adequação dos modelos as estatísticas de validação. Esse diagnóstico prévio é ratificado pela análise da correlação de *Spearman* e para o teste de *Wilcoxon Rank Sum Test* para as amostras pareadas de biomassa observada e estimada pelos modelos (Figura 19).

A hipótese de que a associação entre as variáveis pareadas é devida ao acaso foi refutada a 95% de probabilidade, de modo que a correlação de *Spearman* tomada para todas as estimativas é estatísticamente válida. Em relação ao teste de *Wilcoxon*, em todos os casos, o valor da probabilidade do teste (p-valor), foi superior ao limiar de aceitação, 0,05, o que indica que a mediana dos valores estimados e observados de biomassa não difere significativamente a 95% de probabilidade.

Visualmente, percebe-se que tanto a dispersão dos pontos, como as curvas de densidade de frequência indicam que os modelos RF e GLM se adequaram melhor ao conjunto de dados de validação. A grosso modo, a correlação apresentou comportamento inversamente proporcional ao grau de ajuste dos modelos na etapa de treinamento, variando de 0,57 a 0,87 para os modelos ANN e GLM respectivamente.

Figura 19 – Análise da correlação e das curvas de distribuição da biomassa observada e estimada pelos diferentes modelos de regressão nas unidades amostrais de validação



Fonte: Autor.

A análise das estatísticas de validação permite inferir que modelo SVM e especialmente o ANN perderam desempenho na etapa de validação das estimativas. O que pode ser indício da presença do fenômeno conhecido como sobreajuste ou *overfitting*, considerado o grande problema das técnicas de aprendizado de máquina (DOMINGOS, 2012; YEOM et al., NARASINGARAO et al., 2018). De acordo com Facure (2017), o processo de *overfitting* ocorre nos casos em que o modelo além de aprender as regularidades, capta também o ruído presente nos dados de treino, de modo que o erro na etapa de treinamento é extremamente baixo, enquanto o erro de generalização (nas amostras de teste/validação), é alto.

A arquitetura das redes neurais funcionando em um sistema de balanceamento de pesos entre as entradas e saídas, minimizando o erro de predição, pode ser ao mesmo tempo o que explique o melhor desempenho do modelo ANN no ajuste aos dados de treinamento, mas também um fator determinante para ocorrência do processo de *overfitting*. Para Srivastava et al. (2014), esse padrão de aprendizagem por retropropagação acumula co-adaptações frágeis que funcionam bem para os dados de treinamento mas não generaliza dados não vistos. Essa minimização do erro de modo empírico faz com que, comparativamente, os modelos ANN estejam mais propensos ao *overfitting* do que os SVM (YE, 2014).

A escolha inadequada de alguns parâmetros, especialmente do *kernel*, é o principal fator causador de sobreajuste nos modelos SVM (HAN e JIANG, 2014). Ye et al. (2017) relatam que o custo (C) é uma hiperparâmetro que pode ser ajustado para que o modelo SVM treinado não possua a característica de sobreajuste. No entanto, Ye (2014) menciona que custo pode variar de valores muito altos a 0,001, de modo que, havendo necessidade, o ajuste fino deve ser realizado de modo empírico em cada conjunto de dados.

O modelo RF foi o segundo melhor na validação das estimativas, superou o desempenho dos outros dois algoritmos de ML (SVM e ANN), embora tenha apresentado o menor ajuste entre os quatro modelos testados. De acordo com Lucas (2011), a técnica RF apresenta excelentes características de precisão, generalização para outras amostras que não aquelas em que o modelo foi treinado. A explicação para Lopes et al. (2017), é a utilização da técnica *bagging* no processo de treinamento dos modelos RF, o que além de reduzir a variância ajuda a evitar o *overfitting.*

Por sua vez, o modelo GLM foi o que apresentou maior capacidade de generalização, já que apresentou maior correlação com a biomassa observada do subconjunto de treinamento. Em sua investigação sobre a riqueza de espécies arbóreas em áreas de floresta nativa no Chile, Lopatin et al. (2016), atribuíram o melhor desempenho do modelo GLM em relação ao RF, ao reconhecimento prévio do modelo teórico específico que caracteriza a distribuição dos dados.

De acordo com Srivastava et al. (2014), modelos de aprendizado de máquina com grande número de parâmetros são poderosos em termos de predição, mas podem sofrer seriamente com *overfitting*. Sob essa ótica, percebe-se uma certa correspondência entre a importância atribuída as métricas LiDAR (Figura 14) e o desempenho dos algoritmos. Ao diluir a importância das variáveis explicativas de modo mais uniforme, o SVM e especialmente o método ANN apresentaram grande performance na etapa de treinamento dos modelos, ao passo que para validação, houve perda de desempenho. Já o algoritmo RF que condensou a importância em poucas variáveis, teve maior capacidade de generalização. Essa análise pode ser estendida ao modelo GLM, cuja redução prévia das variáveis explicativas, tendo como critério fundamental o grau de associação destas com a biomassa, pode ter resultado no desempenho superior da regressão generalizada na etapa de validação.

Para McClure (2017), sob a perspectiva da dimensionalidade dos dados, a alta variação entre o erro de treino e teste pode ser resolvida reduzindo o número de recursos (variáveis explicativas) nos dados ou aumentando o número de exemplos (unidades amostrais). Nesse sentido, em estudos de redes neurais, NarasingaRao et al. (2018), propõem a adição de dados, redução da complexidade da arquitetura, bem como o uso de arquiteturas internas com maior poder de generalização.

Embora, possa haver alguns indícios de *overffiting* nos modelos ANN e SVM, é importante frizar que estes superaram o critério estatístico de validação das estimativas. De modo que os modelos de ML conseguiram retratar a estrutura dos dados e gerar predições consistente e acuradas, mesmo com a provável existência de variáveis explicativas redundantes e/ou sem uma relação estreita com a biomassa.

5 CONCLUSÃO

As informações derivadas do levantamento LiDAR aerotransportado mostraram-se eficientes ao serem cruzadas com dados de campo obtidos via inventário florestal no processo de modelagem de biomassa acima do solo em floresta tropical. De forma geral, os modelos conseguiram diagnosticar bem a estrutura dos dados e fazer predições satisfatórias para um ambiente naturalmente heterogêneo, mas que tem sua complexidade estrutural e fisionômica potencialmente amplificada pela presença da exploração madeireira seletiva na área de estudo.

Os dados de biomassa obtidos pela equação ajustada de Chave et al (2014) mostraram-se em consonância com estudos relacionados para o mesmo tipo de formação florestal, a despeito das questões de sítio. A amplitude e o desvio padrão registrados, retrataram a grande variabilidade presente na área, provavelmente em decorrência do regime de manejo seletivo.

Os modelos de aprendizado de máquina ANN e SVM foram os que apresentaram melhor desempenho em termos de ajuste aos dados, explicando melhor a variabilidade da biomassa a campo. O desempenho superior em todos os indicadores de qualidade do ajuste, R^2, RMSE, Syx, BIAS e DM, conferiram ao modelo ANN a condição de melhor adequação aos dados de treino, seguido dos modelos SVM, GLM e RF. Já na etapa de validação, os modelos com menor qualidade de ajuste, apresentaram maior capacidade de generalização. O modelo RF com pior performance no ajuste dos dados, apresentou o segundo melhor resultado para o índice de correlação de *Spearman* entre a biomassa predita e observada nas amostras de teste, ficando abaixo apenas do modelo GLM. Já os modelos SVM e especialmente ANN, perderam performance em relação aos demais ao realizar predições com a nova base de dados.

A capacidade inferior de generalização dos modelos ANN e SVM na etapa de validação é um indício da presença do fenômeno de *overfitting*, cuja causa provável tenha sido o grande número de variáveis preditoras utilizadas. Essa relação ficou mais evidente ao analisar a importância dada as variáveis explicativas pelos diferentes modelos. Ao condensar a importância em um conjunto menor de variáveis preditoras, o método RF teve o pior ajuste, entretanto apresentou boa capacidade de generalização. Ao passo que o algoritmo ANN ao distribuir a importância de forma mais equânime, obteve grande desempenho na etapa de treinamento, mas teve a pior performance na validação. Já em relação ao modelo GLM, a seleção de variáveis preditoras via correlação de *Spearman* e Análise de Componentes Principais – ACP, permitiram o melhor desempenho na etapa de validação das estimativas e performance superior ao RF em termos de ajuste.

