

RENATA PAULO MACEDO

**MODELAGEM DE ATRIBUTOS BIOFÍSICOS DE FLORESTA ESTACIONAL
SEMIDECIDUAL UTILIZANDO APRENDIZAGEM DE MÁQUINA E DADOS
OBTIDOS COM LIDAR AEROTRANSPORTADO**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência Florestal, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2019

**Ficha catalográfica preparada pela Biblioteca Central da
Universidade Federal de Viçosa - Campus Viçosa**

T

M141m
2019
Macedo, Renata Paulo, 1991-
Modelagem de atributos biofísicos de floresta estacional
semidecidual utilizando aprendizagem de máquina e dados obtidos com
LiDAR aerotransportado / Renata Paulo Macedo. - Viçosa, MG, 2019.
vi, 58f. : il. (algumas color.) ; 29 cm.

Orientador: Cibele Hummel do Amaral.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: 50 - 58.

1. Mata Atlântica. 2. Sensoriamento remoto. 3. Algoritmos. I.
Universidade Federal de Viçosa. Departamento de Engenharia
Florestal. Programa de Pós-Graduação em Ciência Florestal. II. Título.

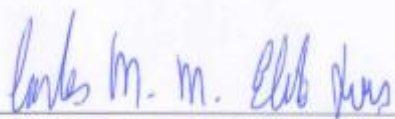
CDO adapt. CDD 634.99221

RENATA PAULO MACEDO

MODELAGEM DE ATRIBUTOS BIOFÍSICOS DE FLORESTA ESTACIONAL SEMIDECIDUAL UTILIZANDO APRENDIZAGEM DE MÁQUINA E DADOS OBTIDOS COM LIDAR AEROTRANSPORTADO


Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência Florestal, para obtenção do título de *Magister Scientiae*.

APROVADA: 26 de fevereiro de 2019.



Carlos Moreira Miquelino Eleto Torres

(Coorientador)



Heho Garcia Leite



Elpídio Inácio Fernandes Filho



José Marinaldo Gleriani



Cibele Hummel do Amaral

(Orientadora)

AGRADECIMENTOS

A professora Cibele, pela orientação e contribuição nessa dissertação.

Aos coorientadores, professor Carlos Pedro Boechat Soares e professor Carlos Moreira Miquelino Eleto Torres.

Aos professores da banca examinadora Helio Garcia Leite, Elpídio Inácio Fernandes Filho, José Marinaldo Gleriani.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa.

A empresa CENIBRA S.A. pelo oferecimento dos dados, em especial aos seus funcionários, o mestre Jacinto Moreira de Lana, aos engenheiros florestais Walter da Silva Costa e Gustavo Bretas Lage, ao funcionário terceirizado Maicon Bruno Nascimento e a estagiária Brenda Barbalho.

Ao professor Antônio Santana Ferraz e ao Laboratório de Engenharia de Agrimensura em especial ao funcionário técnico em cartografia Tiago Oliveira Lopes.

Ao laboratório SIGMA do departamento de Engenharia Florestal, UFV, em especial para o engenheiro florestal Rodrigo Vieira Leite, ao mestre Duberli Geomar Elera Gonzáles, ao mestre Pedro Henrique Santos Mota, ao mestre Lucas Arthur Almeida Teles e a todos os outros que contribuíram de alguma forma.

Aos funcionários do departamento de Engenharia Florestal – UFV, pelo apoio e companheirismo durante essa jornada.

À minha família pelo amor, confiança e dedicação.

Aos meus amigos de todos os cantos do Brasil, pela amizade, força e conselhos.

SUMÁRIO

RESUMO	v
ABSTRACT	vi
1. INTRODUÇÃO.....	1
2. OBJETIVOS.....	2
2.1. Objetivo Geral.....	2
2.2. Objetivos Específicos.....	2
3. REVISÃO DE LITERATURA.....	3
3.1. Floresta Estacional Semidecidual	3
3.2. Métodos quantitativos e qualitativos dos parâmetros florestais	4
3.3. Modelos para estimativas dos parâmetros florestais	6
3.4. LiDAR aplicado às estimativas de atributos e mapeamento florestal.....	7
4. MATERIAL E MÉTODOS.....	10
4.1. Área de estudo	10
4.2. Dados de campo e análise descritiva.....	11
4.3. Georreferenciamento das parcelas	12
4.4. Estimativa de campo dos atributos biofísicos.....	13
4.5. Dados de sensoriamento remoto	14
4.6. Processamento dos dados LiDAR e análise descritiva	14
4.7. Modelagem dos atributos biofísicos	18
4.7.1. Regressão.....	18
4.7.1.1. Seleção de variáveis	18
4.7.1.2. Ajuste do modelo	19
4.7.2. Aprendizado de máquina.....	20
4.7.2.1. Seleção de variáveis	20
4.8. Mapeamento dos atributos biofísicos.....	20
4.9. Validação dos mapeamentos.....	21
5. RESULTADOS	21
5.1. Análise descritiva.....	21
5.1.1. Dados de campo	21
5.1.2. Dados de sensoriamento remoto	22
5.2. Estimativa de campo dos atributos biofísicos.....	26
5.3. Modelagem dos atributos biofísicos	27
5.3.1. Seleção de variáveis e treinamento	27
5.3.1.1. Regressão.....	27
5.3.1.2. <i>Random Forest</i>	30
5.3.1.3. Validação externa.....	32
5.4. Mapeamento dos parâmetros florestais	33
5.5. Validação dos mapeamentos.....	38
6. DISCUSSÃO	45

7. CONCLUSÕES	50
8. REFERÊNCIAS BIBLIOGRÁFICAS.....	50

RESUMO

MACEDO, Renata Paulo, M.Sc., Universidade Federal de Viçosa, fevereiro de 2019. **Modelagem de atributos biofísicos de Floresta Estacional Semidecidual utilizando aprendizagem de máquina e dados obtidos com LiDAR aerotransportado.** Orientadora: Cibele Hummel do Amaral. Coorientador: Carlos Moreira Miquelino Eleto Torres.

A estimativa dos atributos biofísicos de florestas inequiduais tornam aplicáveis para o manejo e para a conservação do ambiente atribuídos a sua existência. A utilização de ferramentas de Inteligência Artificial e de Sistema de Informação Geográfica têm tornado o inventário florestal mais ágil, obtendo a mesma ou maior exatidão das estimativas. Dessa forma, esse trabalho tem como objetivo estimar e mapear a partir de dados *Light Detection And Ranging* (LiDAR), o volume, a biomassa e o carbono acima do solo em fragmentos de Floresta Estacional Semidecidual, localizados a leste de Minas Gerais, Brasil. Bem como, comparar o desempenho de métodos de regressão e método de aprendizado por máquinas. Os dados são provenientes de inventário florestal contínuo da empresa CENIBRA S.A. distribuídos em 29 parcelas. Os dados LiDAR foram obtidos via plataforma aérea na região analisada. Para a regressão o ajuste da equação foi o exponencial e para o aprendizado por máquina o modelo *Random Forest*. Destinou 70% das parcelas para o treinamento e 30% para validação externa. Na equação exponencial adotou-se apenas uma métrica como variável independente, enquanto que para o *Random Forest* foi utilizado o método *Recursive Feature Elimination* para seleção das variáveis mais importantes ao modelo. Os valores dos coeficientes de correlação ajustado (r_{YY}) para as três modelagens (volume, a biomassa e o carbono acima do solo) foram superiores a 0,91 para ambos os métodos. No mapeamento dos atributos, os modelos tiveram suas estimativas médias dentro do Intervalo de Confiança ($\alpha = 0,05$). Porém, o modelo *Random Forest* apresentou valores elevados de RMSE. O teste F de Graybill apresentou semelhança entre os métodos sendo realizado em todas as células da estimativa do mapeamento. Desta forma, conclui-se que é possível estimar e mapear com exatidão o volume, a biomassa acima do solo e o carbono acima do solo fragmentos da Floresta Estacional Semidecidual por meio de ambos os métodos.

ABSTRACT

MACEDO, Renata Paulo, M.Sc. Universidade Federal de Viçosa, February, 2019. **Modeling of biophysical attributes of Semideciduous Seasonal Forest using machine learning and airborne LiDAR data.** Advisor: Cibele Hummel do Amaral. Co-Advisor: Carlos Moreira Miquelino Eleto Torres.

The estimation of biophysical attributes of unequal forests makes them applicable to the management and conservation of the environment attributed to their existence. The use of Artificial Intelligence and Geographic Information System tools have made the forest inventory more agile, obtaining the same or greater accuracy of the estimates. Thus, this work aims to estimate and quantify the volume, biomass and carbon above the soil in fragments of the Semideciduous Seasonal Forest located in the east of Minas Gerais, Brazil, from Light Detection And Ranging (LiDAR) data. As well as, compare the performance of regression methods and method of learning by machines. The data come from continuous forest inventory of the company CENIBRA S.A. distributed in 29 plots. LiDAR data were obtained by aerial platform in the analyzed region. For the regression the adjustment of the equation was exponential and for the learning by machine the model Random Forest. It allocated 70% of the plots for training and 30% for external validation. In the exponential equation only one metric was adopted as independent variable, whereas for Random Forest the Recursive Feature Elimination method was used to select the most important variables to the model. The values of the adjusted correlation coefficients ($r_{\hat{Y}Y}$) for the three models (volume, biomass and carbon above ground) were higher than 0.91 for both methods. In the mapping of attributes, the models had their average estimates within the Confidence Interval ($\alpha = 0.05$). However, the Random Forest model presented high RMSE values. The Graybill F test showed similarity between the methods being performed on all cells of the mapping estimate. In this way, it is concluded that it is possible to estimate and accurately map the volume, the above-ground biomass and the carbon above the soil fragments of the Semideciduous Seasonal Forest by means of both methods.

1. INTRODUÇÃO

O bioma Mata Atlântica é considerado um dos *hotspot* mundiais de biodiversidade, com elevada incidência de espécies endêmicas, e ameaçado pelo desmatamento e pela conversão de áreas para outras finalidades (MYERS, 2000). O monitoramento contínuo dessa região tem despertado interesse de muitos pesquisadores, a fim de quantificar e conservar esses recursos florestais (LAURIN et al., 2016).

Para que haja o monitoramento contínuo das florestas naturais é necessário a quantificação *in situ* de algumas das principais características da estrutura florestal (SOARES, 2011). Para essa quantificação é necessário coletar o número, diâmetro e altura dos indivíduos, além da densidade populacional por área, para o cálculo da estimativa de volume, biomassa e carbono.

Estimativas dos parâmetros biofísicos se baseiam em dados de inventário florestal contínuo. O uso de equações que convertem uma variável de difícil medição, geralmente a variável resposta, em função de outras variáveis de fácil acesso na floresta segue o princípio da alometria (CHAVE et al., 2005). Essas equações alométricas são a base para as estimativas em campo.

Devido a todo dispêndio financeiro e de tempo para obtenção das estimativas desses parâmetros de forma convencional, novas alternativas de técnicas estão surgindo a fim de otimizar e aprimorar os processos (VIERLING et al., 2011; HUDAK et al., 2012). A utilização de dados provenientes do sensoriamento remoto como forma não destrutiva e com boa precisão nos resultados têm se tornado alternativa para estimativas dos atributos biofísicos (LEKSKY et al., 2002; NÆSSET et al., 2008; JENSEN et al., 2009; VIERLING et al., 2011; HUDAK et al., 2012).

Dentre todas as ferramentas de estimativa através do sensoriamento remoto, o perfilamento a laser (*Light Detection And Ranging* - LiDAR), têm mostrado grande potencial para gerar essas estimativas. Além disso, possibilita a obtenção de informações tridimensionais (3D) do alvo de interesse (BABCOCK et al., 2016) que está sendo interceptado. Com essa ferramenta é possível recuperar propriedades estruturais do dossel florestal, bem como distinguir diferentes tipos de espécies e formações florestais (FIGUEIREDO et al., 2016; GORGENS et al., 2016; LAURIN et al., 2016; LUO et al.; 2017; GARCÍA et al., 2018).

A estimativa dos atributos florestais, a nível de povoamento, tem facilitado as informações quando trabalhados com dados LiDAR (SAATCHI et al., 2011;

LEITOLD et al., 2015; BECKNELL et al., 2018). Gerando dessa forma, a caracterização de florestas (GORGENS et al., 2016), podendo ser estimados via regressão – com o ajuste de equações (CAO et al., 2016; PALACE et al., 2015; BABCOCK et al., 2016), como também com a utilização de algoritmos do aprendizado por máquina (LAURIN et al., 2016; LONGO et al., 2016) como é observado em trabalhos recentes a alta precisão das estimativas.

No aprendizado de máquina (*Machine Learning*), as estimativas dos atributos florestais têm facilitado os ajustes, sendo outra técnica que tem sido inserida nos cálculos. Através delas, os cálculos tornam tão ou mais exatos quanto aos modelos convencionais aliada aos dados extraídos em campo, podendo definir padrões elevados dos atributos (FENG et al., 2017; LU et al., 2016; LI et al., 2014) podendo ser empregadas, também, a dados de sensoriamento remoto (MONTAÑO, 2016; FENG et al., 2017).

Raros são os trabalhos na literatura desenvolvidos em ambiente de Mata Atlântica, principalmente em suas florestas interioranas, como as Florestas Estacionais Semidecíduais. Desta forma, o uso de novas tecnologias tem se mostrado promissor para estimativas (e extrapolação para amplas áreas) de parâmetros florestais, como volume, biomassa e carbono, nesses fragmentos florestais.

2. OBJETIVOS

2.1. Objetivo Geral

Estimar os atributos biofísicos florestais de fragmentos da Floresta Estacional Semidecidual utilizando dados Light Detection And Ranging obtidos de plataforma aérea.

2.2. Objetivos Específicos

- I. Modelar os atributos biofísicos florestais com o uso de regressão;
- II. Modelar os atributos biofísicos florestais utilizando aprendizado por máquinas (*Random Forest*);
- III. Mapear e validar os atributos biofísicos pelos métodos supracitados.

3. REVISÃO DE LITERATURA

3.1. Floresta Estacional Semidecidual

As florestas tropicais perderam grande área de sua ocupação original para urbanização e atividades agrícola (SILVEIRA et al., 2008). Com a conversão de áreas, a variação do clima global e as consequências advindas dessas mudanças têm despertado preocupação e alerta no cenário de desmatamentos contínuos, como sendo uma das principais causas para essas alterações na paisagem (NEVES et al., 2017). O conjunto dos elementos florestais possuem papel importante no clima e no equilíbrio do ciclo global do carbono. As espécies florestais em estágios iniciais de crescimento apresentam a capacidade de fixar o dióxido de carbono (CO₂) em uma taxa maior do que a sua emissão para a atmosfera, tornando o saldo positivo de compensação (SILVA, 2013). Essa fixação corresponde ao incremento de biomassa em árvores vivas, decorrentes do processo de fotossíntese, e a emissão representa a liberação de carbono, por respiração e decomposição da biomassa oriunda da morte natural de indivíduos vegetais ou pela combustão deste material (YU, 2004; SILVA, 2013).

