



FLAVIO CARVALHO FERREIRA

**PREDIÇÃO DE VOLUME EM FLORESTAS
PLANTADAS UTILIZANDO ESCANEAMENTO COM
LASER AEROTRANSPORTADO E APRENDIZAGEM
DE MÁQUINAS**

LAVRAS – MG

2022

FLAVIO CARVALHO FERREIRA

**PREDIÇÃO DE VOLUME EM FLORESTAS PLANTADAS UTILIZANDO
ESCANEAMENTO COM LASER AEROTRANSPORTADO E
APRENDIZAGEM DE MÁQUINAS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do programa de Pós-graduação em Engenharia Florestal, área de concentração Ciências Florestais, para obtenção do título de “Mestre”.

Prof. Dr. Luís Marcelo Tavares de Carvalho
Orientador

LAVRAS – MG

2022

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

Ferreira, Flavio Carvalho.

Predição de volume em florestas plantadas
utilizandoescaneamento com laser aerotransportado e aprendizagem
de máquinas / Flavio Carvalho Ferreira. - 2022.

4 p. : il.

Orientador(a): Luís Marcelo Tavares de Carvalho.

Dissertação (mestrado acadêmico) - Universidade Federal de
Lavras, 2022.

Bibliografia.

1. Escaneamento aéreo com laser. 2. Manejo Florestal. 3.
LiDAR. I. Carvalho, Luís Marcelo Tavares de. II. Título.

FLAVIO CARVALHO FERREIRA

**PREDIÇÃO DE VOLUME EM FLORESTAS PLANTADAS UTILIZANDO
ESCANEAMENTO COM LASER AEROTRANSPORTADO E
APRENDIZAGEM DE MÁQUINAS**

**VOLUME PREDICTION IN PLANTED FORESTS USING AIRBORNE
LASER SCANNING AND MACHINE LEARNING**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do programa de Pós-graduação em Engenharia Florestal, área de concentração Ciências Florestais, para obtenção do título de “Mestre”.

APROVADA em 30 de julho de 2021.

Prof. Dr. Luís Marcelo Tavares de Carvalho – UFLA

Dra. Maria Zélia Ferreira – Veracel

Prof. Dr. Natalino Calegário – UFLA

Prof. Dr. Luís Marcelo Tavares de Carvalho

Orientador

LAVRAS-MG

2022

AGRADECIMENTOS

À Universidade Federal de Lavras, especialmente ao Departamento de Ciências Florestais, pela oportunidade.

Ao apoio da Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG), com a concessão da bolsa de mestrado.

Ao professor Luís Marcelo Tavares de Carvalho, pela orientação, paciência e disposição em ajudar.

Aos colegas pós-graduandos do LEMAF/UFLA.

À todos os funcionários do DCF/UFLA e LEMAF/UFLA.

À minha família, pelos conselhos e apoio total em minhas decisões.

MUITO OBRIGADO!

RESUMO

O setor de florestas plantadas demanda cada vez mais informações em quantidade, qualidade e rapidez para basear as suas tomadas de decisão. Em função disso o presente trabalho compara a estimativa volumétrica obtida pelo inventário florestal utilizando o método convencional, baseado na média e desvio padrão das unidades amostradas, e um método que utiliza dados de escaneamento aéreo com laser (ALS – *airborne laser scanning*) dos talhões da floresta para gerar mapas com o volume de madeira. No inventário ALS, foram testados três diferentes métodos de modelagem: modelo linear ajustado pela minimização dos quadrados do erro (OLS – *ordinary least squares*), floresta aleatória (RF – *Random Forest*) e máquinas de vetor de suporte (SVM – *support vector machine*). Os modelos foram treinados e validados pelo método *holdout* por 200 vezes, avaliando no fim a média do coeficiente de determinação (R^2) e da raiz do erro quadrático médio (RMSE). O modelo que obteve melhor resultado foi o modelo linear com R^2 de 53,74% e RMSE de 20,19 m³/ha. Quando comparado os métodos de inventário convencional e ALS, ambos tiveram erros de inventário baixos. Pode-se concluir que a utilização dos inventários ALS mostra-se promissora para setor florestal.

Palavras-chave: Escaneamento aéreo com laser. Manejo Florestal. LiDAR. Inventário Florestal. Aprendizagem de máquinas.

ABSTRACT

The management of planted forest have been increasing the necessity for more and fast information, in quantity and quality, to base its decision-making. Due to it, the present work aiming compare the volumetric estimative obtained by the forest inventory using the conventional method, based on the mean and standard deviation of the sampled units, and a method that uses data from airborne laser scanning (ALS) of the stands of the forest to generate maps with the volume of wood. In the ALS inventory, three different modeling methods were tested: linear model adjusted by ordinary least squares (OLS), Random Forest (RF) and Support Vector Machines (SVM). The models were trained and validated by the holdout method for 200 times, evaluating in the end the mean of the coefficient of determination (R^2) and the root mean square error (RMSE). The model that obtained the best result was the linear model with R^2 of 53.74% and RMSE of 20,19 m^3/ha . When comparing conventional and ALS inventory methods, both had low inventory errors. It can be concluded that the use of ALS inventories is promising for the forest sector.

Keywords: Airborne laser scanning. Forest management. LiDAR. Forest inventory. Machine learning.

SUMÁRIO

1	INTRODUÇÃO	8
2	REVISÃO DE LITERATURA	11
2.1	COLETAS DE DADOS PARA ESTIMATIVA VOLUMÉTRICA EM FLORESTAS	11
2.2	ESCANEAMENTO AÉREO COM LASER EM FLORESTAS.....	12
2.3	MODELOS	13
2.3.1	Modelo linear	13
2.3.2	Random Forest	14
2.3.3	Support Vector Machine.....	15
3	MATERIAL E MÉTODOS.....	17
3.1	DESCRIÇÃO DA ÁREA DE ESTUDO	17
3.2	METODOLOGIA	18
3.2.1	Escaneamento aéreo com sensor laser.....	19
3.2.1.1	Coleta dos dados com sensor laser.....	19
3.2.1.2	Pré-processamento da Nuvem de Pontos	20
3.2.1.3	Extração de métricas da nuvem de pontos na área das parcelas.....	21
3.2.2	Modelagem dos dados ALS	25
3.2.2.1	Coleta e processamento de dados nas parcelas de calibração	25
3.2.2.2	Treinamento e seleção dos algoritmos	26
3.2.2.3	Extração de métricas da nuvem de pontos na área dos talhões	28
3.2.2.4	Aplicação dos modelos	29
3.2.3	Inventário convencional.....	29
3.2.3.1	Coleta e processamento de dados nas parcelas convencionais	29
3.2.4	Mensuração da madeira na fábrica.....	29
4	RESULTADOS E DISCUSSÃO	31
4.1	INVENTÁRIO COM AS PARCELAS DE CALIBRAÇÃO	31
4.2	TREINAMENTO E SELEÇÃO DOS MODELOS	31
4.3	MAPAS DE VOLUME DE MADEIRA DOS MODELOS	35
4.4	COMPARAÇÃO ENTRE OS INVENTÁRIOS.....	43
5	CONCLUSÕES	45
	REFERÊNCIAS	46

1 Introdução

Sabe-se a importância do uso dos produtos florestais para a sociedade, pois estes recursos foram e continuam sendo um condicionante ao desenvolvimento humano. Diariamente, convivemos com produtos provenientes das florestas, seja direta ou indiretamente. Os exemplos são diversos, como os papéis, os objetos que utilizam metais reduzidos através do carvão vegetal, os móveis que têm em sua composição madeira serrada, laminada ou reconstituída e uma variedade de outros produtos.

Devido à vasta utilização dos produtos florestais e considerando as florestas como um recurso escasso, surgiu a necessidade de se fazer a gestão e o controle eficiente deste recurso. Para tanto, são necessárias informações em quantidade e com qualidade adequadas, medidas com tecnologias de precisão e analisadas usando métodos eficientes.

Considerando as florestas plantadas, o Brasil ocupa uma posição de destaque, com aproximadamente 9 milhões de hectares plantados, contribuindo amplamente com o mercado mundial de produtos florestais, sendo o maior exportador de celulose da atualidade. Além disso, o setor tem grande importância econômica, já que é responsável por cerca de 1,2% do Produto Interno Bruto (PIB) nacional, contribuindo positivamente com a balança comercial brasileira e tendo apresentado uma variação positiva de 10,3% do PIB setorial entre os anos de 2018 e 2019 (IBÁ, 2019).

O sucesso brasileiro nos empreendimentos florestais ocorre por diferentes fatores, dentre eles, as condições edafoclimáticas propícias, emprego de tecnologias de ponta e uma boa gestão dos recursos, tornando o país competitivo mundialmente. Pela importância do setor ao país, é necessário comparar, testar e desenvolver técnicas de aprimoramento das tecnologias empregadas.

O modo de se obter informações das florestas como: altura, diâmetro, forma do fuste, volume, taxas de crescimento e qualidade, para possibilitar a melhor gestão do recurso vem evoluindo ao longo do tempo. Iniciou-se com a medição das árvores e criação de relações dendrométricas. Posteriormente, foram adicionadas técnicas estatísticas de amostragem e regressão, técnicas de sensoriamento remoto e técnicas de inteligência artificial. Todos esses avanços tecnológicos tiveram o objetivo de tornar as informações mais confiáveis e com maiores níveis de detalhes para as tomadas de decisão.

Recentemente, o uso do escaneamento com laser aerotransportado (ALS – Airborne Laser Scanning) tem se tornado popular na área florestal. Por meio dele pode-se gerar diferentes informações, desde os precisos modelos digitais do terreno, até mapas que estimam a

quantidade de madeira nas florestas. Este uso vem se popularizando não só no mundo, mas também no Brasil, onde existem várias oportunidades para a aplicação desta tecnologia.

Os métodos para estimar o volume de madeira nas florestas, partindo-se de informações do escaneamento a laser, podem ser classificados em dois tipos. O primeiro é baseado em árvores individuais, onde tenta-se isolar, na nuvem de pontos geradas pelo sensor laser, cada uma das árvores do povoamento. No entanto, este é um método que não se mostrou eficiente em florestas de alta densidade (YU et al. 2010), como é o caso brasileiro. Já o segundo método é baseado em área, analisando métricas estatísticas locais extraídas das nuvens de pontos geradas pelo ALS, este método vem se mostrando mais eficiente em florestas plantadas (VAUHKONEN, MEHTÄTALO, PACKALÉN, 2011).

No início do uso de dados ALS para estimativas de volume em florestas, utilizando o método baseado em área, eram aplicados apenas modelos lineares, principalmente o *Ordinary Least Squares* (OLS), para correlacionar o volume da floresta com as métricas extraídas dos dados provenientes do ALS (NÆSSET, E, 1997, 2002). Embora o modelo OLS seja o mais utilizado por sua clareza e simplicidade, este tipo de modelo é frágil à utilização de variáveis correlatas e, por ser uma técnica paramétrica, exige normalidade, homocedasticidade, independência e linearidade dos dados (GARCÍA-GUTIÉRREZ et al., 2015).

Existem, também os modelos não paramétricos que vem sendo utilizados nas áreas de sensoriamento remoto, mostrando diferentes possibilidades de uso para gerar estimativas cada vez mais precisas. Os principais modelos não paramétricos que vem sendo utilizados são: o *Random Forest* (RF) (BREIMAN, 2001) e o *Support Vector Machine* (SVM). Estes modelos são menos sensíveis às escolhas de variáveis independentes para gerar estimativas precisas, por isso o seu uso pode tornar-se viável para estimativas de volume em florestas plantadas de eucalipto no Brasil.

Nas publicações relacionadas a levantamentos com sensores laser aerotransportados no Brasil (ZANDONÁ, 2006; MACEDO, 2009; ZONETE, RODRIGUEZ, PACKALÉN, 2010; VAUHKONEN, MEHTÄTALO, PACKALÉN, 2011; PACKALÉN, MEHTÄTALO, MALTAMO, 2011; SILVA, KLAUBERG, PÁDUA, 2014; SILVA et al., 2016; SILVA et al., 2017a; SILVA et al., 2017b), que comparam métodos para estimar e, ou mapear algumas características dendrométricas de florestas plantadas de eucalipto ou pinus, não apresentam consenso de qual seria a melhor metodologia a ser seguida para estimar o volume em florestas clonais de eucalipto, evidenciando a necessidade de estudos que complementem o tema.

Sendo assim, neste trabalho foram exploradas, aplicadas e aprimoradas técnicas de análise de dados de escaneamento aéreo com sensor laser (ALS), em florestas clonais de eucalipto do sul da Bahia.

De forma geral, o objetivo do trabalho foi: apresentar uma revisão conceitual sobre o tema, ajustar e comparar diferentes modelos para estimar o volume de madeira em florestas plantadas de eucalipto com dados derivados de ALS e comparar o método de inventário convencional com o método de inventário utilizando os dados do sensor laser.

Mais especificamente buscou-se:

- Testar, entre três métodos de modelagem, OLS, RF e SVM, qual o melhor para estimar o volume de madeira em florestas plantadas de eucalipto com dados derivados de ALS.
- Comparar o método de inventário convencional com o método de inventário utilizando os dados do sensor laser.

2 REVISÃO DE LITERATURA

2.1 Coletas de dados para estimativa volumétrica em florestas

Tradicionalmente as coletas de informações em florestas são realizadas através de inventários florestais, que consistem na amostragem de parte da floresta, com a finalidade de fazer inferências para toda a área. Os inventários florestais podem ser qualitativos e, ou, quantitativos. Estes inventários convencionais utilizam-se da alocação de parcelas e da medição de variáveis independentes nestas, como o diâmetro à altura do peito, a altura, a altura dominante, a área da parcela, e outras. Estas parcelas representam amostras de grandes áreas, gerando médias genéricas do material existente na floresta (DRUSZCZ1 et al., 2012).

Para a coleta dos dados são empregadas equipes de campo, onde podem ser utilizados diferentes equipamentos para as medições. Para florestas clonais de *Eucalyptus spp.* no Brasil é comum o uso de fitas métricas para medir a circunferência à altura do peito das árvores, diferentes modelos de clinômetros para medir altura total, além de GPS, réguas e outras ferramentas. Nos inventários tradicionais, geralmente são medidos os diâmetros de todas as árvores e as alturas de parte das árvores da parcela. Com as informações de diâmetro e altura, ajustam-se equações hipsométricas que são capazes de estimar com boa acurácia a altura para as árvores que tiveram apenas o diâmetro medido. Para a determinação do volume existente nas parcelas, também faz-se a utilização de outra prática de coleta de informações, elas são chamadas cubagens rigorosas que, nas florestas do gênero *Eucalyptus* brasileiras são realizadas de forma destrutiva, sendo uma atividade de grande risco as equipes de campo pelo momento de abate das árvore (SOARES, PAULA NETO, SOUZA, 2006). Nesta etapa são ajustadas equações volumétricas, tendo o diâmetro e a altura das árvores como variáveis independentes. A variável volume é proveniente das cubagens rigorosas, onde árvores em diferentes classes diamétricas são abatidas e medidas em intervalos de altura pré-determinados, fazendo a medição de diâmetro e tendo o comprimento de diversos segmentos da árvore. O volume é calculado para cada segmento e a soma destes constitui o volume total da árvore.

Uma vez obtida a equação volumétrica, pode-se determinar o volume de cada árvore individualmente na parcela e o volume total da parcela. Finalmente, com base no volume da parcela e utilizando-se de técnicas da teoria da amostragem, pode-se extrapolar o volume obtido na parcela para o hectare, talhão e povoamento.

2.2 Escaneamento aéreo com laser em florestas

Dentre os diversos sensores remotos utilizados no meio florestal, seja no Brasil ou no mundo, o escaneamento aéreo com sensor laser (ALS) tem se destacado como uma opção para se estimar variáveis dendrométricas de interesse (volume, área basal, altura média, etc.) em povoamentos florestais naturais ou cultivados.