Por fim, todos os modelos superaram o critério estatístico de *Wilcoxon Rank Sum Test* de validação das estimativas. Dessa forma, mesmo com a provável existência de variáveis explicativas redundantes e/ou sem uma relação estreita com a biomassa, os algoritmos de aprendizado de máquina conseguiram detectar e reproduzir bem a estrutura não paramétrica dos dados e gerar predições consistentes e acuradas. Assim, é possível concluir que os modelos de aprendizado de máquina fizeram frente a regressão generalizada, sem a necessidade da aplicação de técnicas de redução da dimensionalidade dos dados, o que conferiu mais agilidade ao processo de modelagem.

RECOMENDAÇÕES FINAIS E PERSPECTIVAS FUTURAS

Em trabalhos futuros, os modelos ajustados e validados servirão de base para estimativas que extrapolem as unidades amostrais e permitam a elaboração de mapas temáticos que quantifiquem a biomassa acima do solo, bem como a sua conversão em carbono estocado, para toda área recoberta pelo levantamento LiDAR na fazenda Cauaxi. Os modelos também poderão ser utilizados para avaliação da dinâmica temporal dos estoques de biomassa e carbono, tendo em vista que, além de 2014, foram realizados levantamentos LiDAR na área de estudo em 2012 e 2017. Essa periodicidade viabiliza o cruzamento dos produtos temáticos espacializados, subsídio para diagnóstico de locais com maior intensidade de desbaste, bem como de regeneração florestal, com a presença de incremento das variáveis biofísicas.

Tendo em vista o grau de antropização da área de estudo, é pertinente observar que a aplicabilidade dos modelos ajustados se restringe a área de abrangência da fazenda Cauaxi. Nesse sentido, mesmo em áreas adjacentes de mesma formação (Floresta Ombrófila Densa Submontana), convém a tomada de novas unidades que ampliem o universo amostral de treinamento dos modelos, especialmente para a regressão ANN, que apresentou menor capacidade de generalização.

Esse trabalho expôs a potencial suscetibilidade dos modelos de aprendizado de máquina ao processo de *overfitting*. Nesse sentido, em abordagens similares, quando detectada maior severidade no sobreajuste, de modo que as predições não atendam aos critérios estatísticos de validação, convém a observação de alguns aspectos, como: Proporção entre o número de repetições e o número de variáveis explicativas dos modelos; Avaliação da necessidade ou não da redução de dimensionalidade dos dados, a partir da seleção de variáveis, e; Proporção entre as amostras de treino e validação. Esses são alguns elementos apontados na literatura como reguladores do equilíbrio entre nível de ajuste e capacidade de generalização dos modelos, além naturalmente, do “ajuste fino” dos hiperparâmetros inerentes a cada arquitetura utilizada.

REFERÊNCIAS

ALI, I. et al. Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data. **Remote Sensing**, 4 dez. 2015. v. 7, n. 12, p. 16398–16421. Disponível em: <http://www.mdpi.com/2072-4292/7/12/15841>. Acesso em: 12 set. 2018.

ALMEIDA, O. T.; UHL, C. **Planejamento do uso do solo do município de Paragominas utilizando dados econômicos e ecológicos**. Série Amazônia N° 9- Belém: Imazon, 1998. Disponível em: <http://www.ciflorestas.com.br/arquivos/doc_planejamento__23567.pdf>. Acesso em: 21 mar. 2019.

ANDERSEN, H. E. et al. Monitoring selective logging in western amazonia with repeat lidar flights. **Remote Sensing of Environment**, 2014. v. 151, p. 157–165.

ANDERSON, K. et al. Is waveform worth it? A comparison of LiDAR approaches for vegetation and landscape characterization. **Remote Sensing in Ecology and Conservation**, 2016. v. 2, n. 1, p. 5–15.

ANTHOM, R. et al. Artificial Intelligence Models to Estimate Biomass of Tropical Forest Trees. **Polibits**, 2017. v. 56, n. 1, p. 29–37.

ASNER, G. P.; KELLER, M.; SILVA, J. N. M. Spatial and temporal dynamics of forest canopy gaps following selective logging in the eastern Amazon. **Global Change Biology**, 2004. v. 10, n. 5, p. 765–783. Disponível em: <http://doi.wiley.com/10.1111/j.1529-8817.2003.00756.x>. Acesso em: 7 nov. 2017.

ASNER, G. P. et al. Selective logging in the Brazilian Amazon. **Science**, 2005. v. 310, n. 5747, p. 480–482.

ASNER, G. P. et al. Environmental and biotic controls over aboveground biomass throughout a tropical rain forest. **Ecosystems**, 2008. v. 12, n. 2, p. 261–278.

BASAK, D.; PAL, S.; PATRANABIS, D. C. Support Vector Regression. **Neural Information Processing-Letters and Reviews**, 2007. v. 11, n. 10, p. 203–224. Disponível em: <https://pdfs.semanticscholar.org/c5a9/67eaded74a9fc414de4ad5120b0b66acd2c3.pdf>. Acesso em: 4 jan. 2019.

BASTOS, T. X.; SILVA, G. de F. G. da; PACHECO, N. A.; FIGUEIREDO, R. de O. Informações agroclimáticas do município de Paragominas para o planejamento agrícola. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 14., 2006, Florianópolis. **Anais**... Florianópolis: SBMET, 2006.

BATISTA, J. L. **Biometria Florestal segundo o Axioma da Verossimilhança Com Aplicações em Mensuração Florestal**. 2014. 391 f. Tese (Livre- docência). Departamento de Ciências Florestais, Escola Superior de Agricultura “Luiz de Queiroz”. Universidade de São Paulo. Piracicaba. SP, 2014.

BAYER, F. M. **Tese. Modelagem e Inferência em Regressão Beta**. 2011. 115 f. Tese (Doutorado em Estatística). Pós-Graduação em Estatística. Universidade Federal de Pernambuco. Recife. PE, 2011.

BEASLEY, C. R. **Bioestatística usando R: Apostila de exemplos para o Biólogo**. Bragança: Universidade Federal do Pará. Campus de Bragança. Laboratório de Moluscos, 2004. Disponível em: <https://cran.r-project.org/doc/contrib/Beasley-BioestatisticaUsandoR.pdf>. Acesso em: 4 jan. 2019.

BECK, M. W. NeuralNetTools: Visualization and Analysis Tools for Neural Networks. **Journal of Statistical Software**, 2018. v. 85, n. 11, p. 1–20.

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

BURNHAM, K. P.; ANDERSON, D. R. **Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach**. Second ed. Fort Collins: Springer-Verlag, 2002, 488p.

BUSTAMANTE, M. M. C. et al. Toward an integrated monitoring framework to assess the effects of tropical forest degradation and recovery on carbon stocks and biodiversity. **Global Change Biology**, 2016. v. 22, n. 1, p. 92–109. Disponível em: <http://doi.wiley.com/10.1111/gcb.13087>. Acesso em: 7 nov. 2017.

CHANG, Y.-W.; LIN, C.-J. **Feature Ranking Using Linear SVM**. **JMLR: Workshop and Conference Proceedings**. 2008. Disponível em: <http://proceedings.mlr.press/v3/chang08a/chang08a.pdf>. Acesso em: 27 dez. 2018.

CHAVE, J. et al. Improved allometric models to estimate the aboveground biomass of tropical trees. **Global Change Biology**, 2014. v. 20, n. 10, p. 3177–3190. Disponível em: <http://doi.wiley.com/10.1111/gcb.12629>. Acesso em: 17 set. 2018.

CHEN, Q. et al. Modeling and mapping agroforestry aboveground biomass in the Brazilian Amazon using airborne lidar data. **Remote Sensing**, 2016. v. 8, n. 1, p. 1–17.

CHICCO, D. Ten quick tips for machine learning in computational biology. **BioData mining**, 2017. v. 10, p. 35. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/29234465>. Acesso em: 24 dez. 2018.

COOPS, N. C. et al. Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. **Trees - Structure and Function**, 2007. v. 21, n. 3, p. 295–310.