O bioma Mata Atlântica, que inicialmente correspondia a 15% de todo o território brasileiro, é representado por uma vegetação contígua não homogênea possuindo características peculiares. Suas florestas geralmente são densas com presença de árvores de médio a grande porte (FUNDAÇÃO SOS MATA ATLÂNTICA, 2014). A distribuição geográfica desse bioma abrange 17 estados brasileiros, favorecendo a elevada diversidade de espécies da flora e da fauna em função da variabilidade altimétrica, longitudinal e latitudinal do país (NEVES et al., 2017).

As formações florestais desse bioma são denominadas de: Floresta Ombrófila Densa, Floresta Ombrófila Aberta, Floresta Ombrófila Mista, Floresta Estacional Decidual e Floresta Estacional Semidecidual (IBGE, 1992). As Florestas Estacionais Semidecíduais ocupam ambientes transitórios entre zona úmida costeira e o ambiente semiárido (IBGE, 1992). Uma característica importante de sua fisionomia é a perda de folhas no estrato arbóreo durante o período seco, sendo que, em épocas de chuva, pode ser confundida com a fisionomia de Floresta Ombrófila Densa (IBGE, 1992).

3.2. Métodos quantitativos e qualitativos dos parâmetros florestais

Os inventários realizados em florestas inequiâneas e equiâneas fornecem informações necessárias para a realização do manejo florestal e para políticas governamentais de monitoramento que se baseiam em diferentes escalas para obtenção das estimativas (PICARD et al., 2018). A Organização das Nações Unidas para Alimentação e Agricultura (FAO, 2017) tem dado ênfase à necessidade de monitoramento florestal em níveis nacionais, estabelecendo políticas mitigadoras e de adaptação às mudanças climáticas ao nível de biodiversidade e de manutenção do ecossistema.

A mensuração dos indivíduos florestais em campo é a principal fonte para quantificar a biomassa e o carbono estocado dos indivíduos arbóreos. Esses cálculos são importantes para os mecanismos mitigadores (CHEN et al., 2011). Estudos demonstram que com o aumento exponencial dos gases de efeito estufa, entre eles o dióxido de carbono (CO₂), a biomassa florestal total também é aumentada, tornando necessário a implementação de políticas de monitoramento desses componentes florestais (YAO et al., 2018). Saatchi et al. (2011) estudando a concentração do carbono nos diferentes continentes do planeta, observaram que a América Latina possui uma extensa área de floresta e nela compreende 49% de toda a biomassa global estudada no trabalho.

Os métodos para a obtenção de volume, de biomassa e de carbono podem ser destrutivos ou não destrutivos. Nos métodos destrutivos ou diretos, os indivíduos são selecionados sistematicamente para que possam ser abatidos, segmentados e pesados. Depois de passarem por esse processo, seus componentes florestais são separados, selecionados e destinados ao laboratório para continuação das etapas do processo (SOARES et al., 2011). Nos métodos indiretos, também denominados de não-destrutivos, a quantificação se baseia estimativas dos dados provenientes do inventário florestal sem o abate dos indivíduos medidos (WENDLING, 1998; WATZLAWICK, 2003).

O método destrutivo tem como desvantagem, o dispêndio de esforço, de tempo e do elevado custo para execução, o que por muitas vezes, torna inviável quando se trabalha em áreas extensas (SOARES et al., 2011). Já os métodos não-destrutivos utilizam modelos de regressão, lineares ou não-lineares, cujas variáveis dependentes são diretamente mensuráveis na árvore-amostra, enquanto que as variáveis

independentes podem ser volume, biomassa ou carbono. Dessa forma, a relação entre tamanhos de um organismo é expressa por um modelo, denominado de modelo alométrico (HUXLEY e TEISSER, 1936).

Muitas relações dendrométricas são estudadas e passíveis de modelagem, tais como as relações entre altura e diâmetro. E através dessas relações diversos problemas florestais são resolvidos utilizando relações matemáticas, que possibilitam obter estimativas através de equações de regressão (SCHNEIDER, 1997).

Tratando de estimativa dos atributos florestais, Ketterings et al. (2001), argumentam que a estimativa de biomassa acima do solo é imprescindível aos estudos do balanço global de carbono.

Assim o teor de carbono alocado na biomassa em diversas partes da árvore constitui uma proporção de aproximadamente 50% da composição da matéria seca da biomassa (SAATCHI et al., 2011; SOARES et al., 2011). Na estimativa de carbono acima do solo, Cerri et al. (2001) concluíram que o carbono estocado em ecossistemas tropicais (solo e vegetação) representa cerca de 20 a 25% do carbono terrestre mundial. Corroborando com o estudo que envolveu o teor de carbono na manta orgânica de uma floresta tropical úmida tendo valor igual a 39,3% da biomassa viva presente na área (HIGUCHI e CARVALHO JR.; 1994). Para as florestas tropicais, Amaro (2010) encontrou valores próximos a 48% de carbono no total de biomassa seca em um fragmento de Floresta Estacional Semidecidual Montana, localizado no município de Viçosa, Minas Gerais.

Em áreas florestais extensas, a mensuração dos parâmetros é feita por amostragem dos indivíduos, uma vez que o censo amostral torna inviável em campo. Portanto, na estimativa desses parâmetros, algumas variáveis de campo que são envolvidas nas equações matemáticas tornam importantes e que devem ser analisadas criteriosamente adotando as peculiaridades de cada região, evitando a sub ou a superestimação do valor da variável preditora (SILVA, 2013).

A consistência da base de dados amostradas é importante para a estimativa e a extrapolação dos parâmetros florestais. Através dela, visualiza a variabilidade entre os indivíduos florestais nos diferentes projetos abordados. Essa visualização torna a análise criteriosa para a compreensão geral dos dados.

3.3. Modelos para estimativas dos parâmetros florestais

A estimação de parâmetros florestais empregando equações alométricas fornece estimativas para os gestores florestais (XIANG et al., 2016). O surgimento de novas técnicas para a estimar os parâmetros, como o aprendizado por máquina (*Machine Learning*) pode também gerar resultados tão ou mais exatos do que as relações alométricas. Essa técnica apresenta redução no tempo de coleta e no processamento dos dados advindos do campo (MONTAÑO, 2016).

Seus algoritmos utilizam dos valores destinados ao treinamento da função alvo, que ao induzir uma determinada função que esteja de acordo com os exemplos da função resposta, essas sejam exemplos não vistos durante o treinamento na generalização. O que difere da estatística clássica, pois nela o método de aprender a aproximar uma função contínua dos dados respostas é chamado de regressão (HAYKIN, 1999).

O *Random Forest* (RF) é um algoritmo de aprendizagem supervisionado e possibilita usar tarefas tanto para classificação quanto para regressão (CART) (HUDAK et al., 2012). É constituído por uma série de decisões baseadas em regras binárias que determinam a variável de entrada de acordo com a relação da variável preditora (WALTON, 2008). Ele tem a capacidade de implementar modelos aleatoriamente enquanto cria as árvores de decisões e, ao invés de procurar pela melhor característica isoladamente, busca em subconjuntos aleatórios as melhores características para a resposta desejada (MONTAÑO, 2016). Essa característica do algoritmo possibilita uma maior diversidade dos modelos, e conseqüentemente, a geração de modelos melhores (GHOSH et al., 2018).

O diferencial do RF é a possibilidade de medir e ranquear a importância das variáveis presentes no modelo, reduzindo os erros provenientes de aproximação dos dados. O algoritmo realiza o cálculo do valor de importância para cada métrica dentro do conjunto, automaticamente, após realizar o treinamento e normalizar os resultados, obtendo a soma de todos os valores de importâncias das métricas sendo iguais a 1 (GLEASON et al., 2012). As ramificações oriundas das árvores são criadas quando uma regra for menor que a outra, produzindo novas regras mais precisas. A árvore finaliza suas ramificações quando obtém o erro mínimo em relação aos dados de entrada (WALTON, 2008).

Os modelos de *Random Forest* tendem a reduzir os viés e os sobreajustes dos dados apresentados à árvore de decisão e, em alguns casos, podem resultar em maior precisão das estimativas em relação as técnicas de regressão para estimar volume, biomassa e carbono (GLEASON et al., 2012; POWELL et al., 2010).

3.4. LiDAR aplicado às estimativas de atributos e mapeamento florestal

Uma alternativa às técnicas tradicionais para a estimativa de parâmetros florestais são as provenientes do sensoriamento remoto que tem se tornado versáteis e bastante aplicadas aos estudos de quantificação de parâmetros biofísicos como volume, biomassa e carbono (NÆSSET, 2007; SILVEIRA et al., 2008; MALTAMO et al., 2012).

O sensoriamento remoto permite a obtenção de informações alvos coletados por ele, através de equipamentos que não estejam, de fato, em contato físico com eles (LILLESAND, KIEFER & CHIPMAN, 2015). Seus produtos podem se dividir em duas categorias, sendo os imageadores (obtem imagem como produto) e os não imageadores (não fornecem imagem como produto). Podem variar de acordo com os tipos de sensores, sendo os passivos (dependem de iluminação externa para captar informações do alvo) e os ativos (produzem sua própria iluminação para captar as informações dos alvos). O suporte da plataforma é outra categoria dos produtos do sensoriamento remoto, classificando-os como terrestres, aéreos ou orbitais (GIONGO et al., 2010).

Light Detection And Ranging (LiDAR), traduzido para o português “perfilamento a laser” é uma técnica de sensoriamento remoto que tem como características: ser um sensor ativo, não imageador (produz nuvem de pontos como produto) e podem estar embarcados em diferentes plataformas (LEFSKY et al., 2002; SILVA, 2013).

O pulso laser é emitido pelo sensor e, após interceptar a superfície do objeto alvo, retorna para o equipamento para ser armazenado a informação obtida no intercepto. A diferença entre o tempo de emissão e o tempo de retorno do sinal, considerando a velocidade da luz, podemos calcular a distância percorrida pelo pulso de laser, obtendo a informação do eixo z (altura) em uma dimensão tridimensional, uma vez que as informações de posicionamento (x e y) são obtidas pelo GPS acoplado

ao sensor (HUDAK et al., 2012; HUMAGAIN et al. 2018). Essas feições da superfície do terreno e das estruturas verticais facilitam e auxiliam detalhadamente nos dados.

Essa tecnologia tornou possível reproduzir, de forma acurada, o Modelo Digital de Terreno (MDT - imagens com informação da elevação do terreno livre de objetos) e o Modelo Digital de Superfície (MDS – imagens com informação da elevação da superfície, incluindo os objetos sobre o terreno). O Modelo Digital de Altura (MDA – modelos que descrevem a altura dos objetos, tendo os pontos da nuvem referentes ao solo normalizados a zero) (GORGENS et al., 2014), ambos modelos baseados na densidade de pontos por metro quadrado.

O MDT é a representação do terreno obtendo uma forma regular ou irregular. Dois pontos importantes são a precisão com que a elevação foi captada pelos retornos e o grau de detalhamento do terreno que está representado pela superfície criada. No MDS é a representação de todos os alvos que estão sobre o MDT. A qualidade do MDS está relacionada com a exatidão e nível de detalhamento da nuvem de pontos. O MDA é obtido após processamento de normalização da nuvem de pontos, obtendo a altura dos alvos sem interferência do relevo (GORGENS et al., 2014).

Além desses modelos citados que são obtidos pela nuvem de pontos, obtém-se informações estatísticas dos alvos interceptados. Essas medidas são relacionadas a contagem (número de retornos totais e por tipo), a elevação referente à altura (máxima, mínima, média, mediana, moda, e seus percentis proporcionais), elevação quanto a dispersão (assimetria e curtose, coeficientes de variação, desvio e variância, distância interquartil da elevação) e a elevação quanto a proporção (número de retornos em relação a algum outro parâmetro calculado, porcentagem referente ao todo) (GORGENS et al., 2014). Essas informações estatísticas obtidas dessas categorias (chamadas de métricas LiDAR) podem ser correlacionadas com medidas advindas do campo, pelo inventário florestal, para a estimação precisa dos parâmetros biométricos da floresta, tais como volume, biomassa e estoque de carbono (ZOLKOS et al., 2013; BABCOCK et al., 2016; CAO et al., 2016; LAURIN et al., 2016; LUO et al., 2017; MILENKOVIC et al., 2017).

A escolha de variáveis explicativas, derivadas das métricas LiDAR, é importante para garantir a modelagem dos parâmetros florestais mais precisa. Em estudos semelhantes ao proposto por Gorgens et al. (2014), foram selecionadas uma lista de possíveis candidatas extraída do LiDAR para utilizar na modelagem. As

variáveis que não explicavam bem o modelo foram eliminadas sistematicamente para não influenciar no ajuste da regressão (GLEASON et al., 2012).

A técnica LiDAR tem grande potencial para fornecer estimativas tratando de extensas áreas. Sua capacidade de coletar diversas informações com alta precisão em pouco tempo e a um baixo custo tem despertado interesse dos pesquisadores quando comparado com métodos convencionais (XU et al., 2018). Nessa tecnologia existem duas principais abordagens para a obtenção de informações florestais, a que trabalha com a distribuição da altura do dossel e outra abordagem que aborda a segmentação de árvore individual (GLEASON et al., 2012).

No Brasil, a utilização dessa técnica com a distribuição da altura do dossel dos dados LiDAR começou a ser implantada através de trabalhos que avaliaram potencial para estimativas de parâmetros biométricos em povoamentos florestais e na floresta Amazônia. Esses trabalhos tiveram como foco a estocagem de carbono como medida mitigadora para o aquecimento e mudanças climáticas (SAATCHI et al., 2011; LONGO et al., 2016), também para outros biomas brasileiros, como a Mata Atlântica (GORGENS et al., 2016; LEITOLD et al., 2015).

A combinação de dados de inventário florestal com técnicas de sensoriamento remoto e modelos estatísticos permite prever e mapear a biomassa acima do solo (BABCOCK et al., 2016). De acordo com CAO et al. (2016), as métricas da variação da altura do dossel e algumas métricas de percentil calculadas tendem a ser sensíveis à mudança de biomassa, enquanto que as métricas de altura máxima e densidade do dossel tem pouco efeito na mudança da biomassa.

Gleason et al. (2012) por sua vez, avaliaram as diversas abordagens de aprendizado por máquina, como *Random Forest*, *Support Vector Regression* (SVR), árvores de regressão Cubist para a estimativa de biomassa florestal usando dados LiDAR, obtendo resultados satisfatórios para a análise.