Com o propósito de reduzir a intensidade amostral, aumentar a acurácia e velocidade de coleta das informações são utilizados levantamentos de dados por meio de sensores remotos, não só no Brasil, como também em outros países. Kangas et al. (2018) apresentam como o setor florestal está se desenvolvendo nos países nórdicos, dando destaque às novas tecnologias utilizadas e que ainda têm o potencial a serem exploradas, principalmente o escaneamento aéreo com sensor laser (ALS – *Airborne Laser Scanning*), também conhecido como escaneamento com sensor LiDAR (*Light Detection And Ranging*). Esta tecnologia consiste em um sensor remoto ativo que mede distâncias baseado no tempo de retorno de um feixe de luz emitido pelo próprio sensor. Além disto, o sensor é integrado a um GPS, que tem como função aumentar a acurácia da nuvem de pontos gerada pelo escaneamento.

Os primeiros registros do uso do escaneamento com sensores laser datam de meados dos anos 60 com Rempel e Parker (1964), no qual foi utilizado o sensor laser para obter informações topográficas e de altura das árvores. Já os estudos de (MACLEAN & MARTIN, 1984; MACLEAN & KRABILL 1986), buscaram desenvolver metodologias para estimar o volume e a biomassa de florestas com base nos modelos de superfície gerados pelo escaneamento com sensores laser.

Este tipo de tecnologia em áreas florestais não é tão recente, pois há estudos que retratam possíveis aplicações há 25 anos, detalhando aplicações e usos práticos dos dados de ALS (NÆSSET, 1997; NÆSSET, 2002; NÆSSET et al., 2004; HYYPPÄ et al., 2008; PACKALÉN, MEHTÄTALO, MALTAMO, 2011).

Dentre as aplicações, destacam-se os ajustes de modelos para estimativas de volume com sensores de laser, os quais se dividem basicamente em dois tipos: ajustes baseados em árvores individuais, nos quais, com o processamento dos dados do laser, tenta-se isolar cada árvore na floresta e fazer uma estimativa para as mesmas; e ajustes baseados em área, onde se tem o volume de parcelas com áreas conhecidas e utilizam-se as informações do sensor laser para estimar o volume para a área (VAUHKONEN et al., 2014).

Yu et al. (2010) compararam os dois métodos de estimar o volume em florestas exploradas na Finlândia e concluíram que o método baseado em área é mais operacional em

grandes áreas florestais, pois demanda um menor poder computacional, faz-se uma generalização das informações, evitando problemas como sobreposição da copa das árvores e exige uma menor densidade de pontos coletados pelo sensor.

Além disso, Vauhkonen, MehtÄtalo e Packalén (2011) utilizaram o método de árvores individuais em florestas clonais de eucalipto no sul da Bahia e concluíram que este não é o melhor método a se utilizar neste tipo de floresta, pois por terem um desenvolvimento uniforme, as árvores nas entrelinhas têm suas copas sobrepostas, não possibilitando que o sensor laser capte as árvores individualmente.

Em trabalho realizado por Silva et al. (2016) utilizando o método de estimativa baseado em área em florestas clonais de eucalipto no Vale do Paraíba, no estado de São Paulo, os autores utilizaram o Software FUSION (MCGAUGHEY, 2018) para processar os dados do sensor e gerar diferentes métricas, com as quais foram testadas equações, pelo método de regressão linear simples, para estimativa do volume, resultando em variáveis que melhor estimaram o volume para a área em estudo. Por fim, os autores criaram mapas com a distribuição espacial do volume nos diferentes talhões da floresta.

No trabalho de Silva et al. (2017a), buscando entender a influência nas estimativas de carbono acima do solo em florestas de *Eucalyptus spp.*, localizadas no Brasil, diferentes densidades de pontos geradas pelo sensor LiDAR foram comparadas. Buscou-se também observar como se comportavam os diferentes tamanhos de células nos mapas criados para as estimativas de carbono. Neste trabalho os autores concluíram que se pode estimar a quantidade de carbono acima do solo com precisão utilizando uma densidade de 5 pontos por metro quadrado e que células de 5x5metros geram bons mapas da variável. Estas conclusões tornam-se interessantes de serem testadas nas estimativas de outras características dendrométricas em florestas de *Eucalyptus spp.*

2.3 Modelos

2.3.1 Modelo linear

Técnicas de análises de regressão lineares ou não-lineares são comumente empregadas nas áreas de ciências biológicas, agricultura, ciências médicas, epidemiologia, economia, dentre outras (PAN; FANG, 2002). No geral têm o objetivo de entender os comportamentos de uma variável dependente em relação a um conjunto de variáveis independentes. No meio florestal as análises de regressão têm um amplo histórico de usos com a modelagem de crescimento,

mortalidade, ajuste de equações que estimam volume, área basal, altura total e comercial (CLUTTER, 1963; SULLIVAN; CLUTTER, 1972; EISFELD et al., 2005; SCOLFORO et al., 2013; GROENENDIJK; BONGERS; ZUIDEMA, 2017).

Os modelos de regressão linear são expressos de maneira que Y , a variável dependente, tenha relações com as variáveis independentes x_i , as quais estão associadas a parâmetros β_i , implícitos de cada modelo. Ademais, esta relação estará sempre associada a um erro aleatório ε . A regressão linear consiste basicamente na tendência central de uma base de dados onde a variável Y está relacionada a variáveis x_i (LEPAGE, 2011). Em forma de equação, as regressões lineares são em geral representadas: $Y = f(x, \beta, \varepsilon)$. Assume-se com o uso da regressão linear que os resíduos de cada uma das observações (ε_i) sejam independentes, para isso devem se comportar de forma que a média de ε seja igual a zero e que apresentem homoscedasticidade, ou seja, uma variância positiva e constante no intervalo dos dados (PAN, FANG, 2002).

O método de ajuste mais comum para regressões lineares é o método dos quadrados mínimos (*Ordinary Least Squares*), o qual busca, para um conjunto de dados, minimizar a soma dos quadrados das diferenças entre o valor estimado e os valores observados (FERREIRA, 2005).

2.3.2 Random Forest

O algoritmo de aprendizagem de máquinas Floresta Aleatória, ou *Random Forest* (RF), é um método de aprendizagem supervisionada que se baseia em uma combinação (*ensemble*) de *n* árvores de decisão criadas a partir de um conjunto de variáveis aleatórias (BREIMAN, 2001). A vantagem da Floresta Aleatória é que sua aplicação pode ser feita tanto para problemas de classificação quanto para aqueles de regressão (BREIMAN, 2001). Outra vantagem com relação à utilização de algoritmos com apenas uma árvores de decisão (*CART: Classification and Regression Trees*), é que a Floresta Aleatória evita o sobreajuste (*overfitting*) sem aumentar consideravelmente o viés do modelo (MURPHY, 2012), i.e., a Floresta Aleatórias garante um melhor *trade-off* entre a variância e o viés.

O algoritmo faz amostragens com reposição no conjunto de treinamento da base de dados (*Bootstrap*), gerando diferentes subconjuntos aleatoriamente, os quais são utilizados para a criação das *n* árvores, número de árvores na floresta. A técnica de *bootstrap* possibilita a criação de diferentes séries históricas (porém de mesmo tamanho), garantindo que cada árvore da floresta seja desenvolvida a partir de uma base de dados única. O segundo aspecto que evita o

sobreajuste é a seleção das variáveis explicativas que irão compor cada árvore de decisão e, assim, definir o seu crescimento (BREIMAN, 2001). A cada nó da árvore de decisão é selecionado aleatoriamente *mtry* atributos para definir o seu crescimento. Além disso, o valor de *mtry* deve ser menor que o número de atributos (variáveis explicativas) para evitar a correlação entre as árvores da floresta (TYRALIS; PAPACHARALAMPOUS; LANGOUSIS, 2019). Dessa forma, dois parâmetros devem ser devidamente ajustados para limitar os erros devido à variância e o viés na Floresta Aleatória: (i) o número de árvores na floresta (*ntree*) e (ii) o número de variáveis explicativas em cada árvore (*mtry*).

Nos casos em que a Floresta Aleatória é utilizado para classificação, cada árvore de decisão contribui com um voto para a classe a ser classificada. Por outro lado, para os casos de regressão, cada árvore de decisão contribui com um valor para a contabilização final do resultado. Neste caso, o resultado final é o valor médio da resposta de cada árvore. Devido a todos os aspectos levantados, a utilização do algoritmo Floresta Aleatória para a modelagem de volume em florestas plantadas vem se destacando com o passar do tempo (REIS et al. 2018, SOUZA et al. 2019, SILVA et al. 2020). O algoritmo se comporta bem para bases de dados com poucas observações e, também, não é sensível à escolha dos hiperparâmetros, o que facilita a sua utilização.

2.3.3 Support Vector Machine

O *Support Vector Machine* (SVM) é um algoritmo de aprendizagem supervisionado que pode ser aplicado tanto para problemas de classificação quanto para regressões. Sua teoria foi originalmente proposta por Vapnik (1995) visando aprimorar a separação da base de dados em classes (BOSER et al., 1992; BURGESS, 1998), sendo, a posteriori, estendido para problemas de regressão no intuito de modelar processos de certa complexidade (e.g. relações não lineares) a partir de variáveis explicativas (Vapnik, 1995). A separação de classes em um banco de dados acontece a partir da construção de hiperplanos no ambiente do SVM (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). A distância entre este hiperplano e o primeiro ponto de cada uma das classes é conhecida como margem e os dados que se encontram nestas margens constituem os vetores de suporte.

Além disso, para os modelos SVM os conceitos de kernels deve ser usado para abordar a não linearidade das variáveis, uma vez que eles simplificam a complexidade do modelo e, portanto, minimiza o erro das previsões (RAGHAVENDRA & DEKA, 2014). Os kernels mais

comuns utilizado no SVM são funções lineares, polinomiais, função de base radial (*radial basis function* - RBF) e sigmóide.

A utilização do algoritmo para a estimativa de volume de madeira e de outras variáveis dendrométricas vem sendo testada em estudos recentes, García-Gutiérrez et al. (2015); Wu et al. (2016); Reis et al. (2018), em cada um destes estudos os autores utilizam diferentes técnicas de sensoriamento remoto para a estimativa das características dendrométricas e em todos fizeram a comparação entre métodos de modelagem, sendo o SVM um dos métodos testados em todos os artigos.

Os autores García-Gutiérrez et al. (2015) testaram diferentes comportamentos de métodos de modelagem de volume de madeira em florestas utilizando dados de escaneamento aéreo com sensor laser, este estudo foi realizado em 2 áreas florestais na Espanha, em condições diferentes das florestas clonais de eucaliptos brasileiras. Dentre os métodos testado no estudo está o clássico modelo linear e alguns modelos de aprendizagem de máquinas: redes neurais, máquinas de vetores de suporte (SVM), vizinho mais próximo, conjuntos como florestas aleatórias (RF). Neste estudo o melhor desempenho foi associado ao método de aprendizagem de máquinas SVM, utilizando o kernel gaussiano.

Já Reis et al. (2018) relatam a utilização do método de aprendizagem de máquinas RF para modelar o volume de madeira em florestas plantadas de eucalipto no norte de Minas Gerais. Porém, diferente de García-Gutiérrez et al. (2015) e deste estudo que utilizaram dados de escaneamento aéreo com sensor laser (ALS), Reis et al. (2018) modelam variáveis derivadas de imagens ópticas coletadas pelo satélite Landsat 8, imagens de radar coletadas pelo satélite Sentinel-1B e modelos digitais de elevação derivados do ALOS Palsar.

3 MATERIAL E MÉTODOS

3.1 Descrição da área de estudo

A área de estudo abrange florestas clonais de eucalipto localizadas no sul do Estado da Bahia, Brasil (Figura 1). As florestas são povoadas por clones do híbrido de *Eucalyptus grandis* com *Eucalyptus urophylla* de primeira rotação, com idades variando de 4 a 5,2 anos e com espaçamento de 4x3 metros. O clima da região é quente, úmido e sem estação seca definida. De acordo com metodologia proposta por Köppen, o clima é classificado como tropical úmido (Af), característico do litoral sul da Bahia. A precipitação média anual é de 1200 mm e apresenta distribuição uniforme ao longo do ano. A temperatura média é de 24°C, com pequena amplitude.

Na área do estudo o relevo é plano, característico do Planalto Costeiro, associado ao relevo de topos tabulares do Grupo Barreiras. São predominantes os solos da classe argissolos amarelos, apresentando horizonte B textural com várias derivações, desde arenosos a muito argilosos, com ocorrência frequente de camada adensada em subsuperfície, com alto grau de coesão. Estas foram informações repassadas pela empresa e foram obtidas em estudos não publicados em revistas acadêmicas.

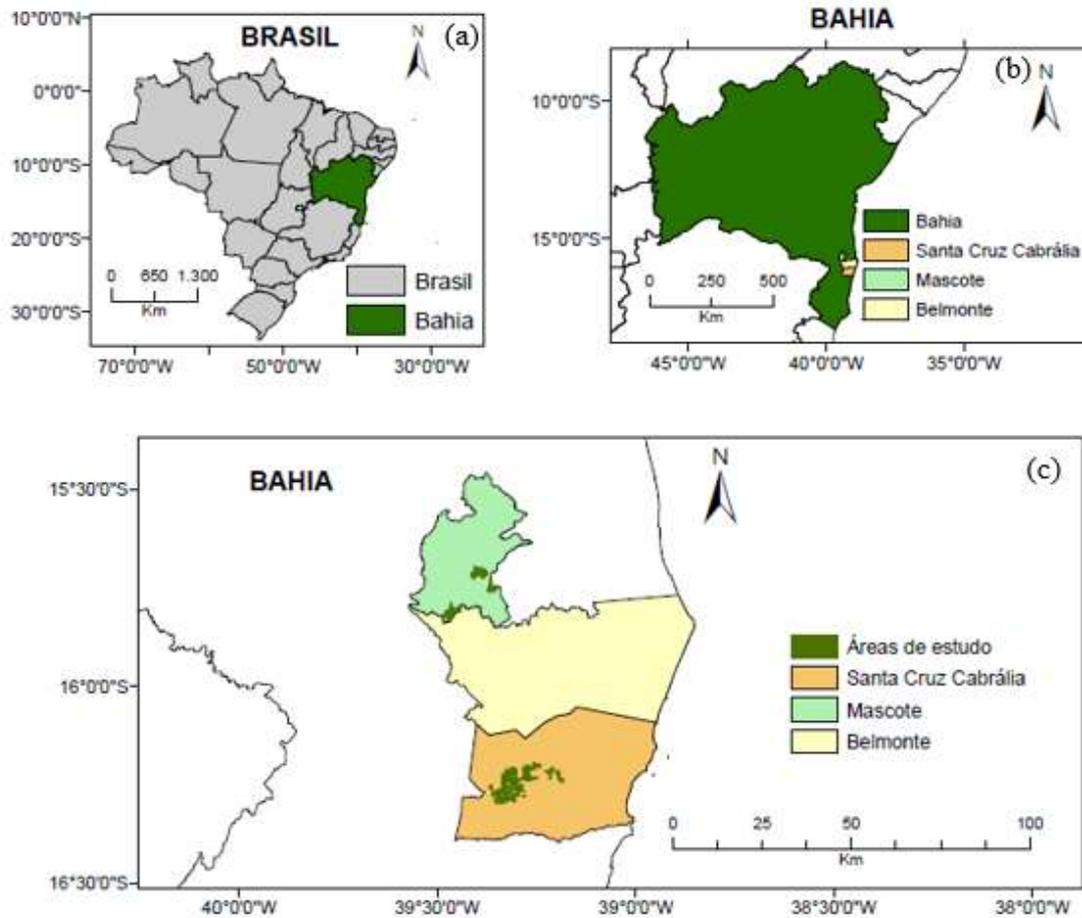
Para melhor detalhar, as florestas estudadas foram divididas em três diferentes áreas, separadas no espaço, sendo a Área 1 e Área 2 com florestas na idade de 5,2 anos e Área 3 com 4 anos de idade. A idade média, a área total e o número de unidades amostrais por área estão especificados na Tabela 1.