CORDEIRO, G. M.; DEMÉTRIO, C. G. B. **Modelos Lineares Generalizados e Extensões**. Piracicaba/SP. 2008. 392p. Disponível em: <http://www.ufjf.br/clecio_ferreira/files/2013/05/Livro-Gauss-e-Clarice.pdf>. Acesso em: 24 dez. 2018.

CORRALES, M. et al. **Machine Learning: How Much Does It Tell about Protein Folding Rates?** 2015. Disponível em: <http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC4659572&blobtype=pdf>. Acesso em: 31 dez. 2018.

CORTES, C.; VAPNIK, V.; SAITTA, L. Support-Vector Networks Editor. Machine Leaming. **Kluwer Academic Publishers**, 1995. Disponível em: <https://link.springer.com/content/pdf/10.1007%2FBF00994018.pdf>. Acesso em: 10 set. 2018.

DI MAIO, A.; RUDORFF, B. F. T.; MORAES, E. C.; et al. **Sensoriamento Remoto. Formação Continuada de Professores. Curso Astronáutica e Ciências do Espaço**. Ministério da Ciência e Tecnologia (MCT) e Agência Espacial Brasileira (AEB), 2008. 78 p.

D'OLIVEIRA, M. V. N.; FIGUEIREDO, E. O.; PAPA, D. De A. **Uso do Lidar como Ferramenta para o Manejo de Precisão em Florestas Tropicais**. 1.ed. Brasília. Embrapa. 2014, 130 p.

DOBSON, A. J.; BARNETT, A. G. **An Introduction to Generalized Linear Models**. 3. ed.: CHAPMAN & HALL/CRC Texts in Statistical Science Series, 2008, 225 p.

DOMINGOS, P. A few useful things to know about machine learning. **Communications of the ACM**, 2012. v. 55, n. 10, p. 78. Disponível em: <http://dl.acm.org/citation.cfm?doid=2347736.2347755>.

DUARTE, J. F. Dos S.; CARNEIRO, R. S. G. S. **Análise de Vulnerabilidade Erosiva no Município de Paragominas-PA**. São José dos Campos: [s.n.], 2017. Disponível em: <http://wiki.dpi.inpe.br/lib/exe/fetch.php?media=ser300:alunos2017-ser300:grupo_monografia:trabalho_final_-_jessyca_e_rebeca.pdf>. Acesso em: 11 dez. 2018.

EASYFIT. **EasyFit - Distribution Fitting Software**. 2018. Mathwave Data Analysis and Simulation. Disponível em: <http://www.mathwave.com/easyfit-distribution-fitting.html>.

EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária. **Sustainable Landscapes Brazil**. 2016. Disponível em: <https://www.embrapa.br/busca-de-solucoes-tecnologicas/-/produto-servico/3862/paisagens-sustentaveis>. Acesso em: 11 set. 2018.

FACELI, K. et al. **Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina**. LTC Editora. Rio de Janeiro, RJ. 2011, 378 p.

FACURE, M. **Aprendizado de Máquina: Essencial**. 2017. Disponível em: <https://matheusfacure.github.io/AM-Essencial/>. Acesso em: 31 dez. 2018.

FENG, Y. et al. Examining effective use of data sources and modeling algorithms for improving biomass estimation in a moist tropical forest of the Brazilian Amazon. **International Journal of Digital Earth**, 2017. v. 10, n. 10, p. 996–1016. Disponível em: <http://www.tandfonline.com/action/journalInformation?journalCode=tjde20>. Acesso em: 2 jan. 2019.

FERNEDA, E. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ciência da Informação**, abr. 2006. v. 35, n. 1, p. 25–30. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652006000100003&lng=pt&tlng=pt>. Acesso em: 12 jan. 2019.

FISHER, R. A. On the Mathematical Foundations of Theoretical Statistics. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, 1922. v. 222, n. 594–604, p. 309–368. Disponível em: <http://rsta.royalsocietypublishing.org/cgi/doi/10.1098/rsta.1922.0009>. Acesso em: 2 jan. 2019.

FRITSCH, S.; GUENTHER, F. **neuralnet: Training of Neural Networks. R package version 1.33.** 2016. Disponível em: <https://cran.r-project.org/package=neuralnet>.

GIONGO, M. et al. LiDAR: princípios e aplicações florestais. **Pesquisa Florestal Brasileira**, 2010. v. 30, n. 63, p. 231–244.

GUISAN, A.; EDWARDS, T. C.; HASTIE, T. Generalized linear and generalized additive models in studies of species distributions: setting the scene. **Ecological Modelling**, 2002. v. 157, p. 89–100. Disponível em: <http://www.dpi.inpe.br/referata/arq/10_Ilka/GuisanEtAl_EcolModel2002.pdf>. Acesso em: 23 dez. 2018.

HAN, H.; JIANG, X. Overcome support vector machine diagnosis overfitting. **Cancer informatics**, 2014. v. 13, n. Suppl 1, p. 145–58. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/25574125>. Acesso em: 2 jan. 2019.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning**. Second ed. Springer. 2017, 446p.

HAWBAKER, T. J. et al. Improved estimates of forest vegetation structure and biomass with a LiDAR-optimized sampling design. **Journal of Geophysical Research: Biogeosciences**, 1 jun. 2009. v. 114, n. G2, p. n/a-n/a. Disponível em: <http://doi.wiley.com/10.1029/2008JG000870>. Acesso em: 24 dez. 2018.

HONGYU, K. et al. Análise de Componentes Principais: resumo teórico, aplicação e interpretação. **EE&S - Engineering and Science** , 2015. v. 1, n. 5. Disponível em: <http://periodicoscientificos.ufmt.br/ojs/index.php/eng/article/viewFile/3398/2623>. Acesso em: 25 dez. 2018.

HUNTER, M. O. et al. Structural dynamics of tropical moist forest gaps. **PLOS ONE**, 2015. v. 10, n. 7, p. 1–19.

IBAÑEZ, M. M. **Uso de redes neurais nebulosas e florestas aleatórias na classificação de imagens em um projeto de ciência cidadã.** 2016. 96 p. Dissertação (Mestrado em Computação Aplicada). Curso de Pós-Graduação em Computação Aplicada. Instituto Nacional de Pesquisas Espaciais - INPE, 2016.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Geologia da Amazônia Legal.** 2003a. Disponível em: <https://downloads.ibge.gov.br/downloads_geociencias.htm>. Acesso em: 11 dez. 2018.

IBGE. Instituto Brasileiro de Geografia e Estatística. 2003b. **Geomorfologia Amazônia Legal**. Disponível em: <https://downloads.ibge.gov.br/downloads_geociencias.htm>. Acesso em: 11 dez. 2018.

IBGE. Instituto Brasileiro de Geografia e Estatística. 2003c**. Pedologia Amazônia Legal**. Disponível em: <https://downloads.ibge.gov.br/downloads_geociencias.htm>. Acesso em: 11 dez. 2018.

IBGE. Instituto Brasileiro de Geografia E Estatística. **Manual Técnico da Vegetação Brasileira**. Ed. 2. Rio de Janeiro. 2012, 271 p.

IBGE. Instituto Brasileiro de Geografia e Estatística**. IBGE Cidades. Paragominas**. 2017. Disponível em: <https://cidades.ibge.gov.br/brasil/pa/paragominas/historico>. Acesso em: 21 mar. 2019.

IBM. IBM Knowledge Center. **SPSS Statistics V24.0. KMO and Bartlett's Test**. 2016. Disponível em: <https://www.ibm.com/support/knowledgecenter/en/SSLVMB_24.0.0/spss/tutorials/fac_telco_kmo_01.html>. Acesso em: 25 nov. 2018.

IKEDA, K.; SHIBATA, Y. Effect of Number of Hidden Neurons on Learning in Large-Scale Layered Neural Networks. **ICROS-SICE International Joint Conference 2009 (ICCASSICE '09).** 2009, p. 5008–5013.

INPE. Instituto Nacional de Pesquisas Espaciais. **Monitoramento da floresta amazônica é tema de seminário de cooperação franco-brasileira**. 2016. Disponível em: <http://www.inpe.br/noticias/noticia.php?Cod_Noticia=4220>. Acesso em: 13 jan. 2019.