Tem se tornado frequentes os mapeamentos dos parâmetros florestais, em especial a biomassa e o estoque de carbono, para tomada de medidas mitigadoras frente ao aquecimento global e as mudanças climáticas (YAO et al., 2018). Saatchi et al. (2011) criaram mapas de referência com o estoque de carbono para três continentes mundiais, aliando dados de campo, imagens hiperespectrais e dados de LiDAR orbital para o gerenciamento e monitoramento do desmatamento e da degradação (REDD+).

García et al. (2018) integraram imagens multiespectrais e dados de satélite Radar juntamente com dados LiDAR para predizer e mapear as variações em escala da estrutura florestal em dois biomas da América do Norte, obtendo resultados de coeficiente de determinação relevantes ao fundir dados LiDAR com dados de Radar

4. MATERIAL E MÉTODOS

4.1. Área de estudo

O trabalho foi realizado em dois fragmentos de Floresta Estacional Semidecidual, denominados projetos Lagoa do Piau e São José, pertencentes a empresa CENIBRA SA, localizados no leste de Minas Gerais, Brasil (FIGURA 1).

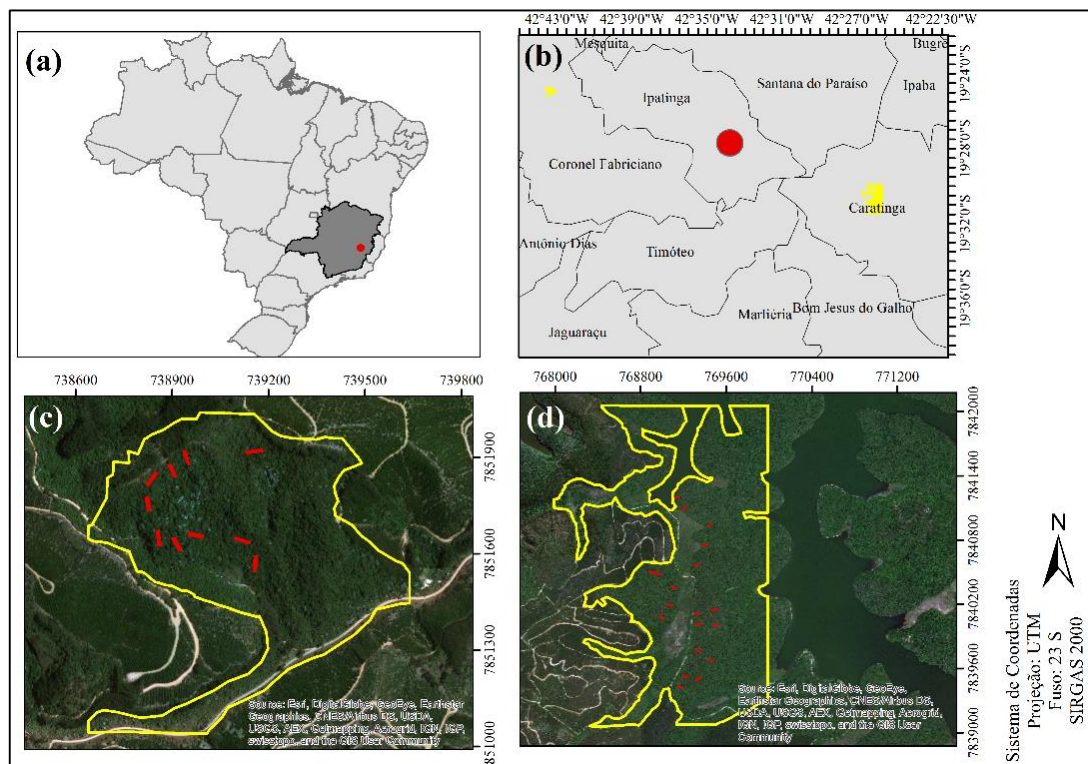


FIGURA 1: Mapa de localização dos fragmentos da Floresta Estacional Semidecidual na região leste de Minas Gerais, Brasil. Mapa do Brasil com destaque para o estado de Minas Gerais e a cidade de Ipatinga como referência (a); localização dos projetos nos municípios de estudo próximos a cidade de Ipatinga (b); Projetos da empresa CENIBRA S.A. em que (c) é o projeto São José, com a demarcação de área e a distribuição das parcelas em campo e; (d) referente ao projeto Lagoa do Piau com o contorno de sua extensão em área e a distribuição das parcelas em campo.

Projeto Lagoa do Piau

O projeto possui uma área de 373,95 hectares e está inserido na região de Ipaba, de propriedade da empresa CELULOSE NIPO-BRASILEIRA S.A. (CENIBRA), no município de Caratinga – Minas Gerais. A altitude da área de estudo varia de 150 a 450 metros. O clima é bem definido do tipo Aw – clima tropical úmido, com inverno seco e temperaturas variando entre 25° C e com verão úmido e temperaturas de 37° C. A umidade relativa média é 59,8% com variações de máxima próximo a 65,5% e com mínima de 51,8% (SOUZA et al., 2011). A precipitação pluvial média mensal é de 106,6 milímetros tendo a máxima de 271,1 mm (dezembro) e a mínima de 13,1 mm (junho) (SOUZA et al., 2011).

Projeto São José

A área desse projeto é de 60,81 hectares de florestas inequiâneas e também pertence a empresa CENIBRA S.A., encontra-se na região de Cocais, no município de Coronel Fabriciano – Minas Gerais (SOUZA et al., 2011). A altitude varia de 450 a 950 metros. A vegetação apresenta mosaico de regeneração remanescentes entre estágios inicial, médio e avançado, com afloramentos rochosos na variação altitudinal. O clima da região é classificado segundo Köppen por Cwb – tropical de altitude, com verões chuvosos e quente. A umidade relativa média foi de 65,4%, apresentando sua máxima de 69% em julho e a mínima de 61,7% em fevereiro. A precipitação pluvial média mensal registrou um valor de 85,58 mm (dezembro – máxima de 306,57 mm e agosto – mínima de 0,75 mm) (SOUZA et al., 2011).

4.2. Dados de campo e análise descritiva

Os dados de campo são provenientes da medição feita em 2012 do inventário florestal contínuo, pertencente a empresa CENIBRA S.A. O ponto central das parcelas retangulares (10 x 50 m), foram registrados com o uso de GPS (*Global Position System*). Para o projeto Lagoa do Piau foram instaladas 19 parcelas enquanto que para o projeto São José, 10 parcelas.

As informações de campo utilizadas no estudo foram: circunferência a 1,30 metros de altura (cap); altura total das árvores (ht); o número da parcela e; o projeto onde a parcela está alocada.

Para as análises subsequentes (análise descritiva, modelagem e mapeamento) foram utilizados os dados agrupados nos projetos Lagoa do Piau e São José.

4.3. Georreferenciamento das parcelas

O georreferenciamento das parcelas foi realizado com dois receptores de sinais GNSS (*Global Navigation Satellite System*), de dupla frequência, ambos do Laboratório de Engenharia de Agrimensura e Cartográfica (LEA) da Universidade Federal de Viçosa (Figura 2).

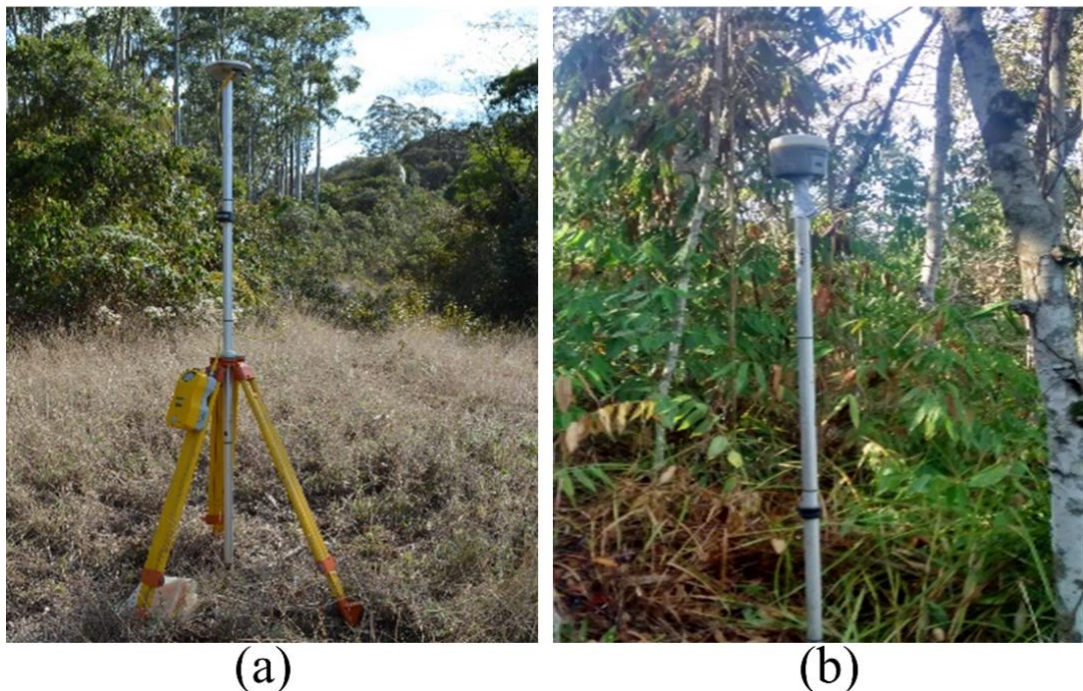


FIGURA 2: Equipamentos utilizados no georreferenciamento dos vértices das parcelas em campo, em que (a) representa o receptor geodésico fixo e (b) representa o receptor geodésico móvel.

Para isto, implantou-se pontos próximos às parcelas, situados em posições com boas condições de rastreamento para serem utilizados como base. As coordenadas destes pontos foram obtidas através do método de triangulação utilizando os dados coletados com um receptor Trimble R7-5700 (instalado nos pontos) bem como os dados de rastreamento em estações fixas pertencentes a Rede Brasileira de Monitoramento Contínuo (RBMC) dentre elas as estações localizadas nas cidades de Belo Horizonte e Governador Valadares.

Os dados dos pontos da RBMC são disponibilizados diariamente pelo IBGE (Instituto Brasileiro de Geografia e Estatística).

As coordenadas dos pontos de base foram utilizadas como referência para a obtenção das coordenadas dos pontos das parcelas, nos quais foi efetuada a coleta de dados, pelo método estático e tempo médio de ocupação de 25 minutos, com o receptor *Topomap T10*.

Para o processamento dos dados tanto dos pontos base quanto dos pontos das parcelas foi utilizado o software *Topcon Tools v8.2.3*“.

4.4. Estimativa de campo dos atributos biofísicos

Para esse trabalho utilizou as equações de volume ajustado para a formação de Mata Secundária no estado de Minas Gerais (CETEC, 1995). O volume do fuste com casca (equação 1), o volume do fuste sem a casca (2), o volume da casca (3). Para o ajuste da biomassa do fuste (4) utilizou dados referente ao volume estimado e a densidade básica da madeira dos indivíduos (SOARES et al., 2012) obtidos em trabalhos de revisão com a mesma formação florestal:

$$VFCC_j = 0,000039 \cdot DAP_j^{1,70764} \cdot HT_j^{1,32032} \quad [1]$$

$$VFSC_j = 0,000024 \cdot DAP_j^{1,575082} \cdot HT_j^{1,582642} \quad [2]$$

$$VC = VFCC - VFSC \quad [3]$$

$$BN_m = VFSC_j \cdot d \quad [4]$$

em que:

$VFCC_j$ = volume do fuste com casca do j-ésimo indivíduo, em m³;

$VFSC_j$ = volume do fuste sem casca do j-ésimo indivíduo, em m³;

DAP_j = diâmetro a altura do peito do j-ésimo indivíduo, em cm;

HT_j = altura total do j-ésimo indivíduo, em metros.

VC = volume da casca por diferença entre os volumes do fuste;

BN_m = biomassa de madeira, em Mg;

d = densidade básica da madeira Mg/ m³;

n = número de espécies.

O estoque de carbono da parte aérea foi assumido como sendo 48% da biomassa calculada na equação 4. Todos os cálculos foram realizados em nível de árvore e, posteriormente, somados por parcela. Como as parcelas apresentam tamanhos diferentes, essas foram ajustadas para 500 m², a fim de padronizar as estimativas por parcela, previamente à modelagem. Para o cálculo da superfície efetiva de cada parcela, utilizou-se os Modelos Digitais de Terreno, com 0,5 m de resolução,

e os arquivos vetoriais de cada parcela, utilizando o comando “*Add Surface Information*” encontrado no *software* ArcGIS versão 10.3 (ESRI – Redlands, CA, EUA).

4.5. Dados de sensoriamento remoto

O sobrevoo com o sensor *Light Detection And Ranging* (LiDAR) foi realizado por uma empresa contratada pela CENIBRA SA, no ano de 2014, que repassou todas as informações do voo. A aeronave utilizada foi *Cessna* de modelo 206, com velocidade do voo equivalente a 55 m/s e a uma altura de 618 m. A largura da faixa de voo foi de 713 m, com ângulo de abertura – campo de visada (FOV) de 60 graus. O *footprint* proveniente do sensor foi de 31 cm e com uma frequência de varredura equivalente a 300 kHz. Assim, a densidade de pontos na nuvem oriunda do aerolevanteamento é de 5 pontos por metro quadrado.

4.6. Processamento dos dados LiDAR e análise descritiva

No processamento da nuvem de pontos LiDAR utilizou-se o *software* FUSION, que é um *software* livre desenvolvido pelo Serviço Florestal Americano (McGAUGHEY, 2018).

Da nuvem de pontos foi obtido dois modelos de elevação: Modelo Digital de Terreno (MDT) e Modelo Digital de Superfície (MDS). Para tanto, foi realizada a separação da nuvem de pontos em duas categorias: pontos de terreno e de não terreno. A resolução espacial para a construção dos modelos foi de 0,5 metros, conforme definido pela literatura (GORGENS et al., 2011). Na geração do MDT utilizou-se o comando *GridSurfaceCreate*, enquanto que para a obtenção do MDS, o comando utilizado foi *Canopymodel* seguindo a metodologia proposta por Gorgens et al. (2014). Os modelos (MDT e MDS) foram visualizados no ArcGIS versão 10.3 (ESRI – Redlands, CA, EUA). De posse do MDT, a nuvem de pontos foi normalizada, retornando à altura dos alvos (eixo z) desconsiderando o relevo.

A separação em categorias dos pontos de terreno e de não terreno através da nuvem de pontos foi importante para realizar a normalização dos alvos, o que posteriormente nos dá a estimativa das alturas provenientes dos retornos considerando o terreno plano.