Tabela 1 – Características das Áreas com floresta de eucalipto estudadas.

Floresta	Idade (anos)	Área total da floresta (ha)	Número de unidades amostrais
Área 1	5	217,2	10
Área 2	5	791,0	19
Área 3	4	277,4	10

Fonte: do autor (2022)

Figura 1 – Mapas com a localização das áreas de estudo, (a): mapa do Brasil, (b): mapa do estado da Bahia e (c): mapa destacando os municípios onde as florestas do estudo estão localizadas.

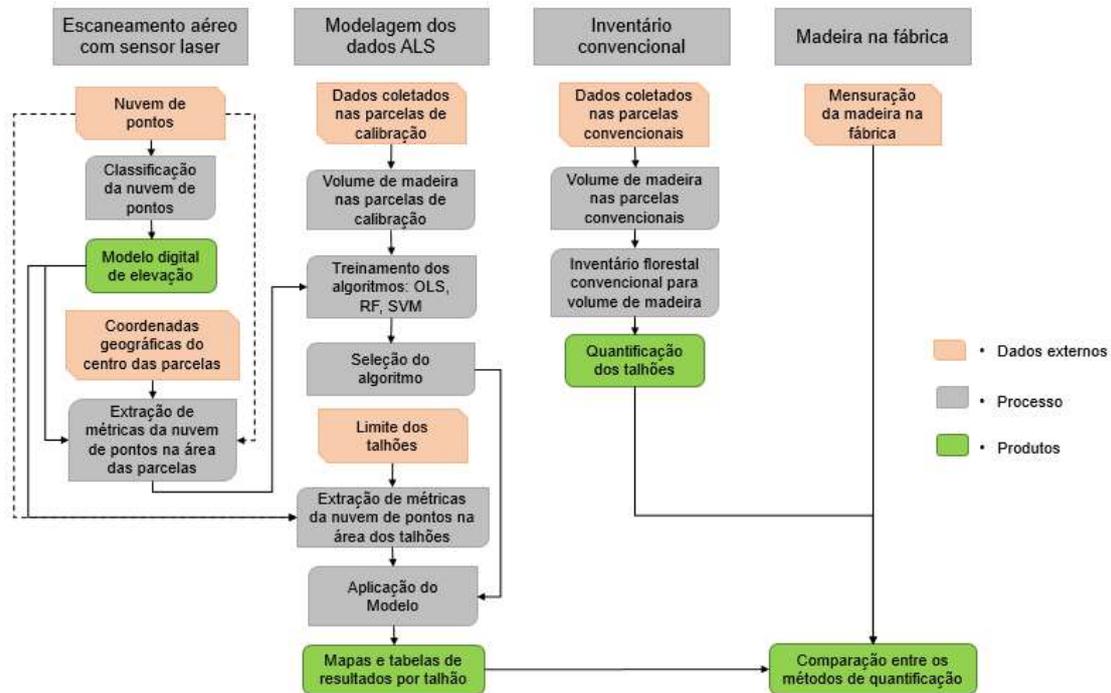


Fonte: do autor (2022)

3.2 Metodologia

O procedimento metodológico foi dividido em quatro etapas. A primeira está relacionada com o processamento dos dados do escaneamento aéreo com laser. A segunda, com o inventário para calibração das informações do ALS. A terceira, com o inventário convencional e, a última, com a mensuração da madeira na fábrica e comparações entre os diferentes métodos de inventário florestal. O fluxograma apresentado na Figura 2 é uma visão geral das etapas do trabalho.

Figura 2 – Fluxo de trabalho para o processamento dos inventários florestais pelos métodos ALS; e comparação do melhor método ALS em relação ao inventário convencional e a quantidade de madeira colhida em uma das áreas estudadas.



Fonte: do autor (2022)

3.2.1 Escaneamento aéreo com sensor laser

3.2.1.1 Coleta dos dados com sensor laser

Os dados de ALS foram coletados utilizando o Sensor LiDAR Harrier 68i da Trimble; um Sistema Inercial Applanix (IMU+GPS), modelo POS AV 150, e uma Aeronave CESSNA modelo 206. O sobrevoo foi realizado em janeiro de 2019. Dentre as características apresentadas, destacam-se as acurácias absolutas, sendo a vertical menor que 0,15 metros e a horizontal menor que 0,25 metros. Mais detalhes sobre as especificações técnicas do voo estão na Tabela 2.

Tabela 2 – Especificações técnicas do sistema LiDAR Harrier 68i.

Especificações do Voo	Valores
Densidade de pontos	5,5 pts/m ²
Frequência de repetição de pulso	300 Khz
Fov (Campo de visada)	45°
Velocidade	55 m/s
Altura de voo	597 m
Footprint (Pegada)	0,43 m
Largura de faixa	494,93 m
Overlap (Sobreposição)	30%

Fonte: do autor (2022)

Os itens citados na Tabela 2 são detalhes do sensor laser utilizado, sendo: a densidade de pontos é o número médio de primeiros retornos do sensor mensurados a cada metro quadrado sobrevoado; a frequência de repetição de pulso retrata o número de pulsos que o sensor envia por segundo; o campo de visada é o ângulo, no sentido perpendicular a direção de voo, em relação à vertical com que o sensor emite e capta os pulsos de retorno; a velocidade está relacionada ao deslocamento do avião; a altura de voo é a altura em que o voo foi realizado, mensurado pelo altímetro do avião; o *Footprint* é o diâmetro cujo o feixe de luz atinge a superfície; a largura de faixa é a distância na superfície, ortogonal à direção de voo, que o sensor cobre durante um voo; *Overlap* é a sobreposição que existe entre cada largura de faixa.

O resultado da coleta de dados pelo sensor acoplado ao avião foram arquivos digitais com nuvens de pontos dos diferentes talhões sobrevoados. Estes arquivos com as nuvens de pontos foram empregados para o processamento do inventário ALS.

3.2.1.2 Pré-processamento da Nuvem de Pontos

O processamento inicial dos dados coletados pelo sensor LiDAR foi feito junto aos Softwares FUSION (MCGAUGHEY, 2018) e LAsTools (ISENBURG, 2014). A nuvem de pontos é um arquivo digital com as leituras realizadas pelo sensor LiDAR, que emite uma luz infravermelha e capta o tempo de retorno desta luz, calculando assim a distância entre o sensor

e o ponto em que a luz atinge uma superfície. O arquivo digital da nuvem de pontos contém várias informações, sendo as principais: latitude, longitude e altitude de cada um dos pontos. O software LAStools foi utilizado para descompactar o formato no qual os arquivos foram recebidos após o levantamento.

A classificação da nuvem de pontos teve o objetivo de filtrar dentre todos os pontos que a constituam, aqueles que provavelmente atingiram a superfície do solo. Para esta tarefa, a função *Grounfilter* do software FUSION foi utilizada. Foram selecionados os parâmetros *default* do software para a função de filtragem, pois McGaughey (2018) relata o bom desempenho destes parâmetros para nuvens de pontos com densidade maior que 4 pontos por metro quadrado. Os pontos selecionados após a filtragem são conhecidos como *bare-earth points*.

A criação do modelo digital de elevação foi feita no software FUSION (MCGAUGHEY, 2018), criando células de 5x5 metros dos *bare-earth points*, classificados na etapa anterior. O modelo digital de elevação foi utilizado para normalização da superfície onde a floresta se encontrava, assim, após a normalização considerou-se a floresta em um relevo plano, podendo extrair valores padronizados para as diferentes florestas, ou seja, extinguiu-se a influência do relevo sobre a nuvem de pontos de cada uma das áreas estudadas.

3.2.1.3 Extração de métricas da nuvem de pontos na área das parcelas

A extração das métricas da nuvem de pontos consiste em gerar, a partir dos dados do escaneamento aéreo com o sensor LiDAR, informações que sirvam de variáveis independentes para a modelagem das características dendrométricas de interesse na floresta, no caso deste trabalho, o volume de madeira sem casca.

Nesta etapa do trabalho, o modelo digital de elevação já estava concluído e já haviam sido mensuradas as 39 unidades amostrais para a modelagem dos dados. Os detalhes sobre as unidades amostrais estão descritos no item 3.2.2.1. Nesse momento é importante mencionar que foram utilizadas 39 parcelas circulares com área de 400 m² (raio de 11,28 metros), cujos centros foram georreferenciados com o uso do GPS geodésico (Trimble GPS Pro-XRT) com precisão de 20 centímetros e correção em tempo real.

A extração das métricas da nuvem de pontos foi dividida em duas etapas, ambas realizadas no software FUSION (MCGAUGHEY, 2018). A primeira consistiu em determinar

os pontos que estavam dentro dos limites de cada uma das unidades amostrais. A segunda etapa consistiu em gerar as métricas da nuvem de pontos dentro das unidades amostrais.

A primeira etapa foi realizada com as coordenadas dos centros das parcelas e com seus raios. A segunda etapa foi realizada com a função *CloudMetrics*, com os parâmetros *above*: 1.3 e *outlier*: 0.8 a 50. Nesta, foram utilizados apenas os primeiros retornos do sensor (*firstreturn*) e todas as unidades amostrais tiveram o relevo normalizado pelo modelo digital de elevação criado e apresentado na seção 3.2.1.2.

Aplicada a função, foram extraídas e selecionadas 33 métricas para cada uma das parcelas. Essas métricas (Tabela 3) foram utilizadas como variáveis independentes para a predição do volume de madeira sem casca nas florestas.

Tabela 3 – Métricas derivadas do escaneamento aéreo com laser (MCGAUGHEY, 2018), candidatas a variáveis preditivas do volume de madeira na floresta.

(continua)

Variável	Descrição
Elev minimum	Altura mínima
Elev maximum	Altura máxima
Elev mean	Altura média
Elev mode	Moda da altura
Elev stddev	Desvio padrão da altura
Elev variance	Variância da altura
Elev CV	Coefficiente de variação da altura
Elev IQ	Distância interquartis
Elev skewness	Skewness
Elev kurtosis	Kurtosis
Elev AAD	Média do desvio absoluto
Elev L1	L-moment 1
Elev L2	L-moment 2
Elev L3	L-moment 3
Elev L4	L-moment 4
Elev L CV	Coefficiente de variação do L-moment
Elev L skewness	L-moment Skewness
ElevL kurtosis	L-moment Kurtosis
Elev P01	1° Percentil
Elev P05	5° Percentil
Elev P10	10° Percentil
Elev P20	20° Percentil
Elev P25	25° Percentil

Tabela 3 – Métricas derivadas do escaneamento aéreo com laser (MCGAUGHEY, 2018), candidatas a variáveis preditivas do volume de madeira na floresta.

Variável	Descrição
Elev P30	30° Percentil
Elev P40	40° Percentil
Elev P50	50° Percentil
Elev P60	60° Percentil
Elev P70	70° Percentil
Elev P75	75° Percentil
Elev P80	80° Percentil
Elev P90	90° Percentil
Elev P95	95° Percentil
Elev P99	99° Percentil

(conclusão)

Fonte: do autor (2022)

Os valores dos percentis foram calculados com a elevação dos pontos da nuvem. Inicialmente, os valores observados da nuvem de pontos (x_i) foram colocados em ordem crescente, após esta etapa, aplica-se a Equação 1.

$$\text{Percentil} = (n - 1) * p = i + f \quad (1)$$

i é a parte inteira de $(n - 1) * p$

f é a parte fracionária de $(n - 1) * p$

Onde: n é o número de observações e p é o valor do percentil de interesse.

Se $f = 0$ então o $\text{Percentil} = x_{i+1}$

Se $f > 0$ então o $\text{Percentil} = x_{i+1} + f * (x_{i+2} - x_{i+1})$

As métricas de Skewness (assimetria) e Kurtosis (curtose) são conhecidas como terceiros e quartos momentos de produto de uma distribuição. A assimetria fornece alguma indicação de quão assimétrica é a distribuição e a curtose fornece uma indicação de quão “pontaguda” é essa distribuição. Infelizmente, essas estatísticas calculadas usando momentos de produto mostram viés e variância consideráveis, dependendo do tamanho da amostra e da

presença de outliers, isto pode dificultar sua utilização como variáveis preditoras. A Skewness e a Kurtosis foram calculadas pelas Equações 2 e 3, respectivamente.

$$Skewness = \frac{\sum_{i=1}^N (Y - \bar{Y})^3}{(N-1) * S^3} \quad (2)$$

$$Kurtosis = \frac{\sum_{i=1}^N (Y - \bar{Y})^4}{(N-1) * S^4} \quad (3)$$

Onde: \bar{Y} é a média, S é o desvio padrão e N é o número de observações.

A métrica Média do Desvio Absoluto (do inglês *Average Absolute Deviation* AAD) foi calculada de acordo com a Equação 4.

$$AAD = \frac{\sum_{i=1}^N (|Y_i - \bar{Y}|)}{N} \quad (4)$$

Os L-moment 1, 2, 3 e 4 são um conjunto de estatísticas mais robusto, baseado no método dos momentos L, foi apresentado por Hosking (1990). Os momentos L são calculados usando combinações lineares de valores de dados ordenados (elevação do primeiro retorno LIDAR). As razões dos momentos L fornecem estatísticas que são comparáveis à variância, assimetria e curtose descritas acima. O primeiros quatro momentos L (λ_1 , λ_2 , λ_3 , λ_4) foram estimados usando os estimadores de amostra direta propostos por Wang (1996) e as razões de momento L correspondentes ao Coeficiente de variação (τ_2), Skewness (τ_3) e Kurtosis (τ_4) que foram calculados da forma apresentada nas Equações 5, 6 e 7, respectivamente.

$$\tau_{2=\frac{\lambda_2}{\lambda_1}} \quad 0 < \tau_2 < 1 \quad (5)$$

$$\tau_{3=\frac{\lambda_3}{\lambda_2}} \quad -1 < \tau_3 < 1 \quad (6)$$

$$\tau_{4=\frac{\lambda_4}{\lambda_2}} \quad -\frac{1}{4}(5\tau_3^2 - 1) \leq \tau_4 < 1 \quad (7)$$

3.2.2 Modelagem dos dados ALS

3.2.2.1 Coleta e processamento de dados nas parcelas de calibração

A obtenção dos dados para o processamento do inventário ALS iniciou-se pela alocação de unidades amostrais em campo, que foram distribuídas de maneira aleatória nos diferentes talhões da floresta.

Após a alocação das parcelas, foram realizadas medições de variáveis dendrométricas, sendo circunferência à altura do peito (CAP) — a 1,3 metros de altura do solo — de todas as árvores, altura total das 10 primeiras árvores da parcela e a altura dominante de 4 árvores, elas estavam incluídas, ou não, nas dez primeiras. A altura coodominante da parcela foi a média de altura das 4 árvores dominantes medidas na parcela. Foram consideradas dominantes as árvores que possuíam o maior diâmetro dentro da parcela, conforme conceito de Assmann. Estas árvores apresentavam boa forma e bom aspecto geral, não sendo bifurcadas. Os levantamentos nas parcelas foram realizados em abril e maio de 2019.