INPE. Instituto Nacional de Pesquisas Espaciais. **PRODES - Monitoramento da Floresta Brasileira por Satélite.** 2018. Disponível em: <http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes>. Acesso em: 7 jan. 2019.

ISA. Instituto Socioambiental. **Atlas de pressões e ameaças às Terras Indígenas na Amazônia brasileira.** 2009. Disponível em: <https://www.socioambiental.org/banco_imagens/pdfs/Atlas.pdf.pdf>. Acesso em: 18 mar. 2019.

JENSEN, J. R. **Sensoriamento Remoto do Ambiente: Uma perspectiva em Recursos Naturais**. São José dos Campos - São Paulo: Paêntese, 2009, 598 p.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Analysis**. 6. ed. Upper Saddle River: Pearson Education, Inc. 2007, 773 p.

JÚNIOR, J. De A. **Métodos de otimização Hiperparamétrica: Um estudo comparativo utilizando árvores de decisão e florestas aleatórias na classificação binária**. 2018. 60 p. Dissertação (Mestrado em Engenharia Elétrica). Programa de Pós-Graduação em Engenharia Elétrica. Universidade Federal de Minas Gerais, 2018.

JUNIOR, L. C. G. **Avaliação automática da qualidade de escrita de resumos científicos em inglês**. 2007. 144 p. Dissertação (Mestado em Ciências da Computação e Matemática Computacional). Instituto de Ciências Matemáticas e de Computação - ICMC-USP. Universidade de São Paulo, 2007.

KOTSCHOUBEY, B. et al. Caracterização e gênese dos depósitos de bauxita da província bauxitífera de Paragominas, noroeste da Bacia do Grajaú, nordeste do Pará/oeste do Maranhão. **In: Caracterização de depósitos minerais em distritos mineiros da amazônia**. Brasília/DF: DNPM - CT/MINERAL – ADIMB. 2005, p 691-782.

LARY, D. J. et al. Geoscience Frontiers Machine learning in geosciences and remote sensing. **Geoscience Frontiers**, 2016. v. 7, n. 1, p. 3–10.

LE, S.; JOSSE, J.; HUSSON, F. FactoMineR: An R Package for Multivariate Analysis. **Journal of Statistical Software**, 2008. v. 25, n. 1, p. 1–18.

LEITOLD, V. et al. Airborne lidar-based estimates of tropical forest structure in complex terrain: opportunities and trade-offs for REDD+. **Carbon Balance and Management**, 2015. v. 10, n. 1, p. 3.

LEITOLD, V. et al. El Niño drought increased canopy turnover in Amazon forests. **New Phytologist**, 2018. v. 219, n. 3, p. 959–971.

LEONI, R. C.; NILO, A. D. S. S.; CORRÊA, S. M. Estatística Multivariada Aplicada ao Estudo da Qualidade do Ar. **Revista Brasileira de Meteorologia**, 2017. v. 32, n. 2, p. 7. Disponível em: <http://dx.doi.org/10.1590/0102-77863220005>. Acesso em: 25 dez. 2018.

LIAW, A.; WIENER, M. **Classification and Regression by RandomForest**. 2002. Disponível em: <https://www.researchgate.net/publication/228451484>. Acesso em: 10 set. 2018.

LONGO, M. et al. Aboveground biomass variability across intact and degraded forests in the Brazilian Amazon. **Global Biogeochemical Cycles**, 2016. v. 30, n. 11, p. 1639–1660.

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**, 2016. v. 173, n. October 2018, p. 200–210.

LOPES, T. D. et al. Aplicação do Algoritmo Random Forest como Classificador de Padrões de Falhas em rolamentos de Motores de Indução. Porto Alegre: XIII Simpósio Brasileiro de Automação Inteligente, out. 2017. In: **Anais**...Disponível em: <https://www.ufrgs.br/sbai17/papers/paper_98.pdf>. Acesso em: 10 nov. 2018.

LORENA, A. C.; CARVALHO, A. C. P. L. F. De. Uma Introdução às Support Vector Machines. **Revista de Informática Teórica e Aplicada**, 2007. v. 14, n. 2, p. 43–67.

LU, X.; GUO, Q.; LI, W.; FLANAGAN, J. A bottom-up approach to segment individual deciduous trees using leaf-off lidar point cloud data. **ISPRS Journal of Photogrammetry and Remote Sensing**, 2014, v. 94, p. 1–12.

LUCAS, L. C. De S. Árvores, Florestas E Sua Função Como Preditores: Uma Aplicação Na Avaliação Do Grau De Maturidade De Empresas. **Revista Pmkt**, 2011. v. 6, n. 1, p. 6–11.

MATOS, F. D. De A.; KIRCHNER, F. F. Estimativa de biomassa da floresta ombrófila densa de terra firme na amazônia central com o satélite IKONOS II. **Floresta**, 2008. v. 38, n. 1. p. 157-171.

MCCLURE, S. **How many training samples are needed to get a reliable model in ML**? 2017. Disponível em: <https://www.quora.com/How-many-training-samples-are-needed-to-get-a-reliable-model-in-ML/answer/Sean-McClure-3?srid=zGgv>. Acesso em: 2 jan. 2019.

MCCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**. 2. ed. Chapman and Hall. 1989, 511 p.

MCCULLOCH, W. S.; PITTS, W. H. A Logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, 1943. v. 5, p. 115–133. Disponível em: <http://www.cse.chalmers.se/~coquand/AUTOMATA/mcp.pdf>. Acesso em: 11 set. 2018.

MCGAUGHEY, R. J. **FUSION/LDV: Software for LIDAR Data Analysis and Visualization (Manual).** U.S. Department of Agriculture, Forest Service**.** 2016, 206 p.

MCGAUGHEY, R. J. **FUSION/LDV: Software for LIDAR Data Analysis and**

Visualization. The Forest Service of the U.S. Department of Agriculture. 2018.

MELO, M. D. **Um processo de mineração de dados para predição de níveis criminais de áreas geográficas**. 2010. 128 p. Dissertação (Mestrado Acadêmico em Ciência da Computação). Universidade Estadual do Ceará, 2010.

MERSCHMANN, L. H. De C. **Classificação probabilística baseada em análise de padrões**. 2007. 103 f. Tese. (Doutorado em Ciência da Computação). Programa de Pós-Graduação em Computação. Universidade Federal Fluminense, 2007.

MEYER, D. et al. **e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-0.** 2018. Disponível em: <https://cran.r-project.org/package=e1071>.

MEYER, V. et al. Canopy Area of Large Trees Explains Aboveground Biomass Variations across Nine Neotropical Forest Landscapes. **Biogeosciences Discussions**, 2018. p. 1–38.

MMA. Ministério do Meio Ambiente. **Mapas de Cobertura Vegetal dos Biomas Brasileiros**. 2002. Disponível em: <http://mapas.mma.gov.br/mapas/aplic/probio/datadownload.htm>. Acesso em: 8 nov. 2017.

MMA. Ministério do Meio Ambiente. Biomas. **Amazônia.** 2019. Disponível em: <http://www.mma.gov.br/biomas/amaz%C3%B4nia>. Acesso em: 18 mar. 2019.

MONTAÑO, R. A. N. R. **Aplicação de Técnicas de Aprendizado de Máquina na Mensuração Florestal. Universidade Federal do Paraná. Programa de Pós-Graduação em Informática**. [S.l.]: [s.n.], 2016. Disponível em: <https://acervodigital.ufpr.br/bitstream/handle/1884/45346/R - T - RAZER ANTHOM NIZER ROJAS MONTANO.pdf?sequence=1&isAllowed=y>. Acesso em: 12 set. 2018.

MOREIRA, M. A. **Fundamentos do sensoriamento remoto e metodologias de aplicação**. 3 ed. Viçosa: UFV, 2005. 320 p.

MORTON, D. C. Forest carbon fluxes: A satellite perspective. **Nature Climate Change**, 24 mar. 2016. v. 6, n. 4, p. 346–348. Disponível em: <http://www.nature.com/doifinder/10.1038/nclimate2978>. Acesso em: 7 nov. 2017.

MURPHY, K. P. **Machine Learning A probabilistic perspective**. Adaptive computation and machine learning series. Massachusetts Institute of Technology, 2012, 1067 p.