A detecção dos *outliers* presente na nuvem de pontos normalizada foi realizada visualmente no *software* ArcGIS (ESRI – Redlands, CA, EUA), não apresentando pontos destoantes do conjunto total.

A nuvem de pontos foi recortada através do georreferenciamento das parcelas em campo. O *software* R (R Core Team, 2017) pelo pacote *rLiDAR* (SILVA et al., 2017) foi utilizado para essa etapa.

O processamento de extração das métricas foi realizada pelo *software* FUSION (McGAUGHEY, 2018) usando o comando *CloudMetric* que retorna os parâmetros estatísticos descritos da nuvem de pontos LiDAR. Foi considerado como *inputs*, as limitações necessárias para realizar o comando: *minht* de 1,30 (altura mínima para obter o diâmetro a altura do peito – DAP) e o *above* foi igual a 2,0, proposto pela literatura (GORGENS et al., 2011).

As métricas provenientes da extração da nuvem de pontos (TABELA 2) abrangem medidas de Cobertura, de Elevação e de variação entre as taxas de retorno.

TABELA 1: Descrição das métricas extraídas do sensor LiDAR e sua definição para todas as parcelas dos projetos localizados a leste de Minas Gerais, Brasil. ¹ representa a fórmula: ((elevação média – elevação mínima) / (elevação máxima – elevação mínima))

Categoria	Métricas	Definição
Variáveis de cobertura	<i>Total return count</i>	Contagem de todos os retornos
	<i>Total return count above 2.00</i>	Contagem de todos os retornos acima de 2 metros
	<i>Return 1 count above 2.00</i>	Contagem do primeiro retorno acima de 2 metros
	<i>Return 2 count above 2.00</i>	Contagem do segundo retorno acima de 2 metros
	<i>Return 3 count above 2.00</i>	Contagem do terceiro retorno acima de 2 metros
	<i>Return 4 count above 2.00</i>	Contagem do quarto retorno acima de 2 metros
Variáveis de elevação	<i>Elev minimum</i>	Elevação mínima
	<i>Elev maximum</i>	Elevação máxima
	<i>Elev mean</i>	Elevação média
	<i>Elev mode</i>	Elevação modal

<i>Elev stddev</i>	Desvio padrão da elevação
<i>Elev variance</i>	Variância da elevação
<i>Elev CV</i>	Coefficiente de Variação da elevação
<i>Elev IQ</i>	Elevação Inter Quartil
<i>Elev skewness</i>	Assimetria de elevação
<i>Elev kurtosis</i>	Curtose da elevação
<i>Elev AAD</i>	Desvio Médio Absoluto da elevação
<i>Elev MAD median</i>	Mediana da elevação dos desvios médios absolutos
<i>Elev MAD mode</i>	Moda da elevação dos desvios médios absolutos
<i>Elev L1</i>	Elevação no primeiro momento
<i>Elev L2</i>	Elevação no segundo momento
<i>Elev L3</i>	Elevação no terceiro momento
<i>Elev L4</i>	Elevação no quarto momento
<i>Elev L CV</i>	Momento do coeficiente de variação da elevação
<i>Elev L skewness</i>	Momento da assimetria da elevação
<i>Elev L kurtosis</i>	Momento da curtose da elevação
<i>Elev P01</i>	Elevação no percentil 1
<i>Elev P05</i>	Elevação no percentil 5
<i>Elev P10</i>	Elevação no percentil 10
<i>Elev P20</i>	Elevação no percentil 20
<i>Elev P25</i>	Elevação no percentil 25
<i>Elev P30</i>	Elevação no percentil 30
<i>Elev P40</i>	Elevação no percentil 40
<i>Elev P50</i>	Elevação no percentil 50
<i>Elev P60</i>	Elevação no percentil 60
<i>Elev P70</i>	Elevação no percentil 70

	<i>Elev P75</i>	Elevação no percentil 75
	<i>Elev P80</i>	Elevação no percentil 80
	<i>Elev P90</i>	Elevação no percentil 90
	<i>Elev P95</i>	Elevação no percentil 95
	<i>Elev P99</i>	Elevação no percentil 99
	<i>Canopy relief ratio</i>	Relação da copa ¹
	<i>Elev SQRT mean SQ</i>	Elevação quadrática média
	<i>Elev CURT mean CUBE</i>	Elevação cúbica média
Variação entre as taxas de retornos considerando o valor mínimo (above)	<i>Percentage first returns above 2.00</i>	Porcentagem do primeiro retorno acima de 2 metros
	<i>Percentage all returns above 2.00</i>	Porcentagem de todos os retornos acima de 2 metros
	<i>(All returns above 2.00) / (Total first returns) * 100</i>	(Todos os retornos acima de 2) / (retornos do primeiro retorno) * 100
	<i>First returns above 2.00</i>	Primeiro retorno acima de 2 metros
	<i>All returns above 2.00</i>	Todos os retornos acima de 2 metros
	<i>Percentage first returns above mean</i>	Porcentagem do primeiro retorno acima da média
	<i>Percentage first returns above mode</i>	Porcentagem do primeiro retorno acima da moda
	<i>Percentage all returns above mean</i>	Porcentagem de todos os retornos acima da média
	<i>Percentage all returns above mode</i>	Porcentagem de todos os retornos acima da moda
	<i>(All returns above mean) / (Total first returns) * 100</i>	(Todos os retornos acima da média) / (todos os primeiros retornos) * 100
	<i>(All returns above mode) / (Total first returns) * 100</i>	(Todos os retornos acima da moda) / (todos os primeiros retornos) * 100

<i>First returns above mean</i>	Primeiros retornos acima da média
<i>First returns above mode</i>	Primeiros retornos acima da moda
<i>All returns above mean</i>	Todos os retornos acima da média
<i>All returns above mode</i>	Todos os retornos acima da moda
<i>Total first returns</i>	Todos os primeiros retornos
<i>Total all returns</i>	Todos os retornos

O comportamento e a tendência da distribuição da nuvem de pontos foram comparados com a realidade em campo, na análise exploratória dos dados. Plotou-se a relação entre altura média estimada nos diferentes percentis. O *software* utilizado para esse ajuste foi o *Curve Expert* (HYAMS, 2009), na versão acadêmica.

4.7. Modelagem dos atributos biofísicos

4.7.1. Regressão

4.7.1.1. Seleção de variáveis

Com as métricas LiDAR e as estimativas dos parâmetros biofísicos florestais, a partir de dados de campo, realizou-se análises de correlação entre elas, no pacote *Corr* (JACKSON, 2016) do *software* R (R Core Team, 2017). Foram obtidos os coeficientes de correlação linear de Pearson (equação 5) e correlação não paramétrica de Spearman (equação 6):

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^n (x_i - \bar{x})^2] [\sum_{i=1}^n (y_i - \bar{y})^2]}} \quad [5]$$

$$\rho = 1 - \frac{6 \sum_{i=1}^n d^2}{n(n^2 - 1)} \quad [6]$$

em que:

r = coeficiente linear de Pearson;

n = número de observações;

x e y são as medidas;

\bar{x} e \bar{y} são as médias das observações;

ρ = coeficiente de Spearman;

d = diferença entre os postos das variáveis.

A partir dessas análises foi possível selecionar qual métrica LiDAR apresentava maior coeficiente de correlação com o volume, a biomassa e o estoque de carbono acima do solo.

4.7.1.2. Ajuste do modelo

As parcelas foram separadas aleatoriamente, constituindo 70% para treinamento (20 parcelas) e 30% para a validação (9 parcelas). O ajuste equação com as parcelas de treinamento foi realizado no *software Curve Expert* (HYAMS, 2009) que tem como método *Liveberg-Marquardt* para estimação dos parâmetros. Como variáveis de entrada, utilizamos os valores preditos em campo (variável dependente) juntamente com os valores da métrica LiDAR mais bem correlacionada (variável independente) para cada atributo florestal estudado. O modelo ajustado é apresentado abaixo.

$$Y = a \cdot e^{b \cdot x} + \varepsilon \quad [7]$$

em que:

Y = Atributo biofísico estimado (V, B, C);
a e b = Parâmetros de ajuste do modelo;
x = Variável selecionada das métricas LiDAR;
ε = erro aleatório.

O modelo que melhor ajustou às parcelas de treinamento foi o exponencial. A validação do modelo foi realizada com as parcelas restantes, de validação. As medidas de precisão computadas foram o coeficiente de correlação ($r_{\hat{Y}Y}$) (equação 8) e o resíduo das amostras (RMSE) (equação 9).

$$r_{\hat{Y}Y} = \frac{\sum \hat{Y}Y - \frac{\sum \hat{Y} \cdot \sum Y}{n}}{\left[\left[\sum \hat{Y}^2 - \frac{(\sum \hat{Y})^2}{n} \right] \cdot \left[\sum Y^2 - \frac{(\sum Y)^2}{n} \right] \right]^{\frac{1}{2}}} \quad [8]$$

$$S_Y = \frac{s}{n} \quad [9]$$

em que:

$r_{\hat{Y}Y}$ = coeficiente de correlação entre os valores;

Y = valores observados;
 \hat{Y} = valores estimados;
 S_Y = erro padrão;
 s = desvio padrão;
 n = número de observações.

4.7.2. Aprendizado de máquina

4.7.2.1. Seleção de variáveis

Na análise de correlação foram desconsideradas métricas que tinham correlação superior da 90% entre elas. O número de métricas a serem utilizadas na modelagem foi definido devido sua importância adotada pelo comando *Recursive Feature Elimination* do pacote *caret* (KUNH et al, 2018) - *Random Forest* do *software* R. Para cada parâmetro (atributo biofísico) obteve-se o número ótimo de variáveis explicativas considerando um valor mínimo de tolerância. A tolerância adotada foi inferior a um. Essa tolerância equivale a uma perda de 1% do R^2 em nome da simplicidade das métricas no modelo.

A seleção do modelo e todo o processamento foi realizado pelo pacote *labgeo* (FERNANDES FILHO, 2018) no *software* R (R Core Team, 2017), criado pelo laboratório de Geoprocessamento do Departamento de Solos da Universidade Federal de Viçosa. Dentre os modelos propostos pelo pacote, utilizou-se a modelagem via regressão com o algoritmo do *Random Forest* – RF.

4.8. Mapeamento dos atributos biofísicos

Com as variáveis selecionadas, extraiu-se os valores obtidos para cada célula do comando *GridMetric*, do *software* FUSION, para toda a área em estudo, com resolução espacial equivalente a 22,4 m (501,8 m²). Com essas informações estimou-se os parâmetros biofísicos utilizando os métodos supracitados: regressão e *Random Forest*.

Foi utilizada a ferramenta *Raster Calculator* do *software* ArcGIS (ESRI – Redlands, CA, EUA) para realizar o cálculo da estimativa dos parâmetros florestais (volume, biomassa acima do solo e carbono acima do solo), com base na equação de regressão não linear ajustada. Para o modelo *Random Forest* utilizou do comando *GridMetrics* no *software* R (R Core Team, 2017) para composição das variáveis selecionadas na modelagem do aprendizado por máquina.

4.9. Validação dos mapeamentos

Afim de validar as estimativas do mapeamento, calculou os intervalos de confiança (IC) (equação 10) para os três parâmetros biofísicos, com base nos dados de campo. Para o cálculo do IC adotou probabilidade de 95% na tabela *t* de *Student* juntamente com os graus de liberdade. Verificou-se, assim, se as médias das estimativas de mapeamento no interior das 29 parcelas enquadravam-se entre os limites adotados pelo IC.

$$IC = \bar{Y} \pm t . S_{\bar{Y}} \quad [10]$$

em que:

IC = Intervalo de Confiança;

\bar{Y} = média dos valores correspondentes;

t = valor a um nível de probabilidade da tabela de *Student*, usando n-1 graus de liberdade em que n é o tamanho da amostra e;

$S_{\bar{Y}}$ = erro padrão da média para uma população finita.

Outro método de validação foi o cálculo dos resíduos (RMSE) entre os valores de referência com os valores obtidos do mapeamento foi calculado como base a validar as estimativas do mapeamento. O teste f de Graybill (1976) com nível de probabilidade ($\alpha = 0,05$), foi aplicado entre as estimativas dos atributos biofísicos mapeadas (n = 6.121), para verificar o quão similar foram os mapeamentos. De forma complementar, obteve-se a subtração das estimativas de mapeamento da regressão não linear pelo mapeamento do modelo de aprendizado de máquina para visualização espacial da diferença entre estimativas realizadas pelos dois modelos. Por fim, os valores totais de volume (m³. ha⁻¹), biomassa acima do solo (Mg/ha) e de carbono acima do solo (Mg/ha) foram calculados e comparados com as estimativas de campo.

5. RESULTADOS

5.1. Análise descritiva

5.1.1. Dados de campo

O diâmetro máximo, altura média e densidade absoluta de indivíduos para o projeto Lagoa do Piau foram, respectivamente, de 75,0 cm, 10,3 metros e 1.995

indivíduos para todas as parcelas. Já para o projeto São José foram 65,3 cm, 11,0 metros e 986 indivíduos respectivamente (TABELA 3).

TABELA 3: Descrição dos indivíduos medidos em campo nos diferentes projetos; dap = diâmetro a 1,30 m de altura

	Projeto Lagoa do Piau	Projeto São José
Número de parcelas	19	10
Dap máximo (cm)	75,00	65,25
Dap mínimo (cm)	5,00	5,00
Dap médio (cm)	10,28	11,75
Altura máxima (m)	30,00	30,00
Altura mínima (m)	1,50	1,00
Altura média (m)	10,27	10,97
Número de fustes	1995	986

A correlação dos valores estimados e observados ($r_{\hat{Y}Y}$) nas análises gráficas realizadas sobre os dados de campo apresentaram valores superiores a 59%, além de apresentar um comportamento exponencial (linha tracejada) (FIGURA 3). Esse comportamento indica a relação dos indivíduos nas 29 parcelas.

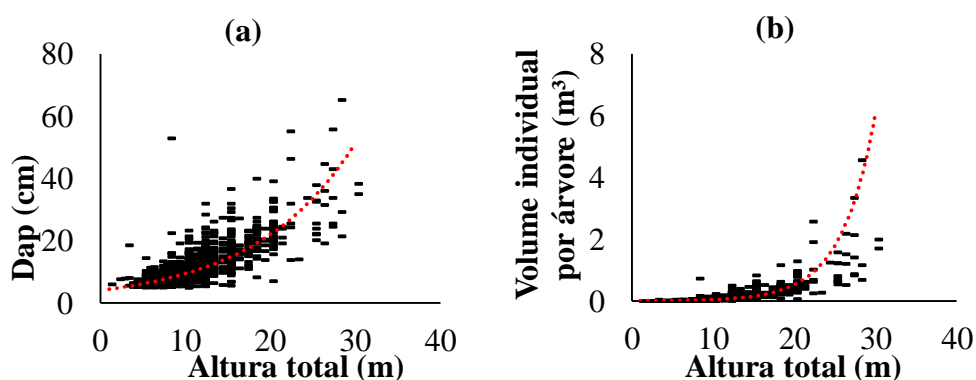


FIGURA 3: Análise gráfica da relação entre altura total (m) e diâmetro a 1,30 m de altura (cm) (a) e volume (m³), dos indivíduos mensurados em campo (n= 2.181), em 29 parcelas dos projetos Lagoa do Piau e São José, da CENIBRA S.A, leste de Minas Gerais, Brasil.