Paralelamente, foram realizadas cubagens rigorosas em árvores nos diferentes talhões da floresta, lembrando que estas cubagens foram realizadas com base em uma amostragem representativa da população alvo. Para esta etapa, foi amostrado um total de 5 árvores por classe de diâmetro, sendo de 3 cm o intervalo de diâmetro de cada classe. O número de classes foi gerado pela distribuição diamétrica das árvores inventariadas. O método de cubagem utilizado foi o de Smalian. Este é um método destrutivo, onde a árvore é abatida e mede-se os diâmetros e as espessuras de casca em 17 alturas na árvore, relativas a 0, 1, 2, 3, 4, 5, 10, 15, 25, 35, 45, 55, 65, 75, 85 e 95% da altura total e no diâmetro comercial que equivale a 6 centímetros. A partir dos dados obtidos nesta etapa, fez-se os ajustes das Equações 8 e 9. A Equação 8 foi o modelo hipsométrico proposto por Campos (1984) e foi utilizada para estimar a altura total. Já o modelo volumétrico foi ajustado para a combinação de região do plantio, clone e classe de idade, utilizando a Equação 9, modelo proposto por Leite, Guimarães e Campos (1995), para estimar o volume das árvores sem a casca em função do diâmetro à altura do peito (DAP) e da altura total.

$$\widehat{Ht}_{ij} = \exp \left(\beta_0 + \frac{\beta_1}{DAP_{ij}} + \beta_2 * (Hdom_j) \right) \quad (8)$$

Onde:

$\widehat{H}_{t_{ij}}$ = Altura estimada de cada árvore em cada unidade amostral;

DAP = Diâmetro Altura do Peito;

\ln = Logaritmo natural;

H_{dom} = Altura dominante da parcela;

β_i = parâmetros a serem estimados.

$$\widehat{V}_i = \beta_0 * Dap^{\beta_1} * Ht^{\beta_2} * \exp\left(\beta_3 * \frac{tx}{Dap}\right) * \left[1 - \left(\frac{d_i}{Dap}\right)^{1+\beta_4*d_i}\right] \quad (9)$$

Onde:

\widehat{V}_i : Volume individual da árvore com ou sem casca;

d_i : Diâmetro comercial;

Dap : Diâmetro a altura do peito;

Ht : Altura total;

$tx = 0$: para volume com casca;

$tx = 1$: para volume sem casca;

β_i : Parâmetros a serem ajustados.

Os coeficientes dos modelos são de propriedade intelectual da empresa, pois para o ajuste dos mesmos utilizou-se não apenas as medições realizadas neste estudo, mas também uma extensa base de dados com os inventários realizados rotineiramente.

Após o ajuste dos modelos volumétricos, estimou-se os volumes sem casca para cada uma das árvores nas unidades amostrais, somando os valores de volume sem casca das árvores foi possível determinar o volume de madeira das unidades amostrais.

3.2.2.2 Treinamento e seleção dos algoritmos

O treinamento dos algoritmos foi realizado a partir das 33 diferentes métricas, obtidas na etapa de Extração de Métricas da nuvem de pontos, as quais estão expressas na Tabela 3. Todas as métricas foram normalizadas com o modelo digital de elevação criado nas etapas anteriores, para que não existisse o efeito do relevo na nuvem de pontos. A manipulação e o

processamento foram realizados nos pacotes Pandas (MCKINNEY,2010) e Scikit-learn (PEDREGOSA, 2011) na linguagem Python.

A estimativa do volume de madeira sem casca e a criação dos mapas de distribuição espacial do volume nas áreas florestais foram realizadas por meio de dois algoritmos de aprendizagem: *Random Forest*, utilizando as 33 métricas; *Support Vector Machine*, utilizando as principais métricas selecionadas pelo *Random Forest Selection*. Os resultados foram comparados ao modelo linear ajustado pelo método *Ordinary Least Squares*, o qual é tradicionalmente aplicado em levantamentos florestais (Silva et al., 2016).

A função *RandomForestRegressor* foi aplicada para ajustar o algoritmo *Random Forest* considerando 100 árvores na floresta. Os valores dos parâmetros foram o padrão do pacote Scikit-learn (PEDREGOSA, 2011) versão 0.24.1. Para o *Support Vector Machine* foi utilizado o Pacote do *Scikit-learn SVM* na versão 0.24.1 como um regressor. Com o objetivo de melhorar a precisão das estimativas, foram utilizadas no modelo SVM apenas as métricas selecionadas pelo modelo *Random Forest* usando o *SelectFromModel* do *Scikit-learn*. Apesar de existirem diferentes funções Kernel para o modelo SVM (polinomial, radial, linear e sigmoide), a função de Kernel utilizada no presente estudo foi a função linear, pois após testes com a base de dados, foi a função que apresentou o melhor resultado. O valor do parâmetro C selecionado foi de 1.

Para seleção das variáveis no modelo linear, foram feitos os testes de normalidade da variável dependente e das variáveis independentes, também foi realizado o teste de homocedasticidade do modelo escolhido. O método utilizado para os teste de normalidade e homocedasticidade foi Shapiro-Wilk (Shapiro & Wilk, 1965), com valor nominal de significância de 5%. Após os testes de normalidade aplicou-se o modelo linear com o método de ajuste *Ordinary Least Squares* utilizado o método de *Stepwise Forward*.

Para determinar a precisão do modelo, foi escolhida a análise do Erro quadrático médio percentual (RMSE%) e o coeficiente de determinação R². A Equação 10 representa o RMSE percentual. A Equação 11 representa o R².

$$RMSE(\%) = 100 * \frac{\sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}}}{\bar{Y}} \quad (10)$$

RMSE(%) : Erro quadrático médio percentual;

Y_i : Valor da observação;

\hat{Y}_i : Valor estimado;

n : Número de observações;

\bar{Y} : Valor médio das observações.

$$R^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} \quad (11)$$

R^2 : Coeficiente de determinação;

Y_i : Valor da observação;

\hat{Y}_i : Valor estimado;

\bar{Y} : Valor médio das observações.

Para a validação do modelo proposto, a base de dados contendo as 39 parcelas estudadas foi dividida de forma aleatória em: 70% das parcelas usadas para o treinamento e 30% para validar o algoritmo treinado. Este método de validação é conhecido como *Hold-one-out*. Este processo foi repetido por 200 vezes e então foi extraída a média percentual, o desvio padrão e o coeficiente de variação dos erros quadráticos médios de cada uma das repetições. O modelo escolhido foi o modelo com o menor RMSE% e maior R^2 após as 200 repetições do treinamento.

3.2.2.3 Extração de métricas da nuvem de pontos na área dos talhões

A extração de métricas da nuvem de pontos na área dos talhões foi a etapa de pré-processamento para posterior expansão das estimativas de volume de madeira para os locais sobrevoados durante o levantamento. Para tanto, foi preciso que todos os povoamento estivessem delimitados em arquivos digitais, com isso, foi feito um recorte na nuvem de pontos captada pelo sensor, eliminando os pontos que não estavam contidos dentro dos limites dos talhões, excluindo, por exemplo, estradas, aceiros, área de proteção permanentes e reservas legais.

A extração de métricas nos talhões foi realizada no software FUSION (MCGAUGHEY, 2018) dividindo a área total dos talhões em células de 20x20 metros. A escolha do tamanho das células foi feita para que não fosse necessário realizar transformações nas métricas extraídas, pois a área de cada uma das células coincide com a área das parcelas utilizadas para a calibração dos modelos. A nuvem de pontos foi dividida de maneira que em cada uma das células existisse um subconjunto da nuvem de pontos que representasse a leitura do sensor na área da célula. Os valores de elevação dos pontos da nuvem foram todos

normalizados pelo modelo digital de elevação criado na Seção 3.2.1.2. Este processo foi realizado com a função *GridMetrics* do software FUSION (MCGAUGHEY, 2018), utilizando os parâmetros *outlier*: 0.8 a 50, *minht*: 1.3 e *first* (usa apenas os primeiros retornos do ALS). As métricas extraídas e selecionadas no *grid* foram iguais às métricas extraídas e selecionadas na Seção 3.2.1.3, todas estão descritas na Tabela 3.

3.2.2.4 Aplicação dos modelos

Decidido o modelo mais preciso, com base no RMSE% e no R^2 , para estimar o volume de madeira sem casca nas florestas. Criou-se os mapas com as distribuições espaciais do volume de madeira nas áreas estudadas utilizando este modelo. As células nos mapas criados têm 20x20 metros, mesma dimensão do *grid* criado na Seção 3.2.2.3. Os modelos estimaram o volume de cada uma das células do *grid*, para definir o volume de madeira dos talhões no método de inventário ALS foram somados os valores de volume de todas as células que estavam dentro ou nas bordas dos limites dos talhões.

3.2.3 Inventário convencional

3.2.3.1 Coleta e processamento de dados nas parcelas convencionais

Posterior a escolha de qual o melhor modelo para a estimativa de volume nas florestas utilizando os dados do sensor laser, foi feito o inventário convencional destas mesmas áreas.

O inventário florestal convencional foi baseado em uma amostragem padrão realizada pela empresa, que consiste basicamente na alocação de parcelas retangulares de área variável com um total de 22 árvores por parcela. Utilizou-se a amostragem casual simples para a alocação destas parcelas, sendo que a intensidade amostral variou para cada área estudada. Foram realizadas as medições e processamentos citados no item 2.2 deste Capítulo. A medição das parcelas para o inventário florestal convencional foi feita em junho de 2019.

3.2.4 Mensuração da madeira na fábrica

A quantidade de madeira colhida foi mensurada através dos carregamentos de madeira que chegavam até a fábrica quantificada em metros cúbicos sólidos, por meio do método do

empuxo. Para tanto, todos os caminhões com madeira foram pesados na entrada e na saída da fábrica. A diferença entre o peso da entrada e o da saída correspondeu ao massa da madeira. Para conversão deste peso em volume sólido, foi determinada a “relação peso volume” – RPV.

O sistema de pesagem é composto por uma balança rodoviária e um tanque com água. A RPV é determinada por amostragem, onde uma grua vazia sobe até a balança e realiza a tara do equipamento. Em seguida, vai até a balança carregada com um feixe de madeira que é pesada no ar. Na sequência o feixe é totalmente mergulhado em um tanque com água. A força da água contrária à madeira é medida na balança e corresponde ao empuxo que, por sua vez, representa o volume da madeira.

A RPV é determinada pela razão entre o peso e o volume da madeira amostrada. A cada caminhão com madeira pesada será aplicada a RPV vigente para o talhão de origem, que converterá o peso da madeira em volume.

A amostragem foi realizada para o primeiro caminhão de cada um dos talhões que entrou na fábrica e o valor obtido foi considerado válido por 72 horas, ou para os próximos 1.350 m³sc de madeira transportados. Após este período/volume uma nova amostragem deve ser realizada. A amostra corresponde a um feixe de madeira de um compartimento do caminhão.

4 RESULTADOS E DISCUSSÃO

4.1 Inventário com as parcelas de calibração

O inventário das parcelas de calibração é uma demonstração de como estavam as florestas de cada uma das áreas estudadas no momento do levantamento. Os detalhes deste inventário estão expostos na Tabela 4.

Tabela 4 – Informações de inventário florestal com as parcelas usadas para a calibração de modelagem do Inventário ALS.

Área estudada	Dap médio (cm)		Altura média (m)		Volume médio (m ³ /ha)		Erro padrão da média (m ³ /ha)	Erro do Inventário	
	Média	CV%	Média	CV%	Média	CV%		(m ³ /ha)	(m ³ /ha)
Área 1	16,2	7,11%	27,4	6,52%	199,97	21,37%	13,52	27,4	13,72%
Área 2	15,5	7,39%	26,4	4,24%	191,62	15,54%	6,83	13,5	7,03%
Área 3	15,9	4,20%	25,5	5,04%	167,60	13,27%	7,04	15,1	9,00%

Fonte: do autor (2022)

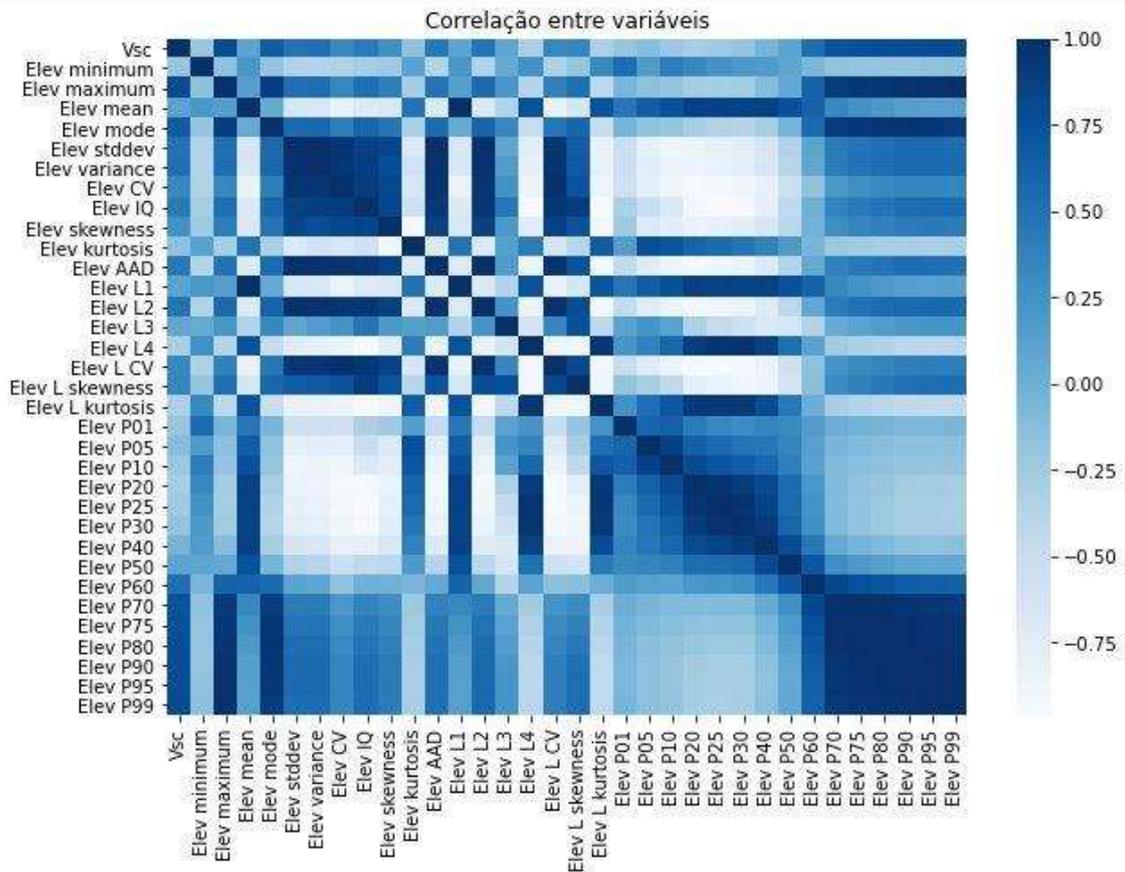
Na Tabela 4 pode-se observar como as florestas das Áreas 1 e 2 tiveram volumes de inventário muito similares. Isso ocorreu devido à semelhança de idade entre as florestas, 5 anos. Já a floresta da Área 3, com 4 anos, apresentou um volume de inventário menor. Apesar da diferença de idade e volume entre as áreas estudadas, as florestas apresentaram algumas características similares, como DAP médio e altura total média. Importante ressaltar que não é necessário que as florestas sejam similares para a aplicação do método de inventário ALS. Essa similaridade influenciará na precisão do resultado. O estudo de Silva et al. (2016) mostrou como o sensor laser reage de maneira diferente com as idades das florestas, mas neste estudo essa diferença entre as áreas estudadas não foi impeditiva à utilização da modelagem utilizando os dados ALS. Caso as áreas fossem muito desiguais, seria interessante separá-las para o processamento do inventário.

4.2 Treinamento e seleção dos modelos

Inicialmente foi realizada uma análise exploratória das métricas do ALS por meio da matriz de correlações, no intuito de avaliar correlações entre as métricas e a variável de interesse

(i.e., volume de madeira sem casca) e as correlações entre as próprias métricas (colinearidade) (Figura 3).

Figura 3 – Matriz de correlação entre as métricas extraídas no ALS e o volume de madeira sem casca.