MUTANGA, O.; ADAM, E.; CHO, M. A. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. **International Journal of Applied Earth Observations and Geoinformation**, 2012. v. 18, p. 399–406.

NAESSET, E. Estimating timber volume of forest stands using airborne laser scanner data. **Remote Sensing of Environment**, 1997. v. 61, n. 2, p. 246–253.

NARASINGARAO, M. R. et al. A survey on prevention of overfitting in convolution neural networks using machine learning techniques. **International Journal of Engineering and Technology(UAE)**, 2018. v. 7, n. 2.32 Special Issue 32, p. 177–180. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-

85050310692&partnerID=40&md5=8afc5c0effa59cefbb859c7c03054cee>. Acesso em: 22 dez. 2018.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized Linear Models. **Journal of the Royal Statistical Society. Series A (General)**, 1972. v. 135, n. 3, p. 370. Disponível em: <https://www.jstor.org/stable/10.2307/2344614?origin=crossref>. Acesso em: 22 dez. 2018.

NILSSON, M. Estimation of tree heights and stand volume using an airborne lidar system. **Remote Sensing of Environment**, 1 abr. 1996. v. 56, n. 1, p. 1–7. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/0034425795002243>. Acesso em: 22 dez. 2018.

NOVO, E. M. L. de M. **Sensoriamento remoto: princípios e aplicações**. 4ª ed. São Paulo: Blucher, 2010. 387 p.

OLIVEIRA, N. DE; AMARAL, I. L. DO. Florística e fitossociologia de uma floresta de vertente na Amazônia Central, Amazonas, Brasil 1. Floristic and phytosociology of a slope forest in Central. **Acta Amazônica**, v. 34, p. 21–34, 2004. Disponível em: <http://www.scielo.br/pdf/aa/v34n1/v34n1a04.pdf>. Acesso em: 19 mar. 2019.

PAISAGENS SUSTENTÁVEIS BRASIL. **Forest Inventory: Fazenda Cauaxi**. 2016. Disponível em: <https://www.paisagenslidar.cnptia.embrapa.br/geonetwork/srv/por/catalog.search#/metadata/386defe3-9c2a-41e2-9392-dec94fbf52e3>. Acesso em: 12 dez. 2018.

PAL, M. Random forest classifier for remote sensing classification. **International Journal of Remote Sensing**, jan. 2005. v. 26, n. 1, p. 217–222. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/01431160412331269698>. Acesso em: 12 set. 2018.

PEDREGOSA, F. et al. **Scikit-learn: Machine Learning in Python. RBF SVM parameters — scikit-learn 0.20.2 documentation**. 2011. Disponível em: <https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html>. Acesso em: 23 dez. 2018.

PIEGORSCH, W. W. **Statistical Data Analytics. Fundations for Data Mining, Informatics, and Knowledge Discovery**. 1. ed. New York: John Wiley & Sons, 2015, 470p.

PINTO et al. 2009. **Diagnóstico Socioeconômico e Florestal do Município de Paragominas**. Relatório Técnico. Belém/PA: Instituto do Homem e Meio Ambiente da Amazônia - Imazon. 65 p. Disponível em: <https://imazon.org.br/PDFimazon/Portugues/outros/iagnostico-socioeconomico-e-florestal-do.pdf>. Acesso em: 21 mar. 2019.

POPESCU, S. C.; WYNNE, R. H.; NELSON, R. F. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. **Canadian Journal of Remote Sensing**, 2 out. 2003. v. 29, n. 5, p. 564–577. Disponível em: <http://www.tandfonline.com/doi/abs/10.5589/m03-027>. Acesso em: 24 dez. 2018.

PRASAD, A. M.; IVERSON, L. R.; LIAW, A. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. **Ecosystems**, 15 mar. 2006. v. 9, n. 2, p. 181–199. Disponível em: <http://link.springer.com/10.1007/s10021-005-

0054-1>. Acesso em: 12 jan. 2019.

R CORE TEAM. **R: The R Project for Statistical Computing**. Disponível em: <https://www.r-project.org/>. Acesso em: 11 set. 2018.

R DOCUMENTATION. R: Family Objects for Models. **R Foundation for Statistical Computing**, [S.l.], 2018. Disponível em: <https://stat.ethz.ch/R-manual/R-patched/library/stats/html/family.html>. Acesso em: 22 dez. 2018.

RAPPAPORT, D. I. et al. Quantifying long-term changes in carbon stocks and forest structure from Amazon forest degradation. **Environmental Research Letters**, 2018. v. 13, n. 6.

RODRIGUES, T. E. et al. **Caracterização e Classificação dos Solos do Município de Paragominas, Estado do Pará**. 1 ed. Belém: Embrapa Amazônia Oriental, 2003, 51 p.

RODRIGUEZ, L. C. E. et al. Inventário florestal com tecnologia laser aerotransportada de plantios de Eucalyptus spp no Brasil. **Ambiencia**, 2010. v. 6, p. 67–80.

RSTUDIOTEAM. **RStudio: Integrated Development Environment for R**. 2016. RStudio, Inc. Disponível em: <http://www.rstudio.com/>.

SANTOS, C. P. F. et al. Mapeamento dos Remanescentes e Ocupação Antrópica no Bioma Amazônia. XIII Simpósio de Sensoriamento Remoto. **Anais...**Florianópolis. 2007. p. 6941–6948.

SATO, L. Y. et al. Post-fire changes in forest biomass retrieved by airborne LiDAR in Amazonia. **Remote Sensing**, 2016. v. 8, n. 10, p. 1–15.

SCHAWLOW, A. L.; TOWNES, C. H. Infrared and optical masers. **Physical Review**. 1958. v. 112, n. 6, p. 1940–1949.

SCHNEIDER, P. R.; SCHNEIDER, P. S. P.; SOUZA, C. A. M. De. **Análise de regressão**. 2. ed. Santa Maria: FACOS – UFSM. 2009, 294 p.

SHAO, Z.; ZHANG, L. Estimating Forest Aboveground Biomass by Combining Optical and SAR Data: A Case Study in Genhe, Inner Mongolia, China. **Sensors (Basel, Switzerland)**, 7 jun. 2016. v. 16, n. 6. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/27338378>. Acesso em: 12 set. 2018.

SHERIDAN, R. et al. Modeling Forest Aboveground Biomass and Volume Using Airborne LiDAR Metrics and Forest Inventory and Analysis Data in the Pacific Northwest. **Remote Sensing**, 24 dez. 2014. v. 7, n. 1, p. 229–255. Disponível em: <http://www.mdpi.com/2072-4292/7/1/229>. Acesso em: 24 dez. 2018.

SHRESTHA, R.; WYNNE, R. H. Remote Sensing Estimating Biophysical Parameters of Individual Trees in an Urban Environment Using Small Footprint Discrete-Return Imaging Lidar. **Remote Sensing**. 2012. v. 4, p. 484–508. Disponível em: <www.mdpi.com/journal/remotesensingArticle>. Acesso em: 10 set. 2018.

SILVA, R. M. da**. Introdução ao Geoprocessamento**: conceitos, técnicas e aplicações. Novo Hamburgo, RS: Feevale, 2007. 176 p.

SILVA, L. de C. T. et al. Mapeamento do uso e cobertura da terra em áreas desflorestadas no município de Paragominas - PA nos anos de 1991 e 2008. 2011. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, XV, Curitiba. **Anais**... São José dos Campos: INPE, 2011, p. 6658-6665.

SILVA, C. et al. Impacts of Airborne Lidar Pulse Density on Estimating Biomass Stocks and Changes in a Selectively Logged Tropical Forest. 2017a. **Remote Sensing**, 23 out. 2017. v. 9, n. 10, p. 1068. Disponível em: <http://www.mdpi.com/2072-4292/9/10/1068>. Acesso em: 4 nov. 2017.

SILVA, C. A.; KLAUBERG, C.; HENTZ, Â. M. K. Predicting aboveground biomass in Pinus taeda L. plantation using airborne LiDAR data. 2017b. **Scientia Forestalis**, 2017. v. 45, n.115, p. 527-539.

SMREČEK, R.; DANIHELOVÁ, Z. Forest stand height determination from low point density airborne laser scanning data in Rožňava Forest enterprise zone (Slovakia). **IForest**, 2013, v. 6, p. 48-54.