5.1.2. Dados de sensoriamento remoto

A variação altimétrica do projeto Lagoa do Piau foi de 190,31 m para o Modelo Digital de Terreno (MDT) e 360,42 m para o Modelo Digital de Superfície (MDS) (FIGURA 4). Para o projeto São José, essa variação foi de 693,63 m para o MDT e de 1002,81 m para o MDS (FIGURA 5).

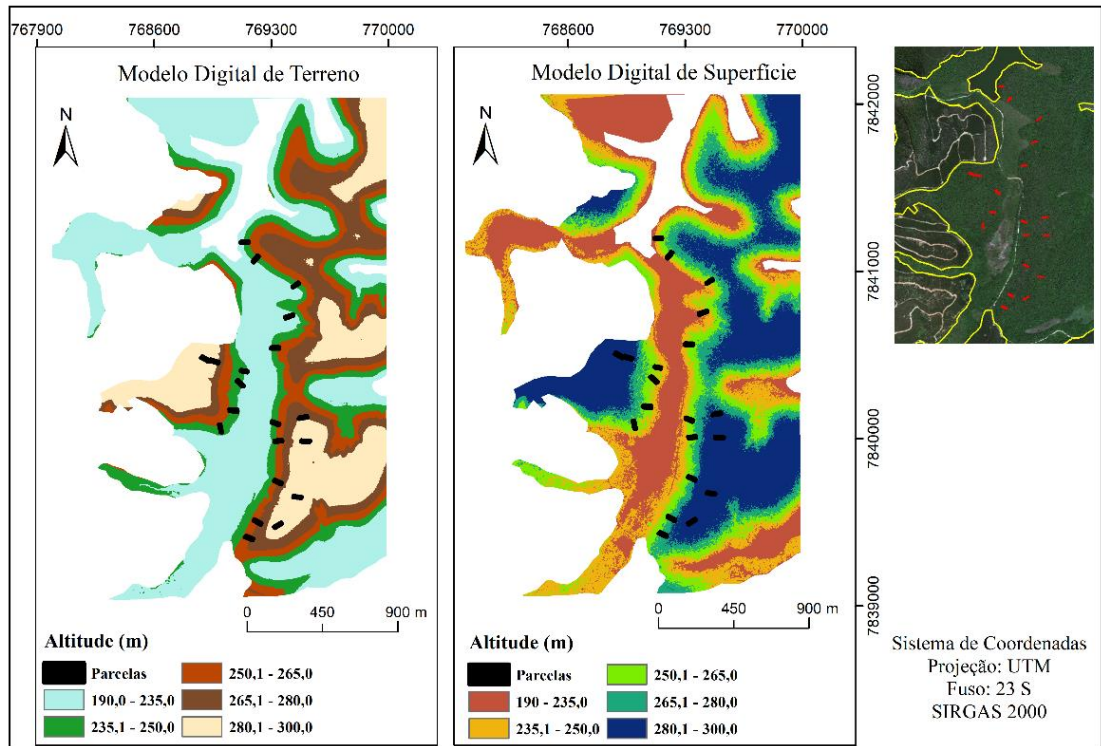


FIGURA 4: Modelo Digital de Terreno e o Modelo Digital de Superfície do projeto Lagoa do Piauí, mostrando sua variação altimétrica, em que os polígonos nas imagens são as parcelas em campo.

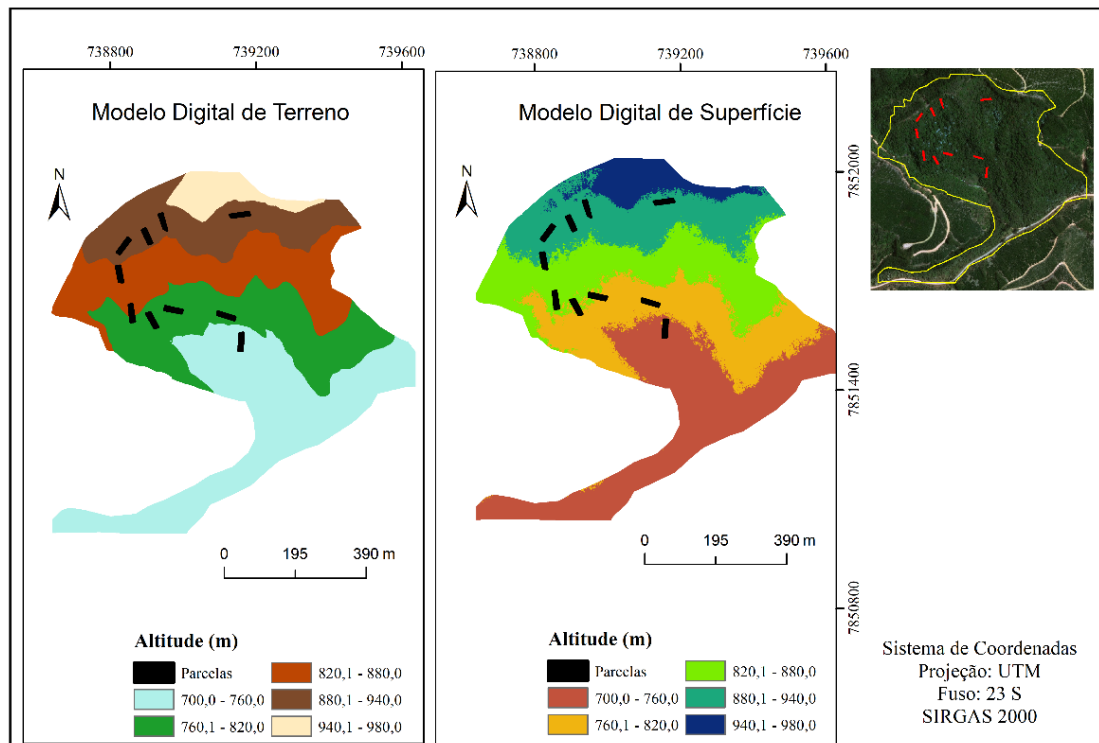


FIGURA 5: Modelo Digital de Terreno e o Modelo Digital de Superfície do projeto São José, mostrando sua variação altimétrica, em que os polígonos nas imagens são as parcelas em campo.

Desconsiderando a altimetria do terreno, os dados normalizados são essenciais para obtenção de métricas da vegetação (FIGURA 6). Com estas nuvens de pontos normalizadas das parcelas pode-se estimar via modelagem os parâmetros de interesse.

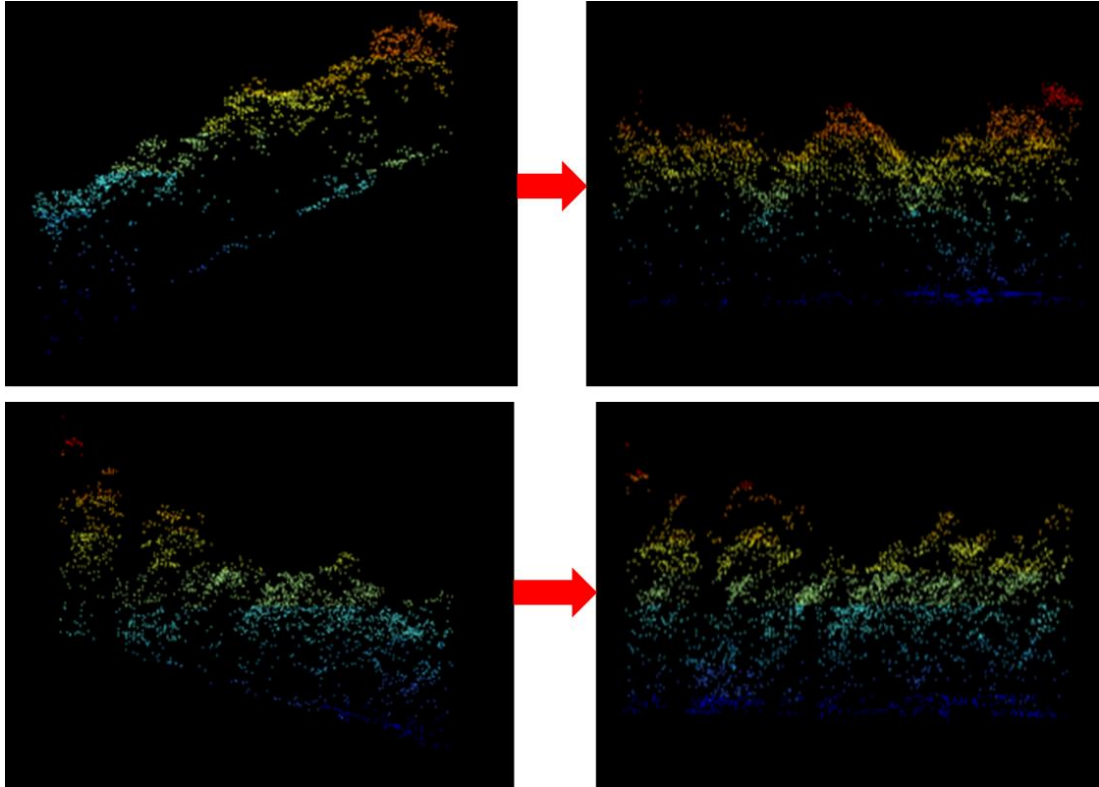


FIGURA 6: Situações da nuvem de pontos extraídas pelo LiDAR de duas parcelas, em que ao lado esquerdo da imagem são as nuvens de pontos antes da normalização e à direita as nuvens de pontos após a normalização.

Existe uma diversidade estrutural tanto entre as parcelas, quanto entre os indivíduos dentro das parcelas. Isso é possível ser observado nas nuvens de pontos normalizadas. Na representação, os tons quentes (laranja/ vermelho) remetem aos estratos superiores da floresta, em contrapartida, os pontos com de coloração fria (verde/ azul), aos estratos inferiores, como é possível ver nas parcelas 06 e 13 (FIGURA 7).

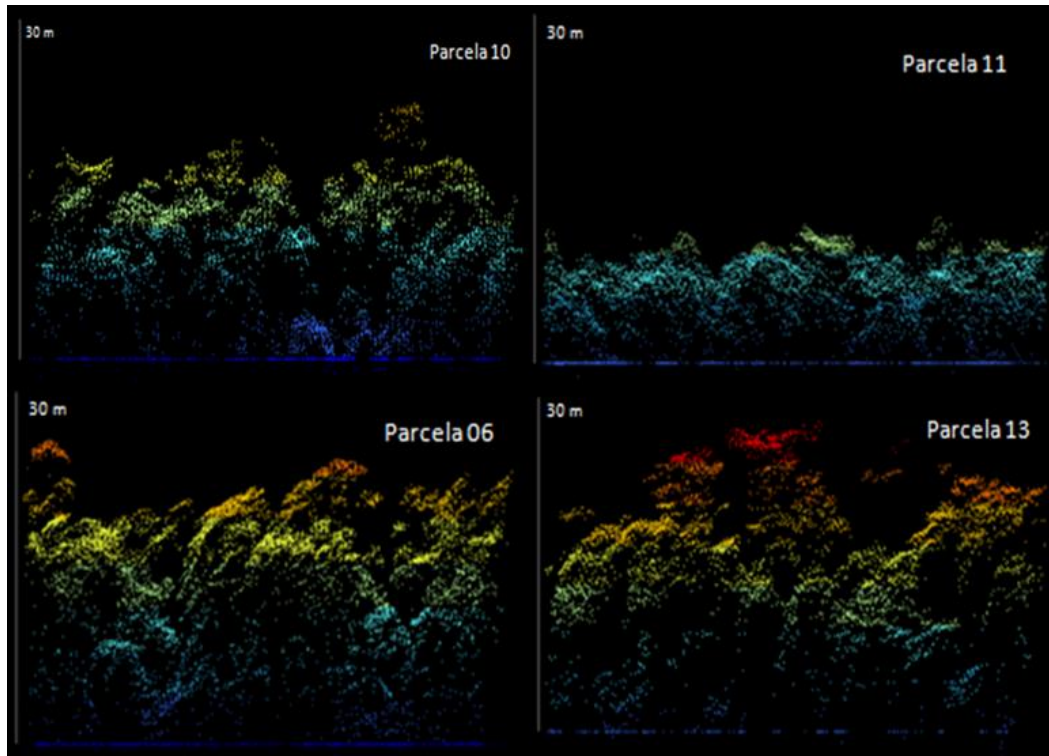


FIGURA 7: Representação da visualização tridimensional da variação estrutural das parcelas nos projetos Lagoa do Piau e São José, pertencente a empresa CENIBRA S.A.

A grande amplitude das alturas médias no percentil 99 demonstra a variabilidade em estrutura vertical das populações amostradas, uma vez que esse percentil tem relação direta com as médias das alturas presentes nas parcelas (FIGURA 8).

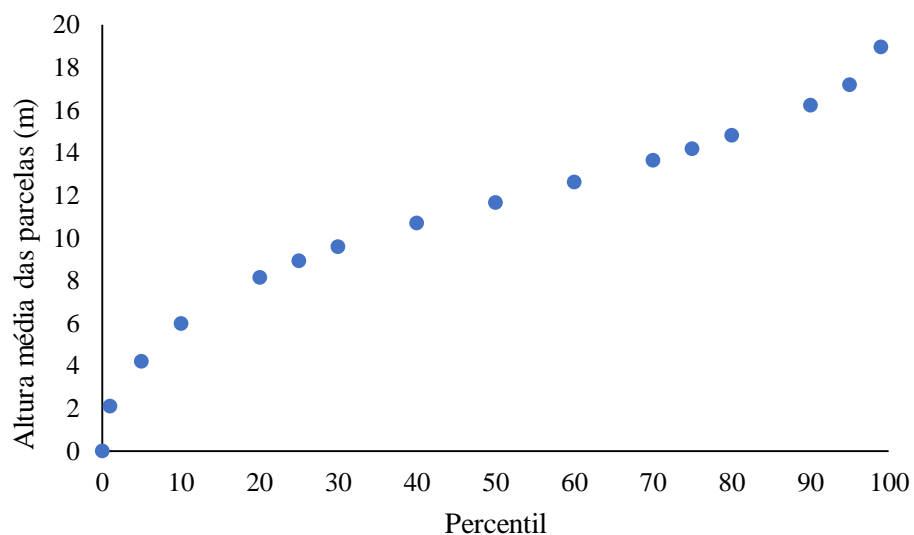


FIGURA 8: Estimativas da altura estimada pelo sensor LiDAR para as 29 parcelas da empresa CENIBRA S.A. distribuídas nos seus respectivos percentis, em que a linha tracejada demonstra a média das alturas de todas as parcelas no percentil.