Legenda das métricas: Vsc: volume de madeira sem casca, Elev minimum: altura mínima, Elev maximum: altura máxima, Elev mean: altura média, Elev mode: moda da altura, Elev stddev: desvio padrão da altura, Elev variance: variância da altura, Elev CV: coeficiente de variação da altura, Elev IQ: distância entre quartis, Elev skewness: altura Skewness, Elev kurtosis: altura Kurtosis, Elev AAD: média do desvio absoluto, Elev L1: L-moment 1, Elev L2: L-moment 2, Elev L3: L-moment 3, Elev L4: L-moment 4, Elev L CV: coeficiente de variação do L-moment, Elev L skewness: L-moment Skewness, ElevL kurtosis: L-moment Kurtosis, Elev P01: 1° percentil, Elev P05: 5° percentil, Elev P10: 10° percentil, Elev P20: 20° percentil, Elev P25: 25° percentil, Elev P30: 30° percentil, Elev P40: 40° percentil, Elev P50: 50° percentil, Elev P60: 60° percentil, Elev P70: 70° percentil, Elev P75: 75° percentil, Elev P80: 80° percentil, Elev P90: 90° percentil, Elev P95: 95° percentil, Elev P99: 99° percentil. Fonte: do autor (2022).

A correlação de Pearson entre o volume de madeira sem casca e o 99° percentil da nuvem de pontos normalizada foi de 0,8005. As demais métricas que apresentaram alta correlação ($> 0,70$) foram os 95°, 90°, 80°, 75°, 70° percentis e altura máxima. Por outro lado, a análise entre as métricas (i.e., entre as variáveis explicativas) mostrou alta correlação do 99° percentil com as métricas Elev maximum (0,995), Elev mode (0,915), Elev P70 (0,922), Elev P75 (0,951), Elev P80 (0,967), Elev P90 (0,986), Elev P95 (0,). Tais correlações indicam

colinearidade entre as variáveis explicativas. Dessa forma, o uso de todas as variáveis fornecidas pelo ALS na modelagem pode diminuir a capacidade preditiva dos modelos (CAMMAROTA, C., & PINTO, A., 2021).

Os testes de Shapiro-Wilk (Shapiro & Wilk, 1965) para as variáveis independentes e para o volume de madeira indicaram, com 95% de confiança, que a distribuição das variáveis segue uma distribuição Gaussiana, possibilitando que as mesmas pudessem ser utilizadas no modelo linear. O método *Stepwise* mostrou que o modelo linear melhor explicaria o volume de madeira com apenas uma das métricas do ALS, o 99º percentil (Elev P99). Este resultado corrobora o estudo de Silva et al. (2016), que encontrou forte relação entre a métrica e o volume de madeira em florestas de eucalipto no estado de São Paulo. Ademais, as ressalvas com relação à colinearidade não são preocupantes, uma vez que apenas uma variável foi selecionada pela metodologia para a construção do modelo.

Feito isso, os três modelos (OLS, RF e SVM) foram treinados e validados por 200 vezes, dividindo a base de dados de forma aleatória em 70% para o treinamento do algoritmo e 30% para a validação em cada uma das vezes. Assim, a capacidade preditiva dos modelos foi averiguada a partir da média entre as 200 repetições das seguintes estatísticas: coeficiente de determinação (R^2), erro quadrático médio (RMSE), desvio padrão percentual e o coeficiente de variação dos erros quadráticos médios, todos estes valores estão expostos na Tabela 5.

Tabela 5 – Erro quadrático médio (RMSE e RMSE%), desvio padrão do erro quadrático médio, coeficiente de variação dos erros quadráticos médios (CV%) e coeficiente de determinação após o treinamento e validação dos algoritmos, utilizando as 39 parcelas.

Modelos (39 parcelas e 200 repetições)			
	RF	OLS	SVM
RMSE (m³/ha)	22,66	20,19	22,54
RMSE% Média	12,37%	10,90%	12,34%
Desvio padrão	1,92%	1,65%	2,29%
CV%	15,51%	15,15%	18,53%
R²	43,36%	53,74%	42,45%

Fonte: do autor (2022)

O modelo que obteve o melhor desempenho foi o OLS, com menor RMSE%, maior R^2 , além de mostrar-se mais estável que os demais, apresentando uma menor variação entre cada treinamento e validação realizada. Quando o modelo OLS é treinado utilizando um maior número de métricas, pode ocorrer um decréscimo em sua precisão devido à sua sensibilidade à colinearidade, sendo isso uma limitação do método OLS. Os autores Silva et al.(2015, 2016)

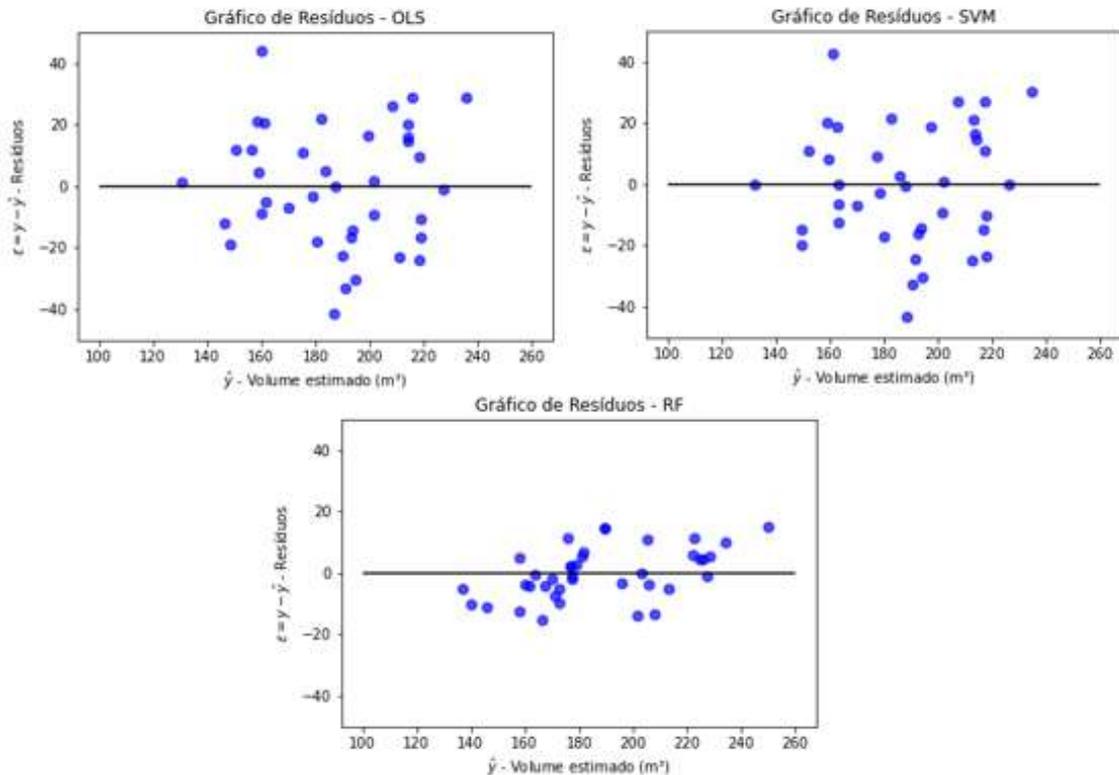
encontraram erros quadráticos médios menores que 15% em florestas clonais de eucalipto, utilizando o método baseado em área e o modelo linear com diferentes variáveis explicativas, sendo que dentre as variáveis que melhor descreviam o volume de madeira estava o 99º percentil.

Com base nos resultados encontrados, o presente estudo corrobora com o trabalho de Cosenza et al. (2020), onde os autores compararam três diferentes métodos — *Ordinary Least Squares*, *Random Forest* e *K-nearest Neighbors* — para estimar o volume em florestas com dados de escaneamento aéreo com sensor laser. Os autores dividiram as florestas estudadas em grupos, existindo um grupo de florestas plantadas de eucalipto no Brasil. Os autores comprovaram que o método que melhor estimou as florestas foi o OLS, com RMSE% próximos de 10%.

Por sua vez, os melhores resultados do modelo linear deste estudo contradizem o estudo feito por GARCÍA-GUTIÉRREZ et al. (2015), realizado com dados de LiDAR em florestas de eucalipto e pinus espanholas, nele os autores constataram que, entre os algoritmos SVM, RF e OLS, o melhor desempenho foi do SVM com kernel gaussiano. No presente estudo, o modelo SVM apresentou comportamento similar ao modelo OLS, principalmente na criação dos mapas, apesar do modelo OLS apresentar resultados de R^2 e RMSE melhores. Isso comprova que ainda não existe consenso sobre o melhor algoritmo para a modelagem do volume de madeira em florestas utilizando dados de ALS, apontando para oportunidades de novos trabalhos.

Para comprovar mais um dos pressupostos da modelagem linear, foi testada a homocedasticidade das estimativas geradas pelo modelo OLS, usando, mais uma vez o teste de Shapiro-Wilk (Shapiro & Wilk, 1965). O teste demonstrou a homogeneidade da variância do erro das estimativas, ou seja, a homocedasticidade do modelo. O gráfico de distribuição dos resíduos foi feito treinando os modelos com toda a base de dados e comparando estes valores com os valores observados, os gráficos estão apresentados na Figura 3.

Figura 3 – Gráficos de Resíduos para os três diferentes modelos estudados.



Fonte: do autor (2022)

Nesta etapa, pode-se constatar como os resíduos do modelo RF apresentam uma característica mais uniforme em relação aos demais modelos, concentrando os erros das estimativas mais próximos de 0. Porém, isso pode ter ocorrido devido um possível *overfitting* do modelo RF, que faz com que o modelo “decore” os valores a serem estimados. A causa do possível *overfitting* pode ser a pequena quantidade de unidades amostrais utilizada neste estudo.

A Equação 12 representa o modelo linear OLS ajustado para estimar o volume de madeira sem casca em relação à métrica 99º percentil do ALS, essa equação foi obtida com o treinamento utilizando toda a base de dados (39 parcelas).

$$V_{sc} = 15,04893668 * Elev P99 - 226,0396998063611 \quad (12)$$

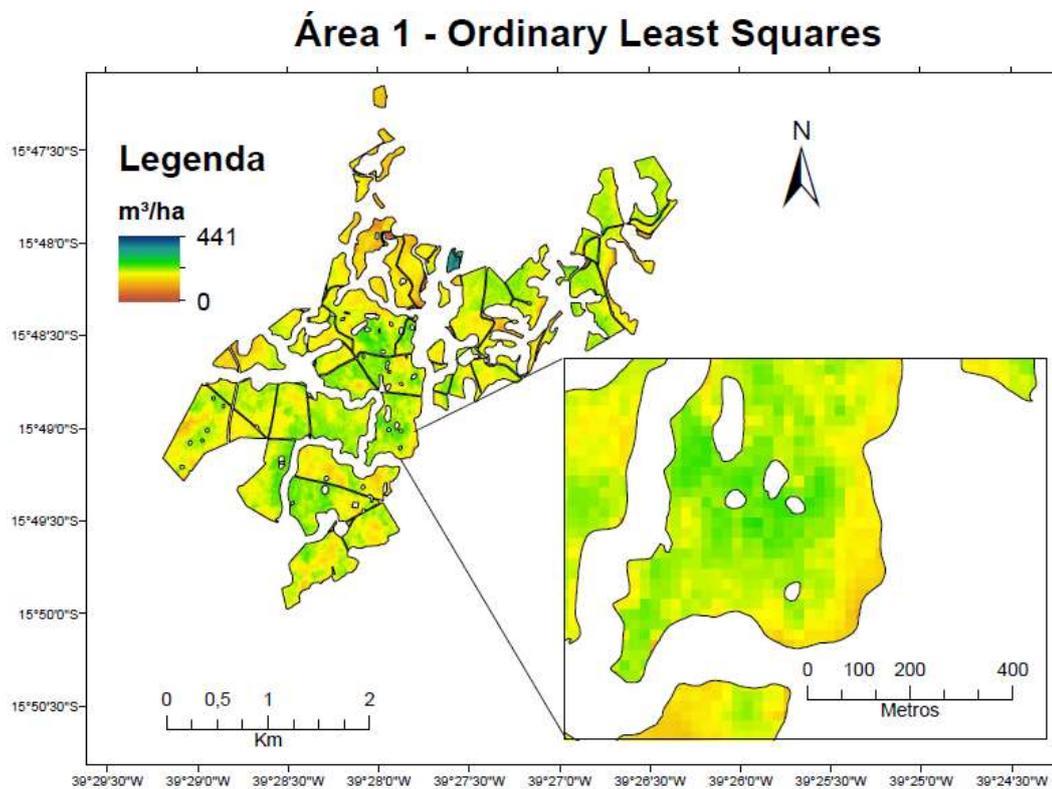
V_{sc} : Volume de madeira sem casca;

$Elev P99$: Métrica de 99º percentil do ALS.

4.3 Mapas de volume de madeira dos modelos

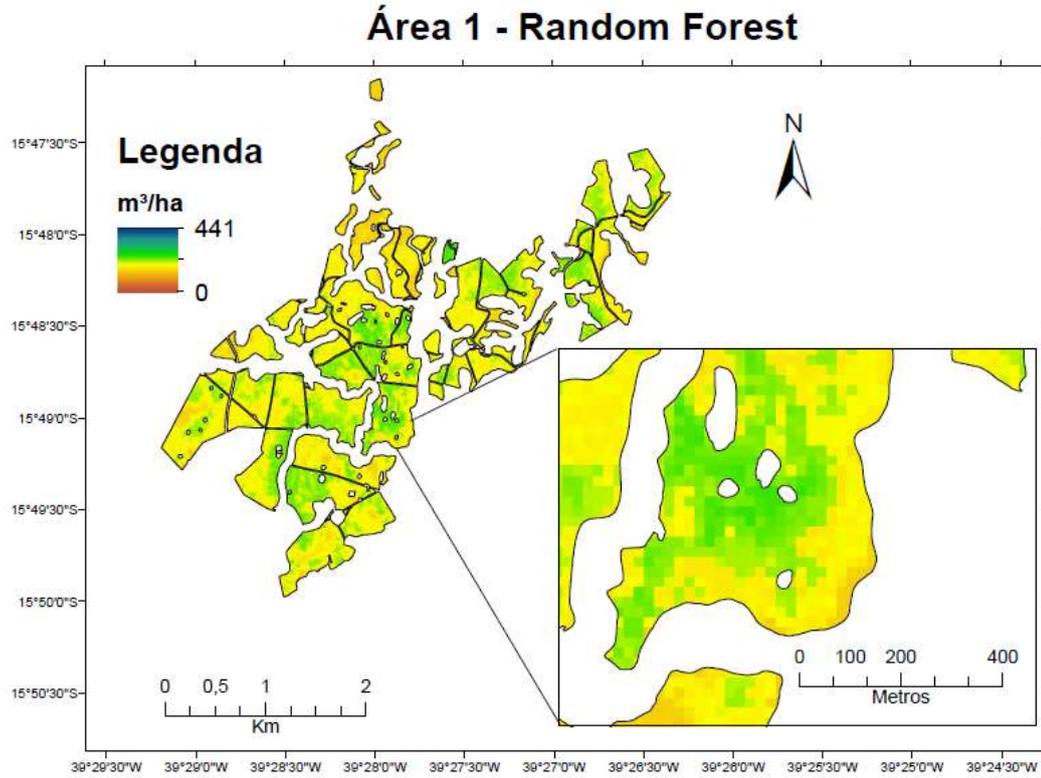
A fim de aplicar os modelos treinados, foram criados os mapas de cada um dos modelos nas 3 áreas estudadas. Os mapas apresentados nas Figuras 4, 5 e 6 são os mapas de volume de madeira para a Área 1 utilizando os três modelos, as Figuras 7, 8 e 9 são os mapas utilizando os três modelos para Área 2 e as Figuras 10, 11 e 12 são os respectivos mapas para a Área 3.