SNIF. Sistema Nacional de Informações Florestais. **Os Biomas e suas florestas**. 2018. Disponível em: <http://snif.florestal.gov.br/pt-br/os-biomas-e-suas-florestas>. Acesso em: 18 mar. 2019.

SNIF. Sistema Nacional de Informações Florestais. **Estoque das Florestas**. 2019. Disponível em: <http://snif.florestal.gov.br/pt-br/estoques-das-florestas>. Acesso em: 18 mar. 2019.

SOARES, F. A. et al. Recursive diameter prediction and volume calculation of eucalyptus trees using Multilayer Perceptron Networks. **Computers and Electronics in Agriculture**, 2011. v. 78, p. 19–27. Disponível em: <http://www.deinfo.uepg.br/~ivo/rna_aplicadas_agricultura/Artigos RN/Recursive diameter prediction and volume calculation of eucalyptus trees using Multilayer Perceptron Networks.pdf>. Acesso em: 12 set. 2018.

SOUSA, C. L.; PONZONI, F. J. Avaliação de índices de vegetação e de bandas TM/Landsat para estimativa de volume de madeira em floresta implantada de Pinus spp. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 9. 1998, Santos. **Anais...** Santos: INPE, 1998. Disponível em: < http://marte.sid.inpe.br/col/sid.inpe.br/deise/1999/02.11.10.35/doc/2_03p.pdf >. Acesso em: 18 mar. 2019.

SRIVASTAVA, N. et al. **Dropout: A Simple Way to Prevent Neural Networks from Overfitting**. **Journal of Machine Learning Research**. [S.l.]: [s.n.], 2014. Disponível em: <http://jmlr.org/papers/volume15/srivastava14a.old/srivastava14a.pdf>. Acesso em: 2 jan. 2019.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition**. 4. ed. Elsevier. 2009, 961 p.

UN. **United Nations. Framework Convention Climate Change. Conference of the Parties (COP). Paris Agreement.** Disponível em: <http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf>. Acesso em: 1 dez. 2018.

VALE, J. R. B. Mapeamento do uso da terra e cobertura vegetal do município de

Paragominas, Sudoeste Paraense. Rio de Janeiro: XXVII Congresso Brasileiro de Cartografia e XXVI Exposicarta, nov. 2017. **Anais...**Disponível em: <http://www.cartografia.org.br/cbc/2017/trabalhos/4/fullpaper/CT04-41_1506692007.pdf>. Acesso em: 21 mar. 2019.

VELOSO, H. P.; FILHO, A. L. R. R.; LIMA, J. C. A. **Classificação da vegetação brasileira, adaptada a um sistema universal**. Rio de Janeiro: IBGE. 1991, 124 p.

VICINI, L. **Análise multivariada da teoria à prática**. Santa Maria: UFSM/CCNE, 2005, 215 p.

VIEIRA, A. F. D. C. **Análise da média e dispersão em experimentos fatoriais não replicados para otimização de processos industriais**. 2004. 204 f. Tese (Doutorado em Engenharia de Produção). Pontifícia Universidade Católica do Rio de Janeiro, 2004. Disponível em: <https://www.maxwell.vrac.puc rio.br/Busca_etds.php?strSecao=resultado&nrSeq=5789@1>.

VOSSELMAN, G.; MASS, H.-G. **Airborne and Terrestrial Laser Scanning**. Whittles Publishing. CRC Press, 1 ed. 2010, 320 p.

WAGNER, W. et al. Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. **ISPRS Journal of Photogrammetry and Remote Sensing**, 2006. v. 60, n. 2, p. 100–112.

WATZLAWICK, L. F.; KIRCHNER, F. F.; SANQUETTA, C. R. Estimativa de biomassa e carbono em floresta com araucaria utilizando imagens do satélite ikonos II. **Ciencia Florestal**, 2009. v. 19, n. 2, p. 169–181.

WATZLAWICK, L. F.; KOEHLER, H. S.; KIRCHNER, F. F. Estimativa de biomassa e carbono em plantios de Pinus taeda L. utilizando imagens do satélite IKONOS II. **Ciência e Natura.** Santa Maria: UFSM, 2006. v. 28, n.1, p. 45-60.

WITTEN, I. H.; FRANK, E. **Data Mining: Pratical Machine Learning Tools and Techniques**. 2. ed. San Francisco: Morgan Kaufmann publications. Elsevier, 2005. 525 p.

YADAV, S. Ground and Non-Ground Filtering for Airborne LIDAR Data. **International Journal of Advanced Remote Sensing and GIS**, 2016. v. 5, n. 1, p. 1500–1506.

YE, F. **Knowledge-driven board-level functional fault diagnosis**. 2014. 184 f. Tese (Doutorado em Filosofia). Duke University, Department of Electrical and Computer Engineering. 2014.

YE, F. et al. **Knowledge-driven board-level functional fault diagnosis**. Springer, ed. 1. 2017. 147 p.

YEOM, S. et al. Privacy Risk in Machine Learning: Analyzing the Connection to Overfitting. **IEEE,** 2018. p. 268–282.

ZANDONÁ, D. F.; LINGNAU, C.; NAKAJIMA, N. Y. Varredura a Laser aerotransportado para estimativa de variáveis dendrométricas. **Scientia Florestales**, 2008. v. 36, n. 80, p. 295–306.

ANEXO A – DADOS DO INVENTÁRIO FLORESTAL

Tabela 1 - Resumo do Inventário Florestal realizado em 88 unidades amostrais da fazenda Cauaxi, município de Paragominas/PA, no ano de 2014

(continua)

Família	Espécie	DAP Médio	Densidade da Madeira	Frequência
Lecythidaceae	*Couratari stellata*	51.8600	0.633	5
	Eschweilera coriacea	37.7957	0.852	141
	Eschweilera grandifolia	32.9500	0.876	2
	Eschweilera ovata	30.0625	0.900	80
	Eschweilera parviflora	26.3333	0.889	6
	Lecythis idatimon	22.3363	0.818	157
	Lecythis lurida	45.6000	0.830	25
	Lecythis pisonis	65.7400	0.852	5
***Lecythidaceae* Total**		**31.3371**	**0.845**	**421**
Sapotaceae	*Chrysophyllum sanguinolentum*	35.1500	0.671	8
	Manilkara amazonica	45.3385	0.884	26
	Manilkara huberi	53.8409	0.930	22
	Pouteria anibifolia	21.6286	0.660	7
	Pouteria gongrijpii	40.8927	0.799	55
	Pouteria guianensis	29.8702	0.930	47
	Pouteria hispida	35.9500	0.870	34
	Pouteria krukovii	36.9000	0.783	5
	Pouteria oppositifolia	49.5875	0.650	16
	Pouteria reticulata	43.1500	0.790	2
	Pouteria retinervis	23.8538	0.783	13
	Pouteria sp.	30.3125	0.782	24
	Pradosia praealta	44.3154	0.731	13
***Sapotaceae* Total**		**38.0482**	**0.827**	**272**
Fabaceae	*Abarema jupunba*	47.8167	0.585	6
	Abarema mataybifolia	18.7571	0.525	7
	Bowdichia nitida	39.4000	0.810	3
	Copaifera reticulata	65.2667	0.608	3
	Dialium guianense	29.7500	0.890	16
	Dimorphandra macrostachya	49.7000	0.600	1
	Dinizia excelsa	98.9667	0.939	3
	Diplotropis sp.	51.2000	0.750	1
	Dipteryx odorata	77.5000	0.920	1
	Enterolobium schomburgkii	62.5333	0.712	3
	Eperua bijuga	26.1867	0.729	30
	Hymenaea courbaril	27.1333	0.787	3
	Hymenaea intermedia	85.1000	0.820	1
	Hymenolobium petraeum	77.7500	0.713	2
	Inga alba	36.9125	0.586	8
	Inga grandis	15.6000	0.576	3
	Inga marginata	18.6889	0.573	9
	Inga sp.	25.6154	0.576	26
	Inga thibaudiana	21.6850	0.576	20
	Mucuna rostrata	13.6200	0.815	5