A união desse gráfico (Figura 7 e 8), com o perfil da nuvem de pontos referente às parcelas 09 e 13, a exemplo, demonstra o comportamento semelhante entre essas duas visualizações (FIGURA 9). A curva retrata a densidade de pontos ao longo da estrutura vertical de cada parcela, em que sua inflexão mais acentuada caracteriza a presença de indivíduos isolados com maior altura (parcela 9). Quando os indivíduos apresentam uma tendência a alturas homogêneas, a curva tende a estabilizar no decorrer dos percentis sem grandes variações em altura (parcela 13).

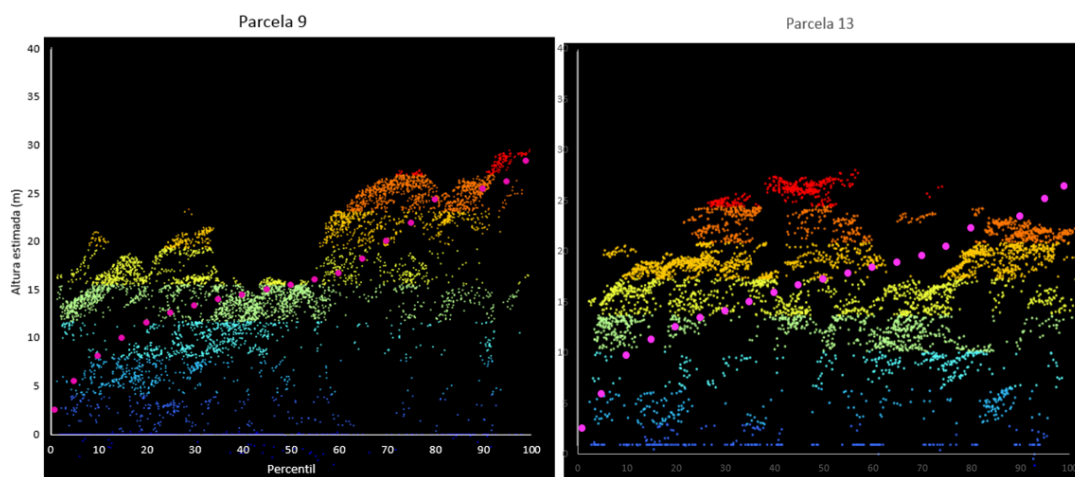


FIGURA 9: Representação gráfica das parcelas 9 e 13 (selecionadas ao acaso) da empresa CENIBRA S.A. e a respectiva imagem 3D da nuvem de pontos.

5.2. Estimativa de campo dos atributos biofísicos

Os maiores valores de volume ($592,644 \text{ m}^3. \text{ ha}^{-1}$), biomassa acima do solo ($324,956 \text{ Mg/ ha}$) e carbono acima do solo ($155,978 \text{ Mg/ ha}$) foram encontrados na parcela 4. Os menores valores encontrados para esses parâmetros foram encontrados na parcela 18 que obteve $53,976 \text{ m}^3. \text{ ha}^{-1}$ para volume, $23,584 \text{ Mg/ ha}$ para biomassa acima do solo e $11,320 \text{ Mg/ ha}$ para carbono acima do solo (TABELA 4).

TABELA 4: Estimativas de volume ajustado pela CETEC – Minas Gerais, biomassa e carbono a partir de equações estimadas através do volume para ambos os projetos da empresa CENIBRA S.A., localizada a leste de Minas Gerais, Brasil

Projeto	Parcelas	Volume (m ³ . ha ⁻¹)	Biomassa (Mg/ha)	Carbono (Mg/ha)
Lagoa do Piau	1	97,0779	64,7571	31,0834
	2	89,1666	60,5660	29,0717
	3	214,7957	140,8197	67,5935
	4	202,7261	130,0990	62,4475
	5	83,5182	55,6885	26,7305
	6	81,0597	49,1162	23,5758
	7	92,6490	60,3578	28,9717
	8	28,6393	16,7019	8,0169
	9	76,2080	44,5330	21,3758
	10	55,6481	34,3695	16,4973
	11	244,5190	158,8745	76,2597
	12	105,2635	70,8257	33,9964
	13	118,5745	80,1943	38,4933
	14	164,2794	109,3457	52,4859
	15	226,6632	148,9576	71,4997
	16	43,9199	25,7263	12,3486
	17	44,9881	26,0707	12,5139
	18	82,2681	55,8801	26,8225
	19	151,4242	102,9880	49,4342
São José	20	79,4547	62,9458	30,2140
	21	93,9188	74,6823	35,8475
	22	103,0970	79,6086	38,2121
	23	367,0474	278,4297	133,6462
	24	212,9292	150,0339	72,0163
	25	220,2246	167,4531	80,3775
	26	122,8253	95,9368	46,0496
	27	159,7006	124,8503	59,9281
	28	148,9706	117,7184	56,5048
	29	153,4599	98,9213	47,4822

5.3. Modelagem dos atributos biofísicos

5.3.1. Seleção de variáveis e treinamento

5.3.1.1. Regressão

Pela análise de correlação linear de Pearson, o percentil 10 apresentou a maior correlação para os três parâmetros florestais de interesse (0,893 para volume e 0,903 para biomassa acima do solo e carbono acima do solo), enquanto que pela análise de correlação de Spearman, a altura estimada no percentil 40 foi a variável que apresentou

a maior correlação monótona, linear ou não, com aqueles parâmetros, sendo 0,932 para volume e 0,958 para biomassa e carbono acima do solo (TABELA 5).

TABELA 5: Coeficientes de correlação de Pearson e de Spearman entre os parâmetros florestais (volume, biomassa e carbono acima do solo) e as métricas LiDAR de elevação em diferentes percentis (P)

Métricas	Volume		Biomassa		Carbono	
	Pearson	Spearman	Pearson	Spearman	Pearson	Spearman
...
Elevação P01	0,8461	0,8177	0,8530	0,8296	0,8530	0,8296
Elevação P05	0,8209	0,7857	0,8309	0,8020	0,8309	0,8020
Elevação P10	0,8933	0,8759	0,9032	0,8847	0,9032	0,8847
Elevação P20	0,8893	0,9020	0,9024	0,9217	0,9024	0,9217
Elevação P25	0,8831	0,9108	0,8974	0,9379	0,8974	0,9379
Elevação P30	0,8836	0,9192	0,8978	0,9468	0,8978	0,9468
Elevação P40	0,8882	0,9320	0,8994	0,9581	0,8994	0,9581
Elevação P50	0,8916	0,9286	0,9012	0,9537	0,9012	0,9537
Elevação P60	0,8894	0,9192	0,8996	0,9458	0,8996	0,9458
Elevação P70	0,8670	0,9079	0,8826	0,9384	0,8826	0,9384
Elevação P75	0,8543	0,9089	0,8712	0,9429	0,8712	0,9429
Elevação P80	0,8454	0,8975	0,8628	0,9325	0,8628	0,9325
Elevação P90	0,8281	0,9064	0,8498	0,9369	0,8498	0,9369
Elevação P95	0,8343	0,8916	0,8590	0,9320	0,8590	0,9320
Elevação P99	0,7722	0,8532	0,8063	0,8926	0,8063	0,8926
...

Uma correlação não linear entre os parâmetros biofísicos e altura no percentil foi encontrada (FIGURA 10).

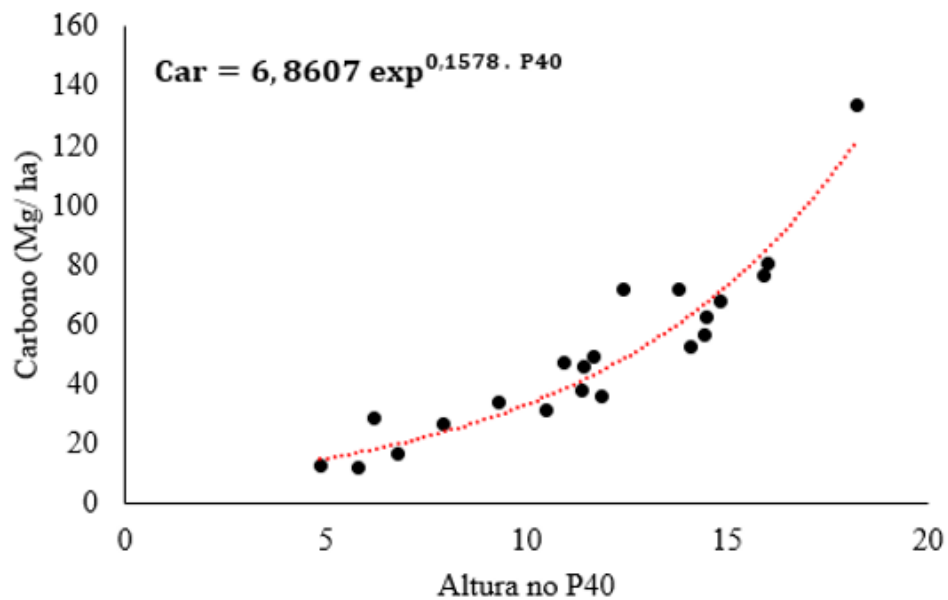
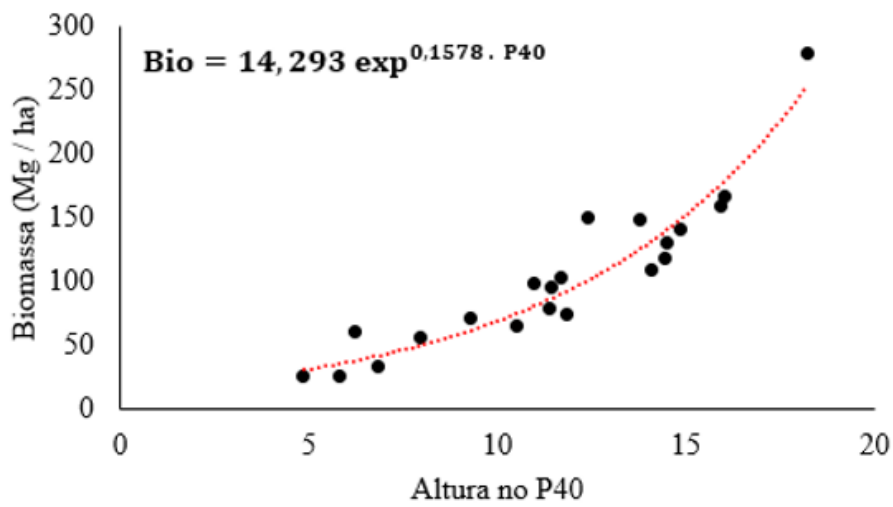
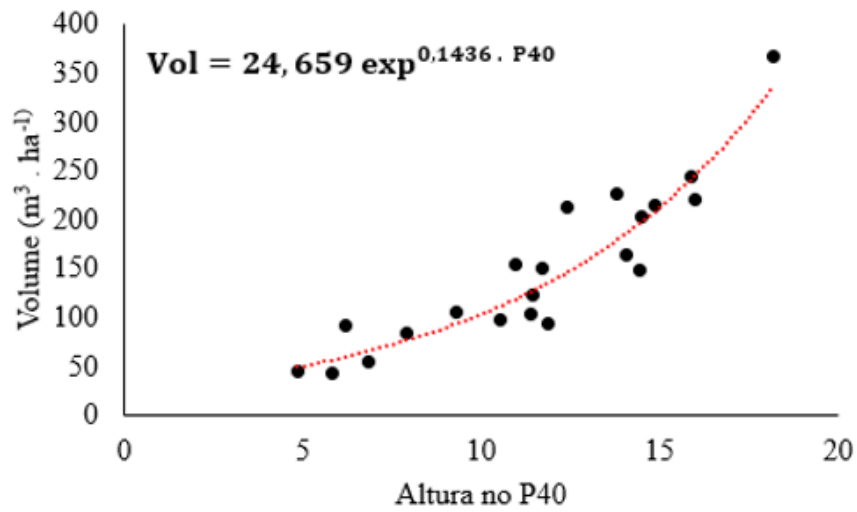
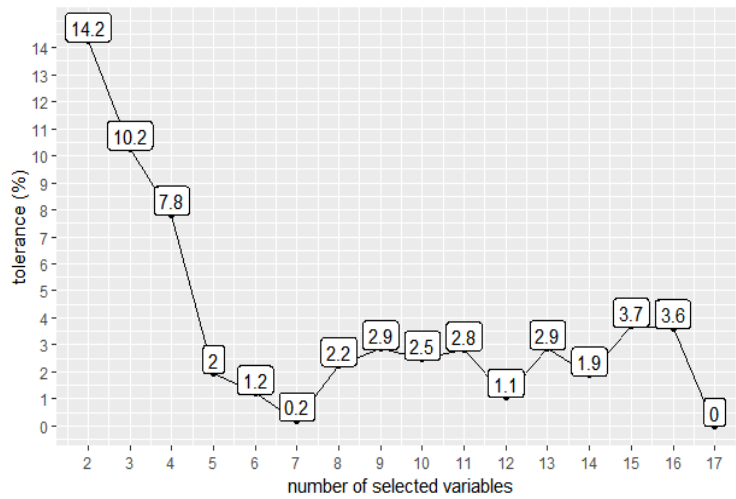


FIGURA 10: Análise gráfica da relação entre volume, biomassa acima do solo e carbono acima do solo com suas respectivas alturas estimadas no percentil 40 (H40), em 29 parcelas de 500 m².