Figura 4 – Mapa da Área 1 estimado pelo modelo linear OLS.



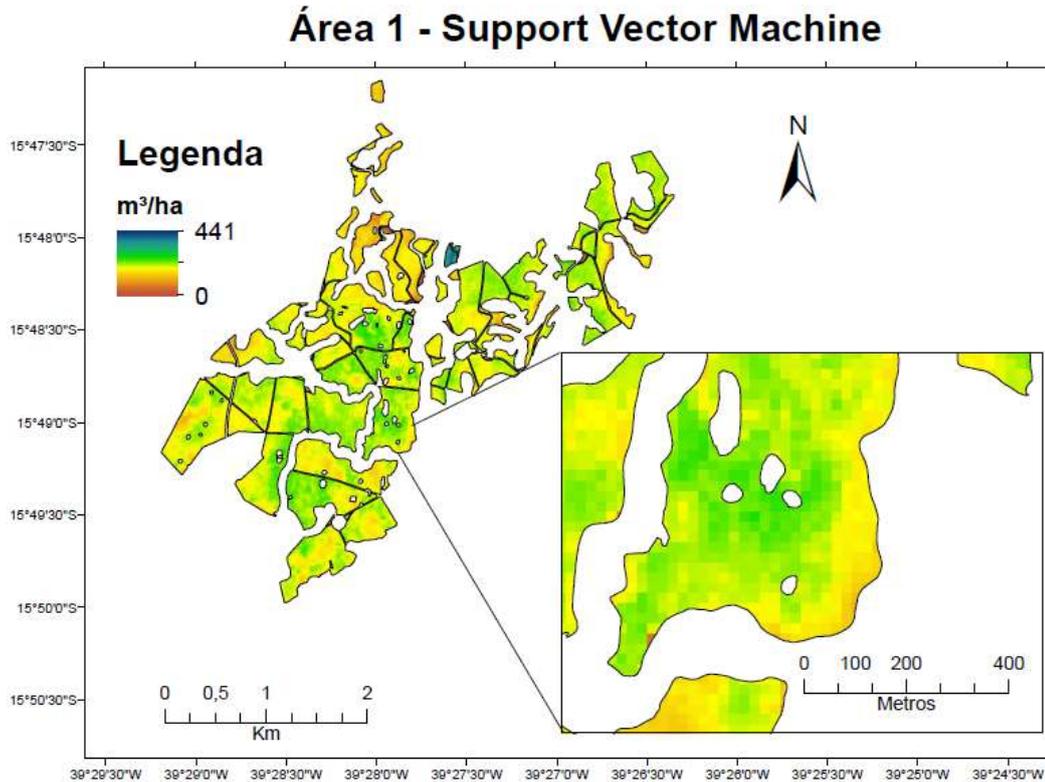
Fonte: do autor (2022)

Figura 5 – Mapa da Área 1 estimado pelo modelo RF.



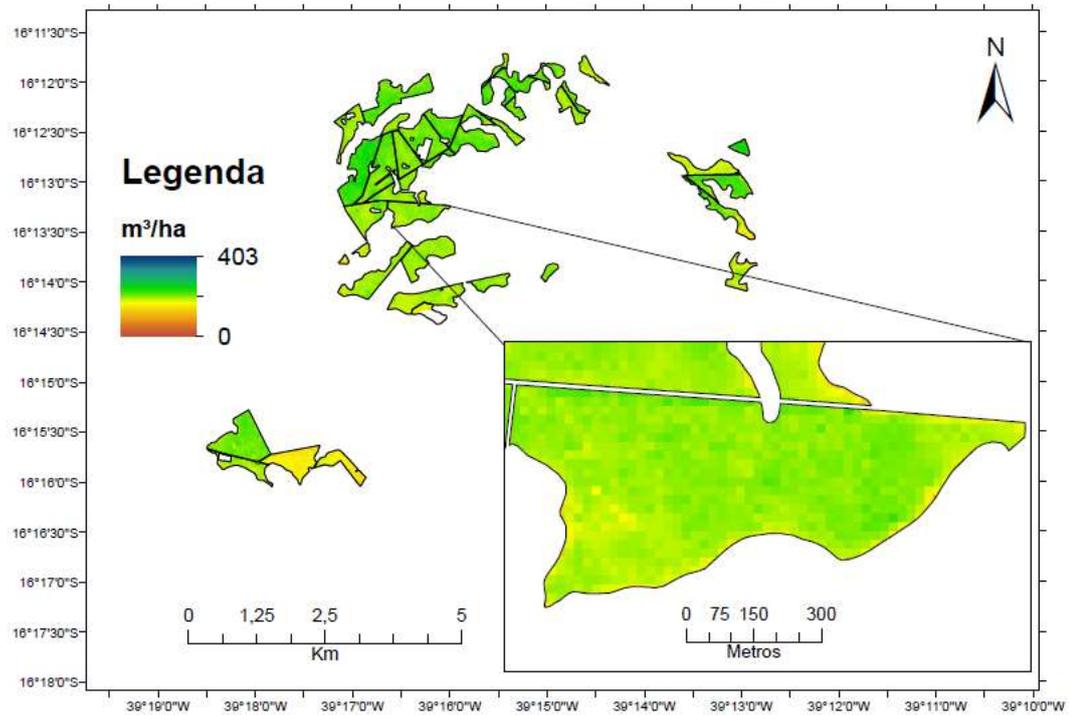
Fonte: do autor (2022)

Figura 6 – Mapa da Área 1 estimado pelo modelo SVM.



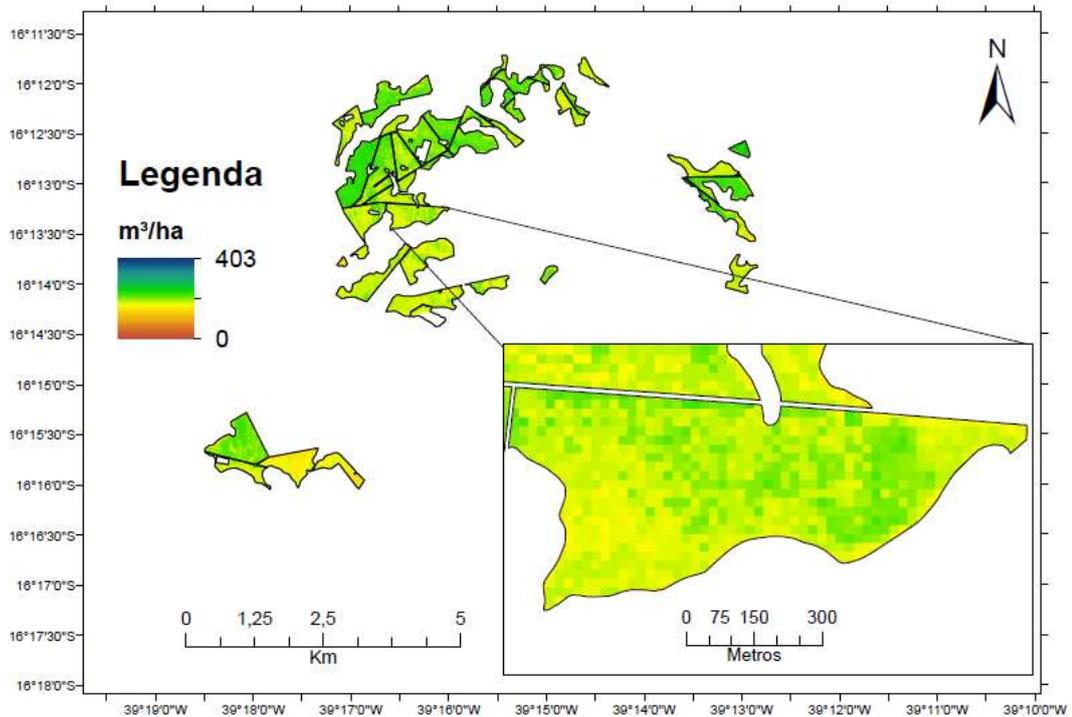
Fonte: do autor (2022)

Figura 7 – Mapa da Área 2 estimado pelo modelo linear OLS.
Área 2 - Ordinary Least Squares



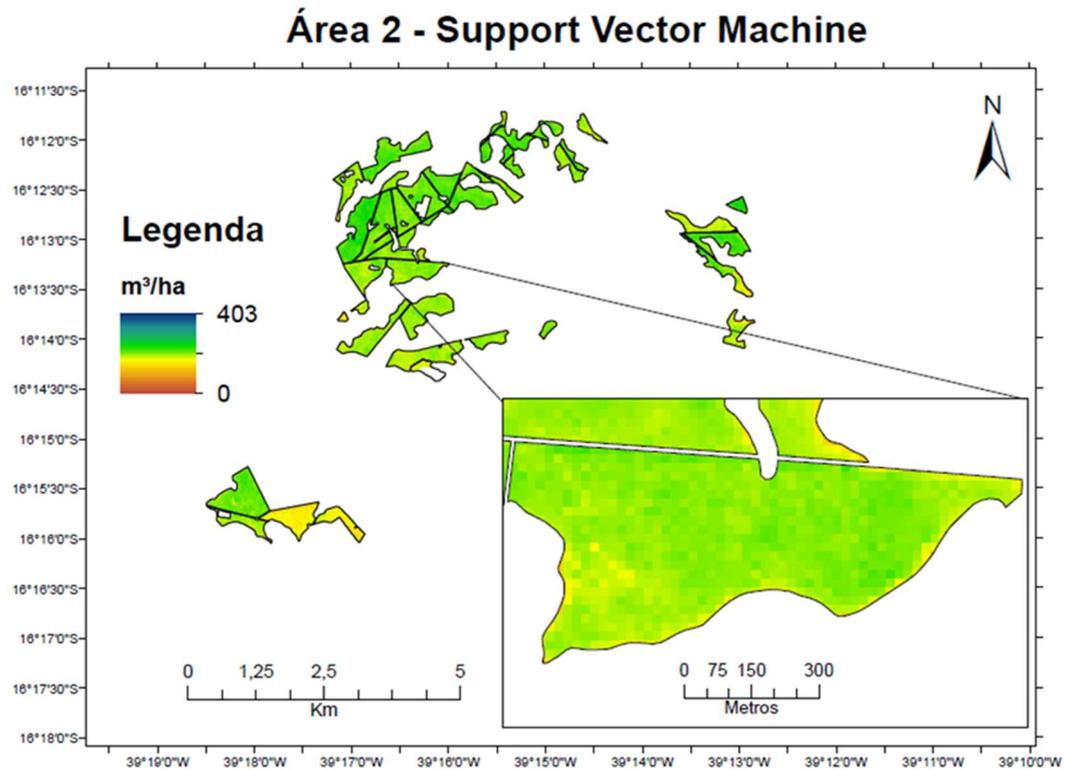
Fonte: do autor (2022)

Figura 8 – Mapa da Área 2 estimado pelo modelo RF.
Área 2 - Random Forest



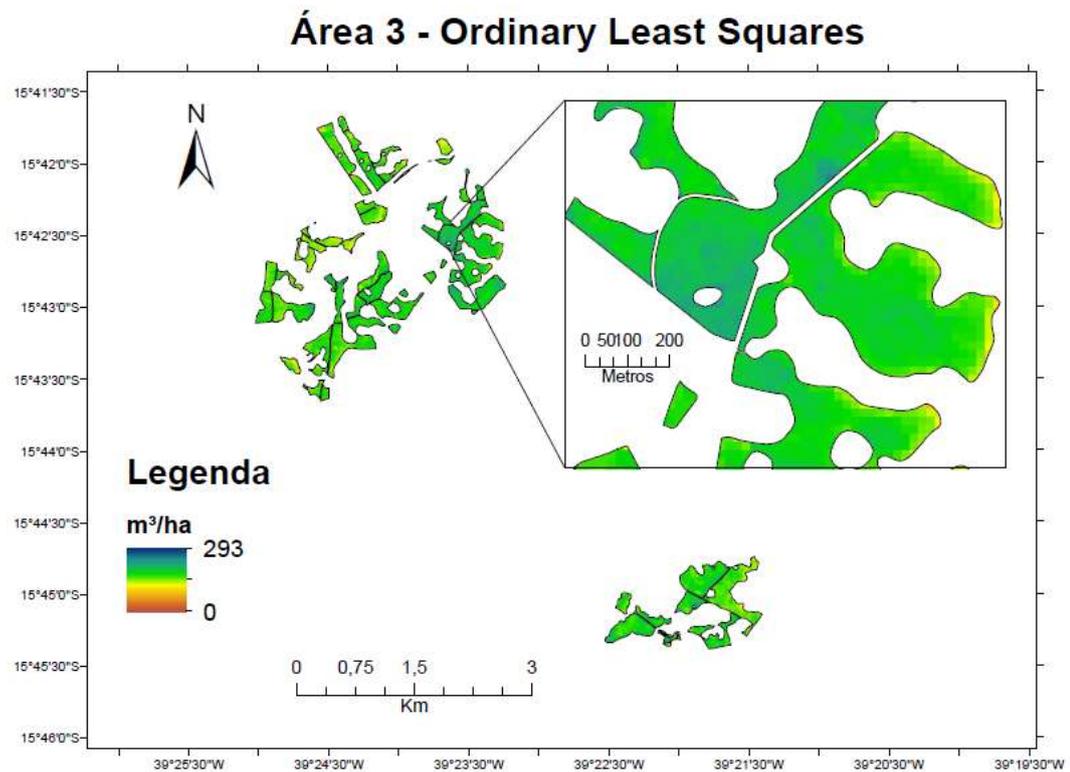
Fonte: do autor (2022)

Figura 9 – Mapa da Área 2 estimado pelo modelo SVM.



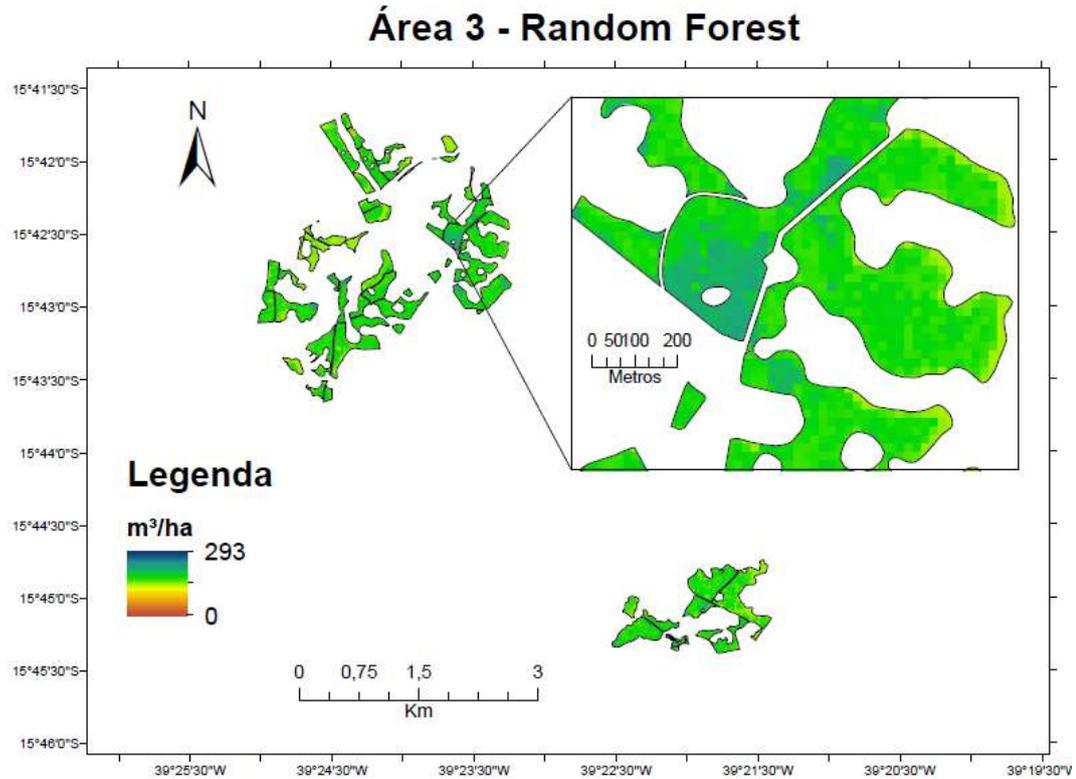
Fonte: do autor (2022)

Figura 10 – Mapa da Área 3 estimado pelo modelo linear OLS.