Tabela 1 - Resumo do Inventário Florestal realizado em 88 unidades amostrais da fazenda Cauaxi, município de Paragominas/PA, no ano de 2014

(continuação)

Família	Espécie	DAP Médio	Densidade da Madeira	Frequência
	Ormosia coccinea	17.0875	0.625	8
	Ormosia flava	40.2000	0.580	2
	Ormosia nobilis	37.0000	0.580	2
	Parkia gigantocarpa	18.3000	0.260	2
	Parkia pendula	82.4000	0.530	2
	Parkia sp.	36.0000	0.449	5
	Peltogyne leicointei	51.1800	0.796	5
	Piptadenia cobi	13.1000	0.780	1
	Platymiscium trinitatis	12.5000	0.830	1
	Poecilanthe effusa	11.7000	0.713	13
	Poeppigia sp.	23.3143	0.690	7
	Pseudopiptadenia suaveo-lens	50.5571	0.680	21
	Pterocarpus sp.	50.0000	0.512	1
	Stryphnodendron panicula-tum	47.6000	0.654	4
	Swartzia corrugata	33.0333	1.057	3
	Tachigali myrmecophila	38.0450	0.530	20
	Tachigali paniculata	61.4500	0.560	2
	Vataireopsis speciosa	14.4000	0.650	1
	Zollernia paraensis	47.5500	0.990	6
	Zygia racemosa	12.9500	0.750	4
***Fabaceae* Total**		**32.8069**	**0.668**	**261**
Burseraceae	*Protium hebetatum*	47.9714	0.576	14
	Protium hepytaphillum	23.8667	0.629	6
	Protium paniculatum	17.7316	0.490	19
	Tetragastris altissima	52.1500	0.708	2
	Tetragastris panamensis	40.1728	0.732	81
	Thyrsodium paraense	24.4000	0.640	1
***Burseraceae* Total**		**36.8650**	**0.671**	**123**
Violaceae	*Rinorea guianensis*	19.2505	0.780	109
	Rinorea passoura	12.2000	0.677	5
***Violaceae* Total**		**18.9412**	**0.776**	**114**
Chrysobalanaceae	*Couepia robusta*	43.1000	0.830	2
	Couepia sp.	36.4000	0.789	1
	Licania canescens	35.9536	0.880	28
	Licania heteromorpha	26.8214	0.816	14
	Licania octandra	29.3133	0.825	15
***Chrysobalanaceae* Total**		**32.4083**	**0.848**	**60**
Euphorbiaceae	*Hevea brasiliensis*	38.0000	0.492	11
	Hevea sp.	35.7733	0.516	15
	Mabea angularis	16.3500	0.616	2
	Sagotia racemosa	11.8875	0.580	16
***Euphorbiaceae* Total**		**26.7614**	**0.538**	**44**

Tabela 1 - Resumo do Inventário Florestal realizado em 88 unidades amostrais da fazenda Cauaxi, município de Paragominas/PA, no ano de 2014

(continuação)

Família	Espécie	DAP Médio	Densidade da Madeira	Frequência
Moraceae	*Brosimum parinarioides*	61.1000	0.580	2
	Brosimum rubescens	26.3000	0.825	4
	Clarisia racemosa	23.8000	0.585	1
	Ficus broadwayi	80.0000	0.415	1
	Helicostylis sp.	33.1083	0.653	24
	Maquira sclerophylla	24.6500	0.530	2
***Moraceae* Total**		**34.5618**	**0.652**	**34**
Humiriaceae	*Endopleura uchi*	37.7667	0.772	6
	Sacoglottis guianensis	30.1464	0.840	28
***Humiriaceae* Total**		**31.4912**	**0.828**	**34**
Malvaceae	*Apeiba echinata*	28.8500	0.276	6
	Bombax paraense	48.7000	0.390	2
	Luehea speciosa	21.3000	0.507	1
	Sterculia pruriens	29.3438	0.486	16
	Theobroma glaucum	10.8000	0.530	2
***Malvaceae* Total**		**28.9963**	**0.436**	**27**
Annonaceae	*Annona duckei*	15.9000	0.413	2
	Annona sp.	10.6500	0.413	4
	Duguetia stelechantha	13.4500	0.849	2
	Xylopia cayennensis	39.0333	0.570	6
	Xylopia sp.	24.2231	0.570	13
***Annonaceae* Total**		**24.0889**	**0.556**	**27**
Urticaceae	*Cecropia engleriana*	30.2500	0.490	2
	Cecropia palmata	18.9833	0.346	6
	Pourouma minor	39.6063	0.445	16
***Urticaceae* Total**		**33.6708**	**0.424**	**24**
NI	*Aiouea sp.*	36.0000	0.670	2
	Galipea sp.	12.1000	0.676	1
	Galipea trifoliata	13.5500	0.676	6
	NI	22.0500	0.640	14
***NI* Total**		**20.6130**	**0.653**	**23**
Lauraceae	*Endlicheria sp.*	47.0000	0.496	1
	Mezilaurus itauba	40.6000	0.720	1
	Nectandra cuspidata	29.5600	0.560	15
	Ocotea canaliculata	50.0000	0.479	1
	Ocotea glomerata	58.5000	0.508	1
	Ocotea sp.	35.7750	0.501	4
***Lauraceae* Total**		**34.0261**	**0.548**	**23**
Boraginaceae	*Cordia goeldiana*	40.6875	0.498	8
	Cordia scabrifolia	24.7091	0.474	11
***Boraginaceae* Total**		**31.4368**	**0.484**	**19**

Tabela 1 - Resumo do Inventário Florestal realizado em 88 unidades amostrais da fazenda Cauaxi, município de Paragominas/PA, no ano de 2014

(continuação)

Família	Espécie	DAP Médio	Densidade da Madeira	Frequência
Anacardiaceae	*Anacardium spruceanum*	33.0500	0.479	4
	Astronium gracile	42.7500	0.730	8
	Spondias sp.	16.9500	0.395	2
	Tapirira guianensis	32.5000	0.457	2
	Thyrsodium paraense	14.3500	0.613	2
***Anacardiaceae* Total**		**33.4333**	**0.594**	**18**
Apocynaceae	*Ambelania acida*	11.8333	0.525	3
	Aspidosperma nitidum	129.8000	0.763	1
	Aspidosperma spruceanum	43.8143	0.753	7
	Couma guianensis	41.0000	0.467	1
	Himatanthus sucuuba	46.1000	0.462	1
	Lacmellea aculeata	11.3000	0.513	1
***Apocynaceae* Total**		**40.7429**	**0.647**	**14**
Nyctaginaceae	*Neea oppositifolia*	33.4462	0.893	13
***Nyctaginaceae* Total**		**33.4462**	**0.893**	**13**
Myristicaceae	*Iryanthera paraensis*	13.4000	0.650	1
	Virola sebifera	28.1100	0.450	10
***Myristicaceae* Total**		**26.7727**	**0.468**	**11**
Meliaceae	*Cedrela odorata*	140.0000	0.430	1
	Guarea guidonia	18.6444	0.548	9
***Meliaceae* Total**		**30.7800**	**0.536**	**10**
Loganiaceae	*Strychnos subcordata*	12.8429	0.540	7
***Loganiaceae* Total**		**12.8429**	**0.540**	**7**
Combretaceae	*Terminalia amazonia*	50.5500	0.680	6
***Combretaceae* Total**		**50.5500**	**0.680**	**6**
Arecaceae	*Oenocarpus bacaba*	16.5667	0.650	6
***Arecaceae* Total**		**16.5667**	**0.650**	**6**
Caryocaraceae	*Caryocar glabrum*	60.3250	0.676	4
	Caryocar villosum	82.2500	0.758	2
***Caryocaraceae* Total**		**67.6333**	**0.703**	**6**
Quiinaceae	*Lacunaria jenmanii*	19.5500	0.804	4
	Quiina florida	41.5500	0.728	2
***Quiinaceae* Total**		**26.8833**	**0.779**	**6**
Salicaceae	*Laetia procera*	45.1800	0.633	5
***Salicaceae* Total**		**45.1800**	**0.633**	**5**

Tabela 1 - Resumo do Inventário Florestal realizado em 88 unidades amostrais da fazenda Cauaxi, município de Paragominas/PA, no ano de 2014

(conclusão)