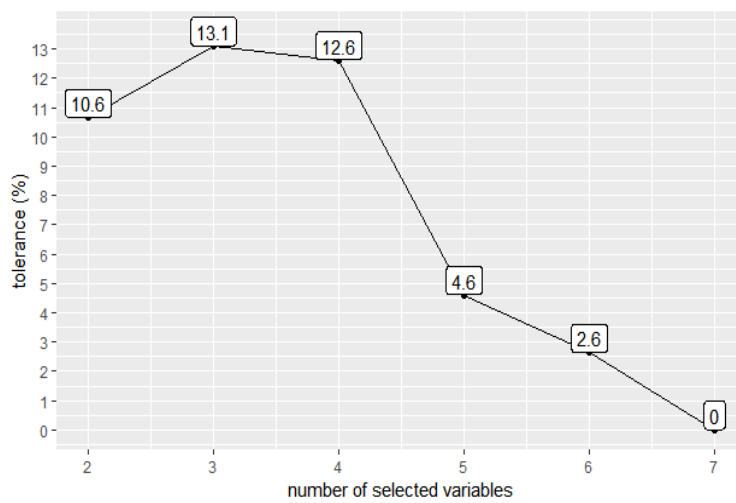
Com as parcelas de treinamento (n= 20) o melhor ajuste de equações para os três atributos biofísicos foi com o modelo exponencial, apresentando forte correlação entre as variáveis, segundo os coeficientes ajustados

5.3.1.2. *Random Forest*

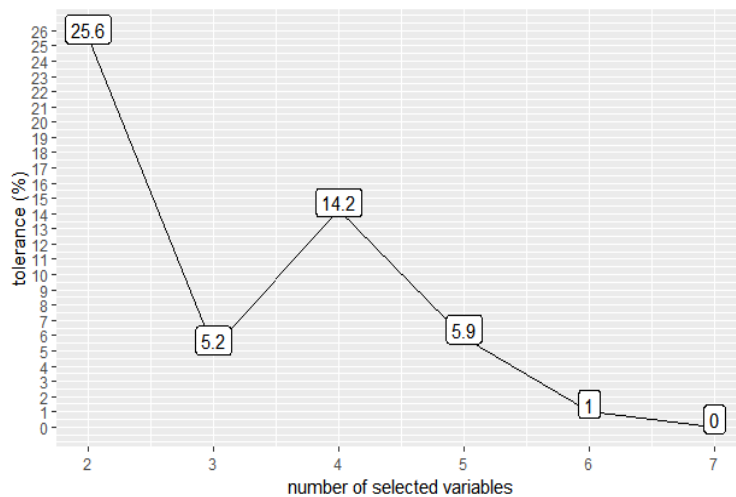
Para o emprego do *Random Forest* foram utilizadas 25 variáveis extraídas do LiDAR. Dessas 25 variáveis candidatas para a modelagem dos atributos biofísicos, apenas sete foram as escolhidas para estimar o volume, a biomassa acima do solo e o carbono acima do solo de acordo com suas importâncias para os modelos (FIGURA 11).



a



b



c

FIGURA 11: Gráficos do número de variáveis com sua perda de precisão de acordo com a o nível de tolerância para o volume (a); a biomassa acima do solo (b) e; o carbono acima do solo (c).

A seleção das sete variáveis escolhidas para o modelo de aprendizado por máquina foram as mesmas entre os parâmetros, contudo, a ordem de importância alterou de um atributo para outro (TABELA 7).

TABELA 7: Variáveis selecionadas para compor a modelagem na estimativa de volume, biomassa acima do solo e carbono acima do solo

Ordem das variáveis	Parâmetros biofísicos		
	Volume	Biomassa	Carbono
1	<i>Elev mode</i>	<i>Elev mode</i>	<i>Elev mode</i>
2	<i>Elev maximum</i>	<i>Elev P01</i>	<i>Elev maximum</i>
3	<i>Elev P01</i>	<i>Elev IQ</i>	<i>Elev L3</i>
4	<i>Elev IQ</i>	<i>Percentage first returns above 2.00</i>	<i>Percentage first returns above 2.00</i>
5	<i>Elev L4</i>	<i>Elev maximum</i>	<i>Elev P01</i>
6	<i>Elev L3</i>	<i>Elev L4</i>	<i>Elev L4</i>
7	<i>Percentage first return above 2.00</i>	<i>Elev L3</i>	<i>Elev IQ</i>

5.3.1.3. Validação externa

Com as parcelas destinadas à validação (30% das parcelas) das equações exponenciais, obteve-se as correlações de estimado e observado superiores a 0,88 e para o ajuste por aprendizado por máquina a correlação entre os valores estimados e observados de 0,96 ambos para os três atributos biofísicos (TABELA 8). Os valores de erro padrão do aprendizado por máquina foram inferiores que o modelo exponencial.

TABELA 8: Estatísticas de validação das equações geradas para estimativa de volume, biomassa e carbono acima do solo a partir de dados LiDAR, em que Exp. é a equação ajustada do modelo exponencial, RF é *Random Forest*, com área de 500 m²

Parâmetros florestais	$r_{\hat{Y}Y}$		S_Y	
	Exp.	RF	Exp.	RF
Volume	0,889	0,967	4,629 m ³	0,505 m ³
Biomassa acima do solo	0,928	0,971	3,358 Mg	0,273 Mg
Carbono acima do solo	0,928	0,970	1,612 Mg	0,135 Mg

5.4. Mapeamento dos parâmetros florestais

O mapeamento do volume utilizando a equação exponencial e *Random Forest* (RF) foi aplicado para os projetos Lagoa do Piau (FIGURA 12), e São José (FIGURA 13). O mapeamento utilizando o ajuste da equação exponencial apresentaram volumes mais elevados quando comparado com o RF.

No projeto Lagoa do Piau cuja área corresponde a 373,95 hectares, a estimativa total foi de 35.478,5417 m³ do valor estimado para o modelo exponencial enquanto que pelo *Random Forest* esse valor foi 13% superior (40.152,3973 m³) da estimativa da equação exponencial.

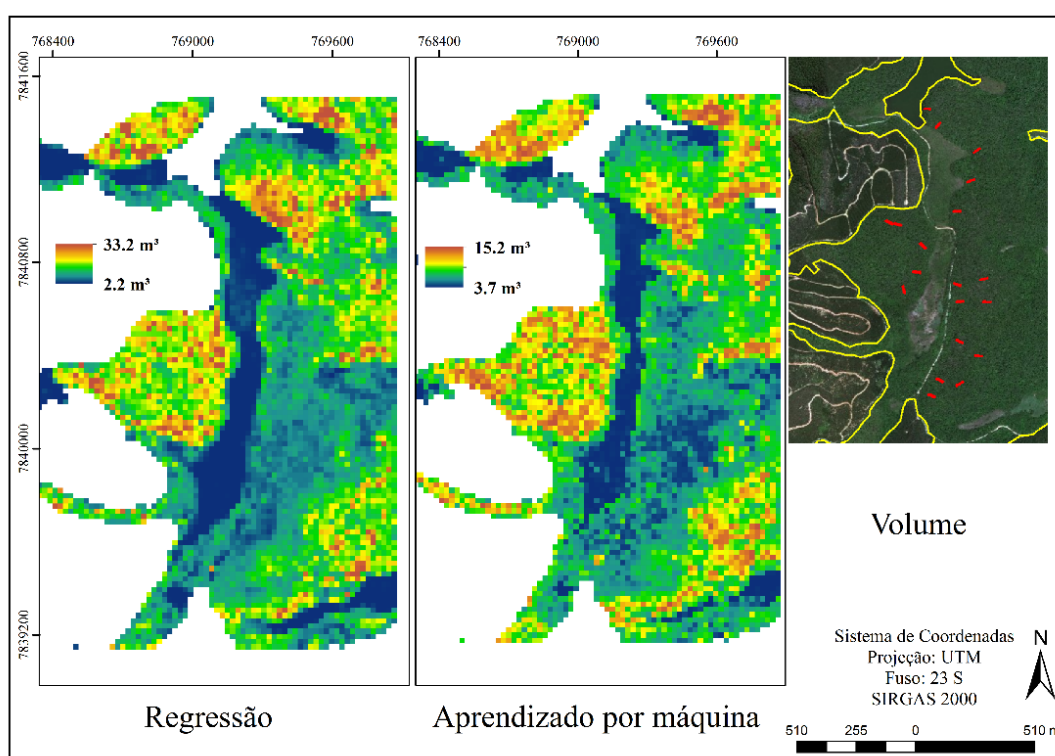


FIGURA 12: Estimativas de volume (m³) para o projeto Lagoa do Piau da empresa CENIBRA S.A., leste de Minas Gerais, Brasil, geradas pelos métodos de regressão (equação exponencial) e aprendizado por máquina (*Random Forest*), onde as cores frias (azul/verde) representam valores baixos e cores quentes (laranja/ vermelho) valores altos. No canto superior direito, imagem em cores reais de alta resolução espacial (*GeoEye, GoogleEarth*®) com delimitação da área mapeada em amarelo e distribuição das parcelas em vermelho.

Para o projeto São José de área equivalente a 60,81 hectares, a estimativa do volume estimado forma semelhantes para os ajustes, tendo 9.519,6211 m³ para o ajuste da equação exponencial em para o *Random Forest* 4% superior (9.907,0213 m³).

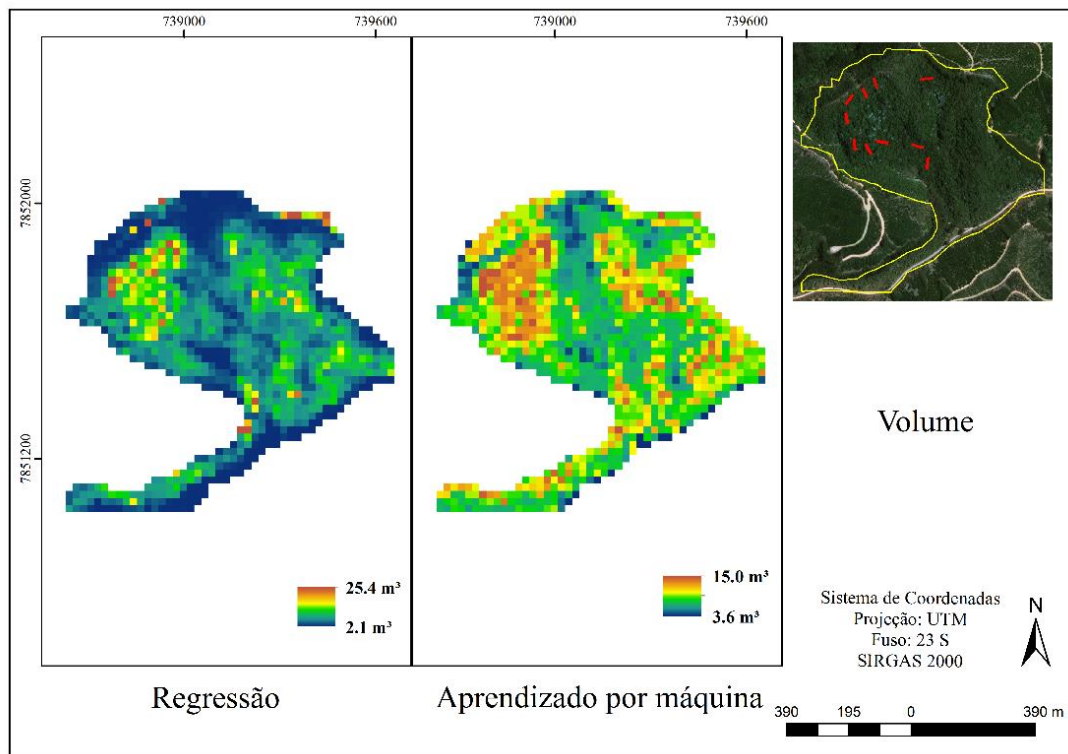


FIGURA 13: Estimativas de volume (m^3) para o projeto São José da empresa CENIBRA S.A., leste de Minas Gerais, Brasil, geradas pelos métodos de regressão (equação exponencial) e aprendizado por máquina (*Random Forest*), onde as cores frias (azul/verde) representam valores baixos e cores quentes (laranja/vermelho) valores altos. No canto superior direito, imagem em cores reais de alta resolução espacial (*GeoEye, GoogleEarth®*) com delimitação da área mapeada em amarelo e distribuição das parcelas em vermelho.

A estimativa do mapeamento da biomassa acima do solo seguiu a mesma tendência das estimativas do volume, em que os valores estimados pelo ajuste da regressão não linear (exponencial) tiveram valores maiores no projeto Lagoa do Piauí (FIGURA 14) e no projeto São José (FIGURA 15).

Considerando a estimativa total de biomassa acima do solo, o projeto Lagoa do Piauí (373,95 ha) apresentou 17.605,5917 Mg para o ajuste da equação exponencial e 13% a mais do valor encontrado para *Random Forest* (19.943,0578 Mg).

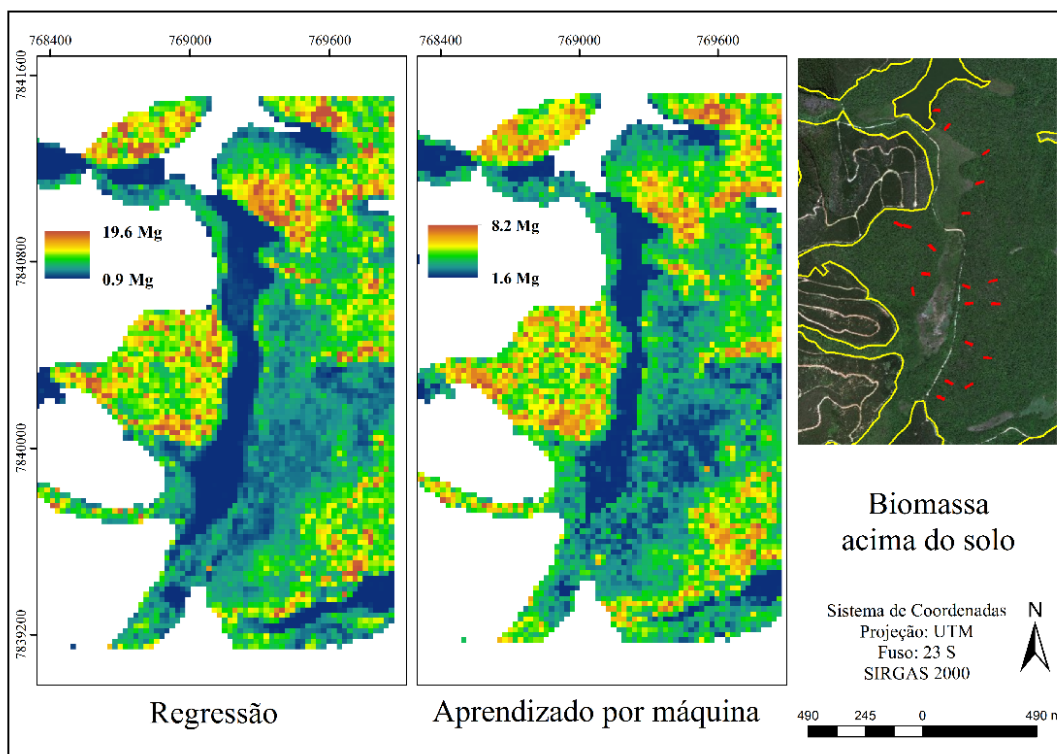


FIGURA 14: Estimativas de biomassa acima do solo (Mg) para o projeto Lagoa do Piauí da empresa CENIBRA S.A., leste de Minas Gerais, Brasil, geradas pelos métodos de regressão (equação exponencial) e aprendizado por máquina (*Random Forest*), onde as cores frias (azul/verde) representam valores baixos e cores quentes (laranja/vermelho) valores altos. No canto superior direito, imagem em cores reais de alta resolução espacial (*GeoEye*, *GoogleEarth*®) com delimitação da área mapeada em amarelo e distribuição das parcelas em vermelho.