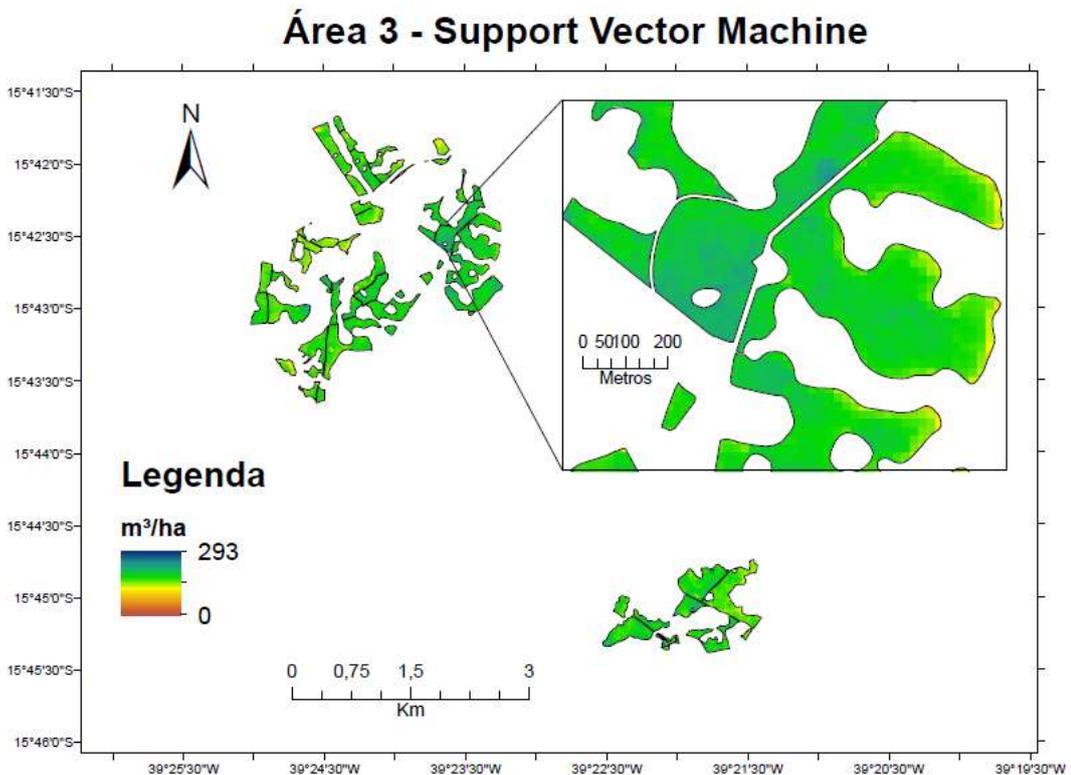
Fonte: do autor (2022)

Figura 11 – Mapa da Área 3 estimado pelo modelo RF.



Fonte: do autor (2022)

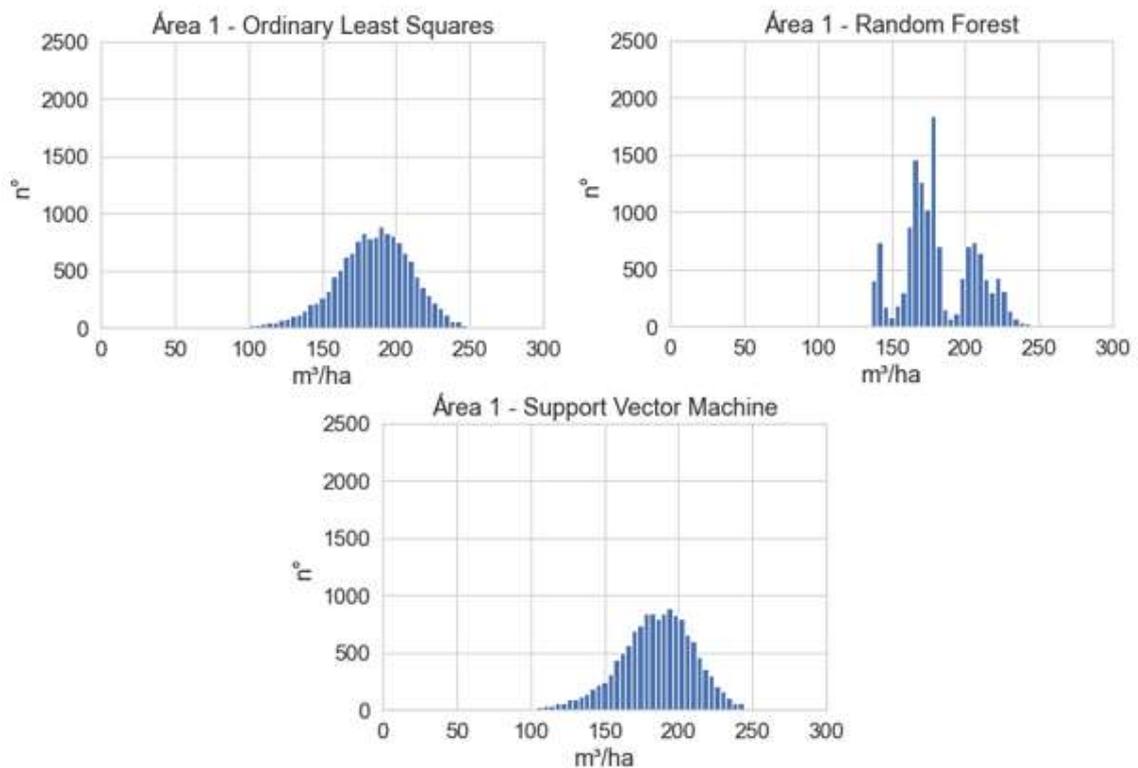
Figura 12 – Mapa da Área 3 estimado pelo modelo SVM.



Fonte: do autor (2022)

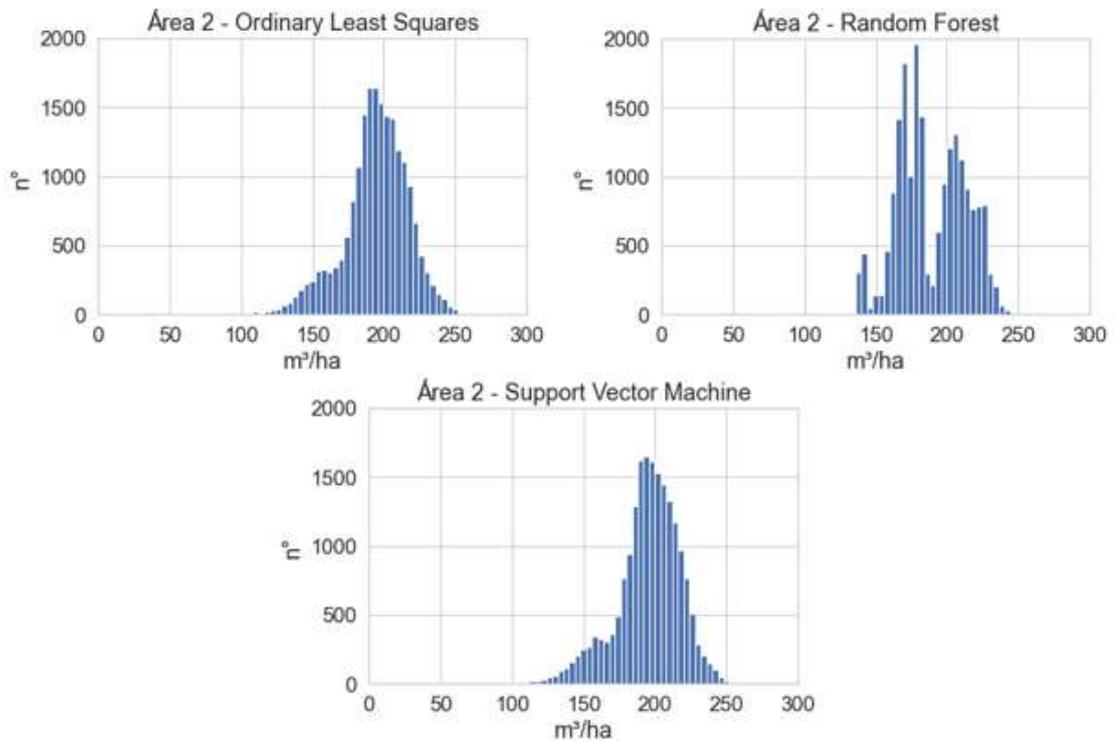
Os mapas de volume criados pelo método de inventário ALS são de grande valia para as tomadas de decisões do setor florestal, pois com eles tem-se a informação espacial dos estoques de madeira nas florestas, demonstrando áreas que estão deficitárias, possibilitando a investigação e solução desses problemas pontualmente, além de apresentar as características de homogeneidade dos diferentes povoamentos. No entanto, visualmente é difícil comparar as diferenças entre os mapas. Portanto, para melhor embasar a indicação de qual modelo deveria ser usado na comparação entre inventários, foi feita a análise dos histogramas dos mapas gerados a partir dos modelos em todas as áreas, eles estão apresentados nas Figuras 13, 14 e 15.

Figura 13 – Histogramas das estimativas de volume de madeira dos três modelos na Área 1 do estudo.



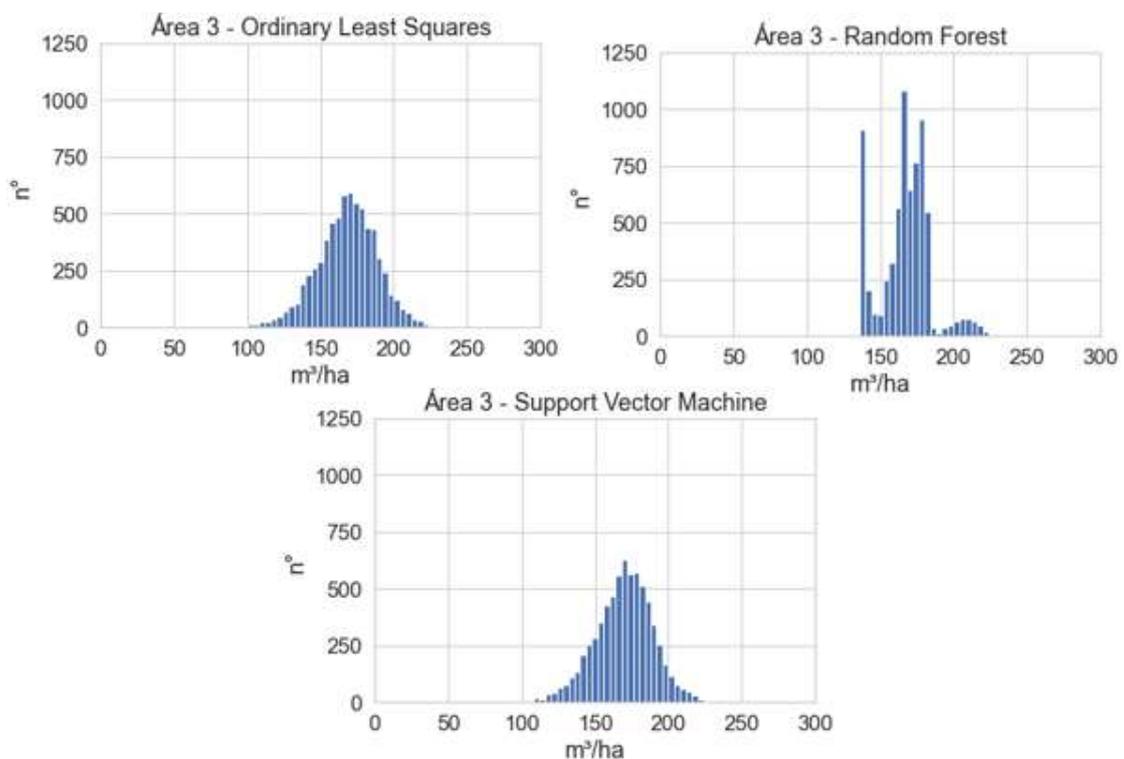
Fonte: do autor (2022)

Figura 14 – Histogramas das estimativas de volume de madeira dos três modelos no Área 2 do estudo.



Fonte: do autor (2022)

Figura 15 – Histogramas das estimativas de volume de madeira dos três modelos no Área 3 do estudo.



Fonte: do autor (2022)

A expectativa era que todos os modelos apresentassem uma distribuição próxima à gaussiana, indicando o comportamento natural das florestas plantadas de eucalipto. Porém isso não foi observado nos histogramas do modelo RF, indicando que ocorreu algum problema nas estimativas do modelo, isso corrobora a ideia de que ocorreu um *overfitting* do modelo RF, indicando que apesar do gráfico de resíduos demonstrar uma assertividade, esta está relacionada a “decora” dos dados de treinamento, exigindo uma maior amostragem para a possível utilização do RF.

Por isso, neste trabalho, o modelo que obteve o melhor desempenho foi o OLS. Assim, o modelo linear OLS foi escolhido para as posteriores comparações entre métodos de inventário florestal.

4.4 Comparação entre os inventários

Os resultados do inventário convencional, do inventário ALS com modelo linear e da quantidade de madeira entregue na fábrica estão apresentados na Tabela 5.

Tabela 5 – Resultados dos inventários florestais pelo método convencional, pelo método ALS/OLS e área onde foi contabilizada a quantidade de madeira através do transporte.

Floresta	Intensidade amostral	Inventário convencional (m³/ha)	Erro do Inventário conv. (m³/ha)	Erro do Inventário conv. (%)	Inventário ALS/OLS (m³/ha)	Madeira transportada (m³/ha)
Área 1	0,453%	182,72	11,183	6,12%	184,74	-
Área 2	0,669%	188,03	4,543	2,42%	194,15	176,98
Área 3	0,348%	143,80	9,427	6,56%	167,99	-

Fonte: do autor (2022)

Os resultados mostram que existe uma grande relação entre os valores dos inventários, demonstrando que a diferença percentual entre o inventário convencional e o inventário ALS nas áreas 1, 2 e 3 foram de respectivamente 1,1%, 3,3% e 16,8%. Com o resultado, pode-se perceber que tanto o inventário convencional quanto o inventário ALS tem pequenos erros associados, no entanto o inventário ALS tendeu a superestimar o volume de madeira nas florestas.

Na expectativa de ter-se informações mais concretas das florestas estudadas, utilizou-se a quantidade de madeira colhida e carregada para a fábrica de uma das áreas, Área 2. Assim, a quantidade de madeira mensurada nos carregamentos foi 6,24% menor que a quantidade

estimada pelo inventário convencional e 9,70% menor que a estimativa do inventário ALS. As diferenças entre os inventários e os carregamentos podem ter ocorrido por possíveis e perdas de madeira durante o manuseio da mesma, como a perda de pontas e erro na altura de corte durante a colheita, sendo detalhes dificilmente controlados.

Apesar de tudo, os inventários são ferramentas indispensáveis a quantificação da madeira, existindo vantagens e desvantagens associadas ao inventário ALS em relação ao inventário florestal convencional que valem ser elencadas. Dentre as desvantagens do inventário ALS há a exigência de um maior poder computacional, sendo um processamento mais lento e não tão difundido pelos técnicos florestais, não existindo nenhum protocolo consolidado no Brasil para a execução do mesmo; além de que ainda existe uma grande dependência das metodologias convencionais de inventário para a calibração dos modelos envolvendo o ALS, sendo impossível eliminar o trabalho das equipes de campo, outro ponto que ainda pode ser uma desvantagem do inventário ALS no Brasil é o preço para a contratação dos levantamentos aéreos, que é maior que o custo de amostragem convencional.