Família	Espécie	DAP Médio	Densidade da Madeira	Frequência
Simaroubaceae	*Simaba cedron*	12.7000	0.474	3
	Simarouba amara	53.0000	0.378	2
***Simaroubaceae* Total**		**28.8200**	**0.436**	**5**
Bignoniaceae	*Jacaranda copaia*	30.9400	0.354	5
***Bignoniaceae* Total**		**30.9400**	**0.354**	**5**
Elaeocarpaceae	*Sloanea guianensis*	25.0000	0.821	5
***Elaeocarpaceae* Total**		**25.0000**	**0.821**	**5**
Clusiaceae	*Caraipa grandifolia*	30.3667	0.780	3
	Symphonia globulifera	39.9500	0.600	2
***Clusiaceae* Total**		**34.2000**	**0.708**	**5**
Ebenaceae	*Diospyros guianensis*	21.9250	0.730	4
***Ebenaceae* Total**		**21.9250**	**0.730**	**4**
Myrtaceae	*Myrcia velutina*	11.8500	0.801	4
***Myrtaceae* Total**		**11.8500**	**0.801**	**4**
Olacaceae	*Heisteria densifrons*	10.3000	0.650	1
	Minquartia guianensis	34.2500	0.770	2
***Olacaceae* Total**		**26.2667**	**0.730**	**3**
Ramnaceae	*Zizyphus itacaiunensis*	38.8333	0.838	3
***Ramnaceae* Total**		**38.8333**	**0.838**	**3**
Rubiaceae	*Amaioua guianensis*	10.2000	0.625	1
***Rubiaceae* Total**		**10.2000**	**0.625**	**1**
Rutaceae	*Euxylophora paraensis*	101.5000	0.656	1
***Rutaceae* Total**		**101.5000**	**0.656**	**1**
Vochysiaceae	*Qualea paraensis*	80.5000	0.689	1
***Vochysiaceae* Total**		**80.5000**	**0.689**	**1**
Dichapetalaceae	*Tapura singularis*	11.0000	0.660	1
***Dichapetalaceae* Total**		**11.0000**	**0.660**	**1**
Melastomataceae	*Mauriri chamissoana*	38.0000	0.640	1
***Melastomataceae* Total**		**38.0000**	**0.640**	**1**
Rhamnaceae	*Colubrina glandulosa*	15.8000	0.648	1
***Rhamnaceae* Total**		**15.8000**	**0.648**	**1**
Malpighiaceae	*Byrsonima chrysophylla*	13.2000	0.618	1
***Malpighiaceae* Total**		**13.2000**	**0.618**	**1**
TOTAL GERAL		**32.1115**	**0.738**	**1649**

Fonte: Adaptado de Paisagens Sustentáveis Brasil (2016).

ANEXO B – MÉTRICAS LIDAR UTILIZADAS NA MODELAGEM

Quadro 1 – Detalhamento das métricas LiDAR obtidas no software FUSION/LDV, versão 3.8
(continua)

Métricas FUSION/LDV		Descrição
1	*Total return count*	Contagem total de retornos (total de pontos)
2	*Total return count above 2.00*	Contagem total de retornos acima de **2 metros**[1]
3	*Return 1 count above 2.00*	Contagem de pontos em cada retorno discretizado acima de 2 metros
4	*Return 2 count above 2.00*	
5	*Elev maximum*	Elevação máxima
6	*Elev mean*	Elevação média
7	*Elev mode*	Moda da elevação
8	*Elev stddev*	Desvio padrão da elevação
9	*Elev variance*	Variância da elevação
10	*Elev CV*	Coeficiente de variação da elevação
11	*Elev IQ*	Distância interquartílica da elevação
12	*Elev skewness*	Assimetria na distribuição da elevação
13	*Elev kurtosis*	Curtose na distribuição de elevação
14	*Elev AAD*	Desvio médio absoluto da elevação
15	*Elev MAD median*	Mediana dos desvios absolutos da mediana geral (para elevação)
16	*Elev MAD mode*	Mediana dos desvios absolutos da moda geral (para elevação)
17	*Elev L1*	Momentos L1, L2, L3 e L4 na distribuição da elevação
18	*Elev L2*	
19	*Elev L3*	
20	*Elev L4*	
21	*Elev L CV*	Momento L de coeficiente de variação para elevação
22	*Elev L skewness*	Momento L de assimetria para elevação
23	*Elev L kurtosis*	Momento L de assimetria para elevação
24	*Elev P01*	Elevação nos diferentes percentis
25	*Elev P05*	
26	*Elev P10*	
27	*Elev P20*	
28	*Elev P25*	
29	*Elev P30*	
30	*Elev P40*	
31	*Elev P50*	
32	*Elev P60*	
33	*Elev P70*	
34	*Elev P75*	
35	*Elev P80*	
36	*Elev P90*	
37	*Elev P95*	
38	*Elev P99*	
39	*Canopy relief ratio*	Razão relevo dossel (elevação média - elevação mínima) / (elevção máxima– elevação mínima)

[1]Elevação mínima definida durante a configuração da função *Cloudmetrics* do FUSION/LDV.

Quadro 1 – Detalhamento das métricas LiDAR obtidas no software FUSION/LDV, versão 3.8 (continuação)

Métricas FUSION/LDV		Descrição
40	*Elev SQRT mean SQ*	Média generalizada para elevação quadrática
41	*Elev CURT mean CUBE*	Média generalizada para elevação cúbica
42	*Int maximum*	Intensidade máxima
43	*Int mean*	Intensidade média
44	*Int stddev*	Desvio padrão da intensidade
45	*Int variance*	Variância da intensidade
46	*Int CV*	Coeficiente de variação da intensidade
47	*Int IQ*	Distância interquartílica da intensidade
48	*Int skewness*	Assimetria na distribuição da intensidade
49	*Int kurtosis*	Curtose na distribuição da intensidade
50	*Int AAD*	Desvio médio absoluto da intensidade
51	*Int L1*	Momentos L1, L2, L3 e L4 na distribuição da intensidade
52	*Int L2*	
53	*Int L3*	
54	*Int L4*	
55	*Int L CV*	Momento L de coeficiente de variação para intensidade
56	*Int L skewness*	Momento L de assimetria para intensidade
57	*Int L kurtosis*	Momento L de assimetria para intensidade
58	*Int P10*	Intensidade nos diferentes percentis
59	*Int P20*	
60	*Int P25*	
61	*Int P30*	
62	*Int P40*	
63	*Int P50*	
64	*Int P60*	
65	*Int P70*	
66	*Int P75*	
67	*Int P80*	
68	*Int P90*	
69	*Int P95*	
70	*Int P99*	
71	*Percentage first returns above 2.00*	Percentual de primeiros retornos acima de 2 metros
72	*Percentage all returns above 2.00*	Percentual de todos os retornos acima de 2 metros
73	*(All returns above 2.00) / (Total first returns) * 100*	Percentual de retornos acima de 2 metros em relação ao primeiro retorno
74	*First returns above 2.00*	Primeiros retornos acima de 2 metros
75	*All returns above 2.00*	Todos os retornos acima de 2 metros
76	*Percentage first returns above mean*	Percentual de primeiros retornos acima da média
77	*Percentage first returns above mode*	Percentual de primeiros retornos acima da moda
78	*Percentage all returns above mean*	Percentual de todos os retornos acima da média
79	*Percentage all returns above mode*	Percentual de todos os retornos acima da moda
80	*(All returns above mean) / (Total first returns) * 100*	Percentual de retornos acima da média em relação ao primeiro retorno
81	*(All returns above mode) / (Total first returns) * 100*	Percentual de retornos acima da moda em relação ao primeiro retorno

Quadro 1 – Detalhamento das métricas LiDAR obtidas no software FUSION/LDV, versão 3.8
(conclusão)

Métricas FUSION/LDV		**Descrição**
82	*First returns above mean*	Primeiros retornos acima da média
83	*First returns above mode*	Primeiros retornos acima da moda
84	*All returns above mean*	Todos os retornos acima da média
85	*All returns above mode*	Todos os retornos acima da moda
86	*Total first returns*	Total de primeiros retornos
87	*Total all returns*	Total de retornos

Fonte: Adaptado de Mcgaughey (2016).