No entanto existem vantagens que mostram quão promissor pode ser o uso dos inventários ALS no futuro, pois o sensor laser coleta informações de toda a floresta, realizando um senso, com isso há uma informação espacial precisa das características dendrométricas, que subsidiarão as decisões na silvicultura de precisão. Também o levantamento ALS é realizado de uma forma muito mais rápida que o tradicional, envolvendo um menor número de pessoas na coleta.

5 CONCLUSÕES

Neste trabalho foram aplicadas e aprimoradas técnicas de análise de dados de escaneamento aéreo com sensor laser (ALS), em florestas clonais de eucalipto plantadas no sul da Bahia. Foram testados três métodos de modelagem (Modelo linear OLS, RF e SVM) e, em termos absolutos, o algoritmo que obteve o melhor desempenho com relação ao erro quadrático médio percentual (RMSE%) e ao coeficiente de determinação (R^2) foi o Modelo linear *Ordinary Least Squares* (OLS) utilizando o 99º percentil da nuvem de pontos.

Os mapas de volume de madeira são ferramentas úteis para os gestores florestais, podendo embasar decisões relacionadas a floresta. Os inventários convencionais e ALS mostraram ter resultados similares para as condições nas quais foram comparados. Contudo, recomenda-se que sejam realizados estudos com bases de dados mais extensas.

6 REFERÊNCIAS

- BOSER, B.E.; GUYON, I.M.; VAPNIK, V.N. A training algorithm for optimal margin classifiers. In: **Proceedings of the fifth annual workshop on Computational learning theory**, p. 144-152, 1992.
- BREIMAN, L. Random forests. **Machine Learning**, vol. 45, Issue 1, pp 5–32, 2001. <<https://doi.org/10.1023/A:1010933404324>> Acesso em: 13 nov 2019.
- BURGES, C. A tutorial on support vector machines for pattern recognition. **Data mining and knowledge discovery**. v. 2, n.2, p.121-167, 1998.
- CAMMAROTA, C., & PINTO, A. Variable selection and importance in presence of high collinearity: an application to the prediction of lean body mass from multi-frequency bioelectrical impedance. **Journal of Applied Statistics**, 48(9), 1644–1658. 2021. <https://doi.org/10.1080/02664763.2020.1763930>
- CAMPOS, J. C. C.; RIBEIRO, J. O.; PAULA NETO, F.. **Inventário florestal nacional, reflorestamento: Minas Gerais**. Brasília: IBDF, 1984.
- CLUTTER J. L., Compatible Growth and Yield Models for Loblolly Pine. **Forest Science**, v. 9, n. 3, p. 354–371, September 1963.
- COSENZA, D. N.; KORHONEN, L.; MALTAMO, M.; PACKALEN, P.; STRUNK, J.; NÆSSET, E.; TOMÉ, M. Comparison of linear regression, k-nearest neighbour and random forest methods in airborne laser-scanning-based prediction of growing stock. **Forestry: An International Journal of Forest Research**, 2020. <https://doi.org/10.1093/forestry/cpaa034>
- DRUSZCZ, J. P. et al. Custos de inventário florestal com amostragem de Bitterlich (PNA) e conglomerado em cruz (CC) em plantação de *Pinus taeda* L. **Scientia Florestalis**, v. 40, n. 94, p. 231-239, 2012.
- EISFELD, R.; SANQUETTA, C.; ARCE, J.; MAESTRI, R.; WEBER, K. Modelagem do crescimento e da produção de pinus taeda l. Por meio de função probabilística. **Floresta**, v. 35, n. 2, p. 317–328, 2005. <<https://doi.org/10.5380/rf.v35i2.4619>> Acesso em: 14 nov 2019.
- FERREIRA, D. F.. **Estatística básica**. Lavras: Editora UFLA, 2005.
- GARCÍA-GUTIÉRREZ, J; MARTÍNEZ-ÁLVAREZ, F; TRONCOSO, A; RIQUELME, J. C. Acomparision of machine learning regression techniques for LiDAR-derived estimation of forest variables. **Neurocomputing**, v. 167, p. 24–31, 2015. <<https://doi.org/10.1016/j.neucom.2014.09.091>> Acesso em: 14 nov 2019.
- GROENENDIJK, P.; BONGERS, F.; ZUIDEMA, P.; A. Using tree-ring data to improve timber-yield projections for African wet tropical forest tree species. **Forest Ecology and Management**, v. 400, p. 396-407, 2017.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. Overview of supervised learning. In: **The elements of statistical learning**, Springer, New York, p. 9-41, 2009.
- HYYPPÄ, J.; HYYPPÄ, H.; LECKIE, D.; GOUGEON, F.; YU, X.; MALTAMO, M. (2008). Review of methods of small-footprint airborne laser scanning for extracting forest inventory

data in boreal forests. **International Journal of Remote Sensing**, v. 29, n. 5, p. 1339–1366. <<https://doi.org/10.1080/01431160701736489>> Acesso em: 29 mai 2019.

IBÁ. Relatório 2019. **Indústria Brasileira de Árvores - IBÁ**. <<https://iba.org/datafiles/publicacoes/relatorios/iba-relatorioanual2019.pdf>> Acesso em: 29 out 2019.

ISENBURG, M.. **LAStools - efficient LiDAR processing software** (version 141017, unlicensed), 2014. obtido em: <<http://rapidlasso.com/LAStools>> Acesso em: 21 ago 2019.

KANGAS, A.; ASTRUP, R.; BREIDENBACH, J.; FRIDMAN, J.; GOBAKKEN, T.; KORHONEN, K. T.; OLSSON, H. . Remote sensing and forest inventories in Nordic countries—roadmap for the future. **Scandinavian Journal of Forest Research**, vol. 33, n. 4, p. 397–412, 2018. <<https://doi.org/10.1080/02827581.2017.1416666>> Acesso em: 13 nov 2019.

LEITE, H.; GUIMARÃES, D.; CAMPOS, J. Descrição e emprego de um modelo para estimar múltiplos volumes de árvores. **Revista Árvore**, Viçosa, MG, v. 19, p. 1-21, nov./dez. 1995.b

LEPAGE, R. Linear regression models. In: LOVRIC, M. **International encyclopedia of statistical science**. Editora: Springer, p.1673, 2011.

MACEDO, R. D. C. **Estimativa volumétrica de povoamento clonal de Eucalyptus spp. através de laser scanner aerotransportado**. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, SP, 2009.

MACLEAN, G. A.; MARTIN, G. L. Merchantable timber volume estimation using cross-sectional photogrammetric and densitometric methods. **Canadian Journal of Forest Research**. v.14, n. 6, p. 803–810, 1984.

MACLEAN. G. A.; KRABILL, W. B. Gross-merchantable timber volume estimation using an airborne lidar system. **Canadian Journal of Forest Research**. v. 12, p. 7–18, 1986.

MCGAUGHEY, R. **FUSION / LDV: Software for LIDAR Data Analysis and Visualization**. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, University of Washington, Seattle, WA, USA. 2018.

MCKINNEY, W. Data Structures for Statistical Computing in Python, **Proceedings of the 9th Python in Science Conference**, 51-56, 2010.

Murphy, K. P. (2012). **Machine learning - a probabilistic perspective**. Cambridge, MA: MIT Press. <https://doi.org/10.1038/217994a0>

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

NÆSSET, E. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. **Remote Sensing of Environment**, v. 80, n. 1, p. 88–99, 2002.

NÆSSET, E.; GOBAKKEN, T.; HOLMGREN, J.; HYYPPÄ, H.; HYYPPÄ, J.; MALTAMO, M.; SÖDERMAN, U. (2004). Laser scanning of forest resources: The nordic experience. **Scandinavian Journal of Forest Research**, v. 19, n. 6, p. 482–499, 2004. <<https://doi.org/10.1080/02827580410019553>> Acesso em: 13 nov 2019.

PACKALÉN, P.; MEHTÄTALO, L.; MALTAMO, M. ALS-based Estimation of Plot Volume and Site Index in a *Eucalyptus* Plantation with a Nonlinear Mixed-Effect Model that Accounts for the Clone Effect. **Annals of Forest Science**. V. 68, n. 6, p. 1085–1092, 2011.

PAN, J.;X.; FANG, K.;T. **Growth curve models and statistical diagnostics**. Springer, Amsterdam, p.387, 2002.

PEDREGOSA F.; VAROQUAUX G.; GRAMFORT A.; MICHEL V.; THIRION B.; GRISEL O.; BLONDEL M.; PRETTENHOFER P.; WEISS R.; DUBOURG V.; VANDERPLAS J.; PASSOS A.; COURNAPEAU D.; BRUCHER M.; PERROT M.; DUCHESNAY E. Scikit-learn: Machine Learning in Python, **Journal of Machine Learning Research**, 12, 2825-2830, 2011.

RAGHAVENDRA, S., DEKA, P. C. Support vector machine applications in the field of hydrology: A review. **Applied Soft Computing Journal**, 19, 372–386, 2014. <https://doi.org/10.1016/j.asoc.2014.02.002>

REIS, A. A.; CARVALHO, M. C.; MELLO, J. M.; GOMIDE, L. R.; FERRAZ FILHO, A. C.; ACERBI JUNIOR, F. W. Spatial prediction of basal area and volume in Eucalyptus stands using Landsat TM data: an assessment of prediction methods. **New Zealand Journal of Forestry Science** , v. 48, p. 1-17, 2018.

REMPEL, R.C.; PARKER A. K.. An information note on an airborne laser terrain profiler for micro-relief studies. In: **Proceedings of the 3rd symposium on remote sensing of environment**, University of Michigan Institute of, Science and Technology, p. 321–337, 1964.

SCOLFORO, J. R. S.; MAESTRI, R.; FERRAZ FILHO, A. C.; MELLO, J. M.; OLIVEIRA, A. D.; ASSIS, A. L. Dominant Height Model for Site Classification of *Eucalyptus grandis* Incorporating Climatic Variables. **International Journal of Forestry Research**, v. 2013, p. 7 2013. <<https://doi.org/10.1155/2013/139236>> Acesso em: 13 nov 2019.

SILVA, C. A.; KLAUBERG, C.; CARVALHO, S. P. C.; HUDAK, A. T.; RODRIGUEZ, L. Mapping aboveground carbon stocks using LiDAR data in *Eucalyptus spp.* plantations in the state of São Paulo, Brazil. **Scientia Forestalis**, v. 42, p. 591–604, 2014.

SILVA, C. A.; KLAUBERG, C.; HUDAK, A. T.; VIERLING, L. A.; LIESENBERG, V.; CARVALHO, S. P. C.; RODRIGUEZ, L. C. E. A principal component approach for predicting the stem volume in *Eucalyptus* plantations in Brazil using airborne LiDAR data. **Forestry**, v. 89, n. 4, p. 422–433, 2016.

SILVA, C. A.; HUDAK, A. T.; KLAUBERG, C.; VIERLING, L. A.; GONZALEZ-BENECKE, C.; CARVALHO, S. P. C.; CARDIL, A. Combined effect of pulse density and grid cell size on predicting and mapping aboveground carbon in fast-growing *Eucalyptus* forest plantation using airborne LiDAR data. **Carbon Balance and Management**, v. 12, n. 1, p. 1–16, 2017a. <<https://doi.org/10.1186/s13021-017-0081-1>> Acesso em: 13 nov 2019.

SILVA, C. A.; KLAUBERG, C.; HENTZ, Â. M. K.; CARVALHO, S. P. C.; CORTE, A. P. D.. Predição da biomassa aérea em plantações de Pinus taeda L. por meio de dados LiDAR aerotransportado. **Scientia Forestalis**, v. 45, n. 115, 2017b. <<https://doi.org/10.18671/scifor.v45n115.10>> Acesso em: 13 nov 2019.

SILVA, V. S., SILVA, C. A., MOHAN, M., CARDIL, A., REX, F. E., LOUREIRO, G. H., KLAUBERG, C. Combined Impact of sample size and modeling approaches for predicting stem volume in *Eucalyptus* spp. forest plantations using field and LiDAR data. **Remote Sensing**, 2020.

SOARES, C. P. B.; PAULA NETO, F.; SOUZA, A. L.. **Dendrometria e inventário florestal**. 1. ed. Viçosa, MG: Ed. UFV, 2006. 276 p.

SOUZA, G. S. A. DE, SOARES, V. P., LEITE, H. G., GLERIANI, J. M., DO AMARAL, C. H., FERRAZ, A. S., SILVA, S. Multi-sensor prediction of *Eucalyptus* stand volume: A support vector approach. **ISPRS Journal of Photogrammetry and Remote Sensing**, 135–146, 2019.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete sample). **Bio-metrika**, Great Britain, v. 52, n. 3, p. 591-611, 1965.

SULLIVAN, A. D.; CLUTTER, J. L. A Simultaneous Growth and Yield Model for Loblolly Pine, **Forest Science**, v. 18, n. 1, p. 76–86, 1972.

TYRALIS, H., PAPACHARALAMPOUS, G., LANGOUSIS, A. (2019). A brief review of random forests for water scientists and practitioners and their recent history. **Water**, 11, 910. <https://doi.org/10.3390/w11050910>

VAUHKONEN, J.; MEHTÄTALO, L.; PACKALÉN, P. Combining tree height samples produced by airborne laser scanning and stand management records to estimate plot volume in *Eucalyptus* plantations. **Canadian Journal of Forest Research**, v. 41, n. 8, p. 1649–1658, 2011. <<https://doi.org/10.1139/x11-083>> Acesso em: 13 nov 2019.

VAUHKONEN, J.; MALTAMO, M.; MCROBERTS, R. E.; NÆSSET, E. Introduction to Forestry Applications of Airborne Laser Scanning. In: MALTAMO M.; NÆSSET E.; VAUHKONEN J. Forestry Applications of Airborne Laser Scanning: concepts and case studies. **Managing Forest Ecosystems**, vol 27. Springer, Dordrecht, Cap. 1. p. 1-16, 2014.

VAPNIK, V. The Nature of Statistical Learning Theory. **Springer**, New York, 1995.

WU, C.; SHEN, H.; SHEN, A.; DENG, J.; GAN, M.; ZHU, J.; XU, H.; WANG, K.. Comparison of machine-learning methods for above-ground biomass estimation based on Landsat imagery. **Journal of Applied Remote Sensing**, v.10, n. 3, 2016. <<https://doi.org/10.1117/1.JRS.10.035010>> Acesso em: 13 nov 2019.

YU, X.; HYYPPÄ, J.; HOLOPAINEN, M.; VASTARANTA, M.. Comparison of area-based and individual tree-based methods for predicting plot-level forest attributes. **Remote Sensing**, v. 2, n.6, p. 1481–1495, 2010. <<https://doi.org/10.3390/rs2061481>> Acesso em: 13 nov 2019.

ZANDONÁ, D. F. **Potencial do uso de dados laser scanner aerotransportado para estimativa de variáveis dendrométricas**. 92 p. Dissertação (Mestrado em Engenharia Florestal) – Universidade Federal do Paraná, Curitiba, 2006.

ZONETE, M. F. **Análise do uso da tecnologia laser aerotransportado para inventários florestais em plantios clonais de *Eucalyptus sp* no sul da Bahia**. 2009. 95 p. Dissertação (Mestrado em Recursos Florestais) – Escola Superior de Agricultura Luiz de Queiroz, 2009.

ZONETE, M. F.; RODRIGUEZ, L. C. E.; PACKALÉN, P. Estimação de parâmetros biométricos de plantios clonais de eucalipto no sul da Bahia: uma aplicação da tecnologia laser aerotransportada. **Scientia Florestalis**, Piracicaba, v. 38, n. 86, p. 225-235, 2010.