No projeto São José (60,81 ha), as estimativas da biomassa acima do solo para os ajustes tiveram semelhanças entre as estimativas. Considerando que *Random Forest* apresentou 2% a mais da estimativa da equação exponencial, 5.056,2163 Mg e 4.949,6211 Mg, respectivamente.

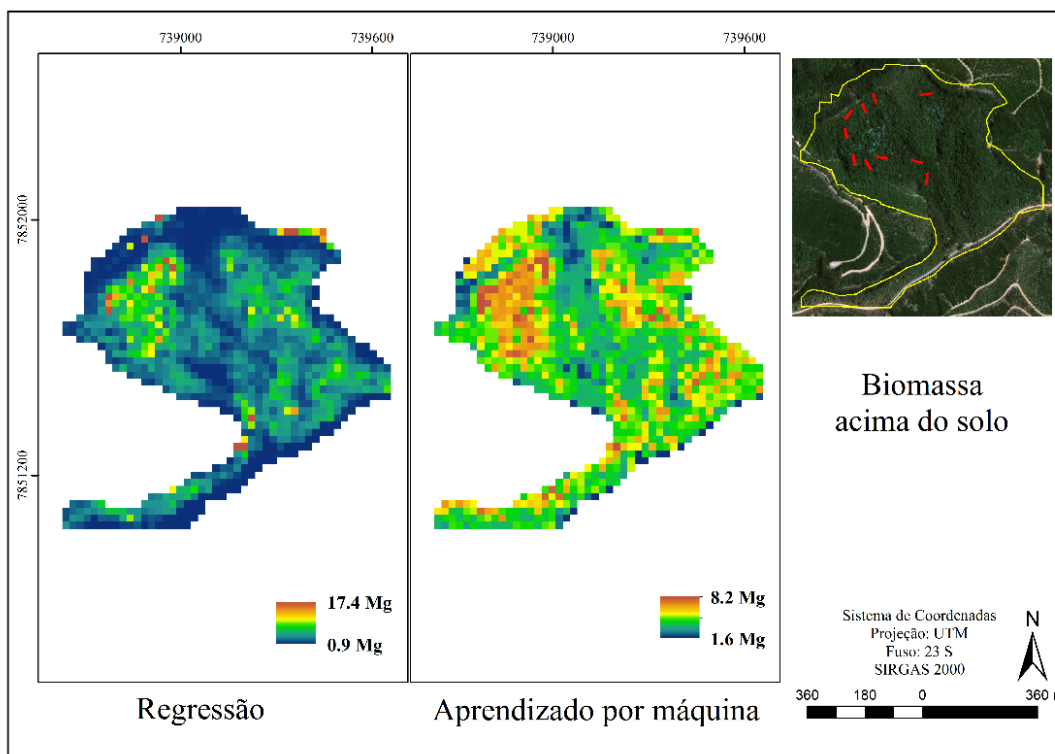


FIGURA 15: Estimativas de biomassa acima do solo (Mg) para o projeto São José da empresa CENIBRA S.A., leste de Minas Gerais, Brasil, geradas pelos métodos de regressão (equação exponencial) e aprendizado por máquina (*Random Forest*), onde as cores frias (azul/verde) representam valores baixos e cores quentes (laranja/vermelho) valores altos. No canto superior direito, imagem em cores reais de alta resolução espacial (*GeoEye*, *GoogleEarth*®) com delimitação da área mapeada em amarelo e distribuição das parcelas em vermelho.

Assim como para o volume e para a biomassa acima do solo, a estimativa de carbono acima do solo obteve o comportamento semelhante no ajuste, em que o ajuste da equação exponencial apresentou, pontualmente maiores valores de estimativas do que o *Random Forest* tanto para o projeto Lagoa do Piau (FIGURA 16) quanto para o projeto São José (FIGURA 17).

Estimativas de carbono acima do solo através do ajuste do *Random Forest* foram 16% superiores do que para equação exponencial 9.838,8699 Mg e 8.449,9958 Mg, respectivamente.

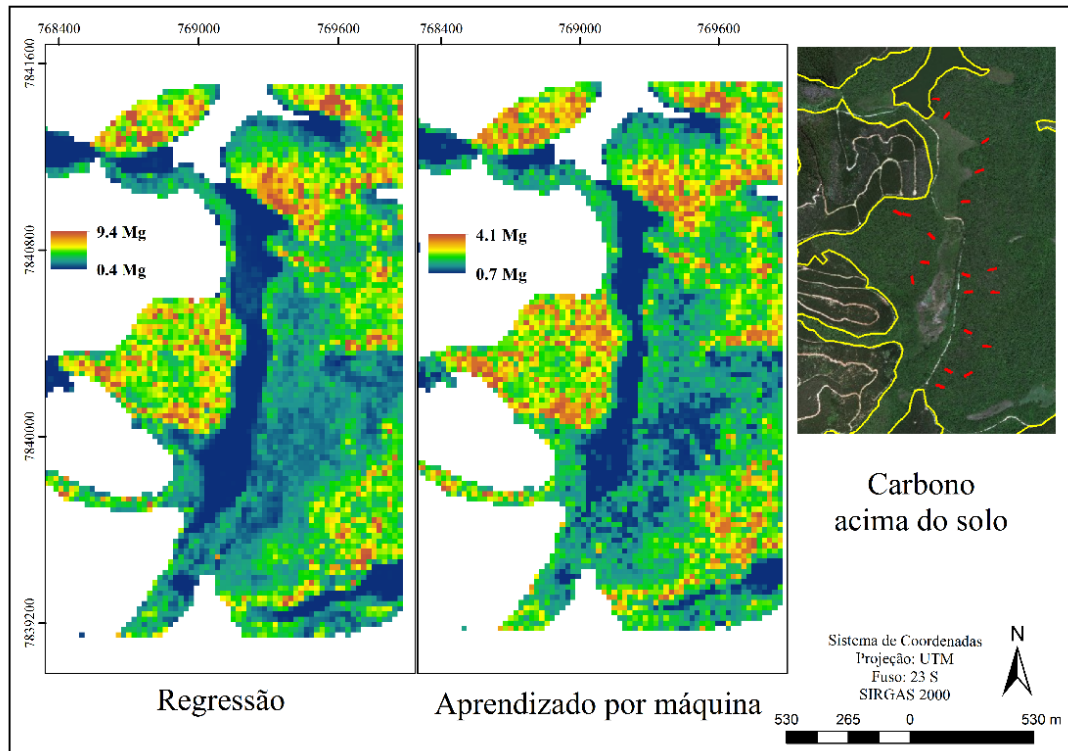


FIGURA 16: Estimativas de carbono acima do solo (Mg) para o projeto Lagoa do Piau da empresa CENIBRA S.A., leste de Minas Gerais, Brasil, geradas pelos métodos de regressão (equação exponencial) e aprendizado por máquina (*Random Forest*), onde as cores frias (azul/verde) representam valores baixos e cores quentes (laranja/vermelho) valores altos. No canto superior direito, imagem em cores reais de alta resolução espacial (*GeoEye*, *GoogleEarth*®) com delimitação da área mapeada em amarelo e distribuição das parcelas em vermelho.

No projeto São José (60,81 ha) obteve a estimativa de carbono acima do solo semelhantes nos métodos ajustado, diferindo apenas em 1% dos valores estimados entre eles. A equação exponencial estimou 2.375,7139 Mg enquanto *Random Forest* obteve 2.408,5671 Mg.

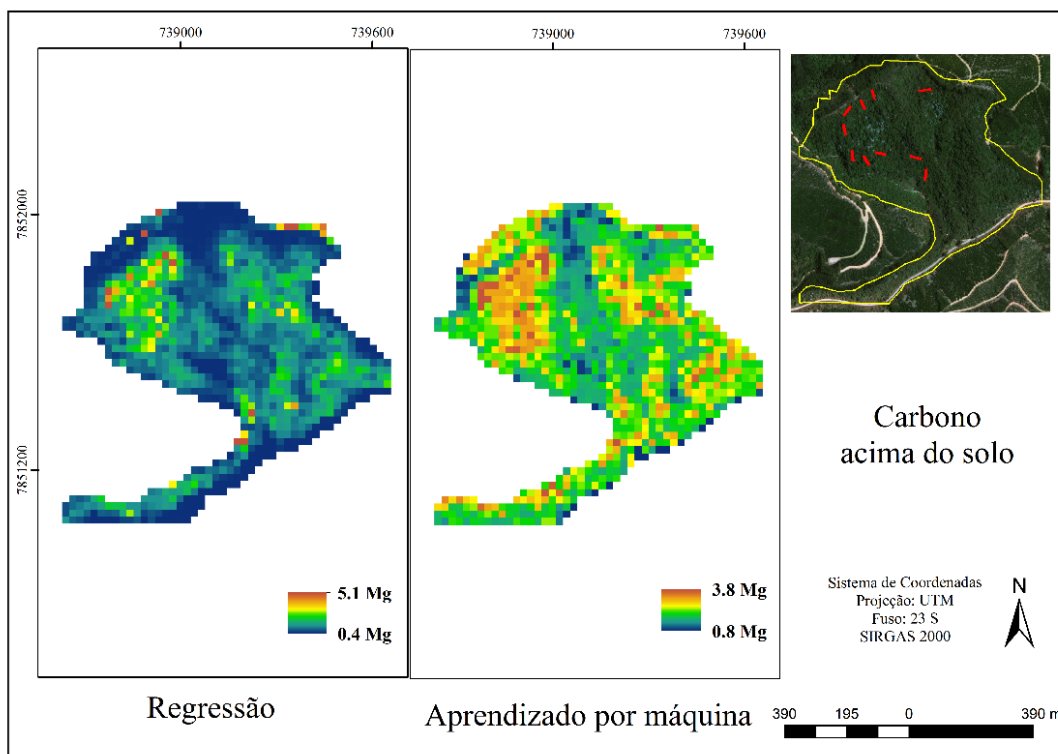


FIGURA 17: Estimativas de biomassa acima do solo (Mg) para o projeto São José da empresa CENIBRA S.A., leste de Minas Gerais, Brasil, geradas pelos métodos de regressão (equação exponencial) e aprendizado por máquina (*Random Forest*), onde as cores frias (azul/verde) representam valores baixos e cores quentes (laranja/vermelho) valores altos. No canto superior direito, imagem em cores reais de alta resolução espacial (*GeoEye*, *GoogleEarth*®) com delimitação da área mapeada em amarelo e distribuição das parcelas em vermelho.

5.5. Validação dos mapeamentos

As estimativas médias, considerando as 29 parcelas, de cada atributo biofísico (volume, biomassa acima do solo e carbono acima do solo) apresentam-se entre os limites do Intervalo de Confiança ($\alpha = 0,05$) calculado para essas mesmas parcelas, mas a partir das estimativas de campo (TABELA 9). Contudo, o RF obteve os maiores valores quando comparado com a regressão não linear para o volume, a biomassa e carbono acima do solo.

TABELA 9: Análise dos limites superiores e limites inferiores para as estimativas de mapeamento via ajuste de equação do modelo exponencial e do *Random Forest* para volume, biomassa acima do solo e carbono acima do solo para os projetos da empresa CENIBRA S.A. localizada a leste do estado de Minas Gerais, Brasil

	Intervalo de Confiança (5%; 28)		Média dos modelos			
	Limite Superior	Limite Inferior	Exponencial		<i>Random Forest</i>	
			Validação	Todas	Validação	Todas
Volume (m ³ . ha ⁻¹)	161,94	104,62	89,12	131,28	143,93	178,39
Biomassa acima do solo (Mg/ha)	113,80	71,48	59,13	91,05	72,40	90,51
Carbono acima do solo (Mg/ha)	54,62	34,31	28,38	43,70	35,85	44,56

A análise gráfica de validação para os dois métodos (Exponencial e *Random Forest*) das 29 parcelas teve aproximação dos valores estimados pelos valores de referência, sendo evidenciado nos pontos próximos a linha criada de 45° (FIGURA 18), em que os marcadores de formato circular remetem as parcelas destinadas ao treinamento e os marcadores de formato triangular as parcelas de validação para ambos métodos.

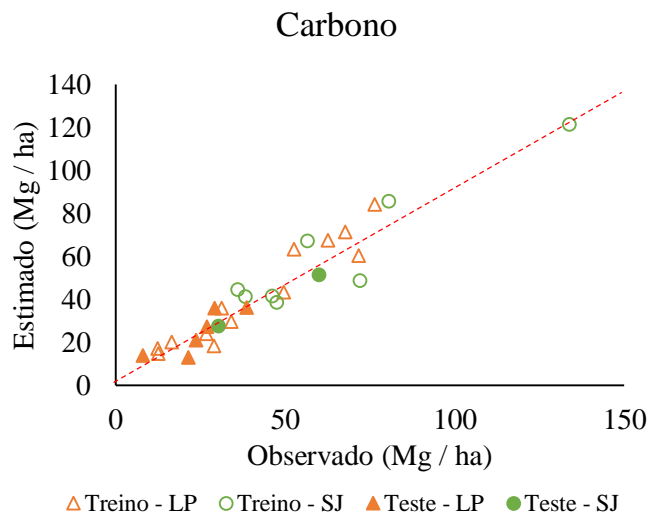
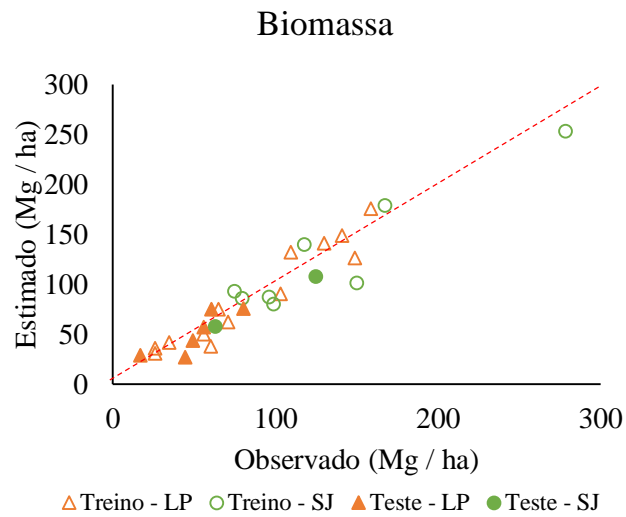
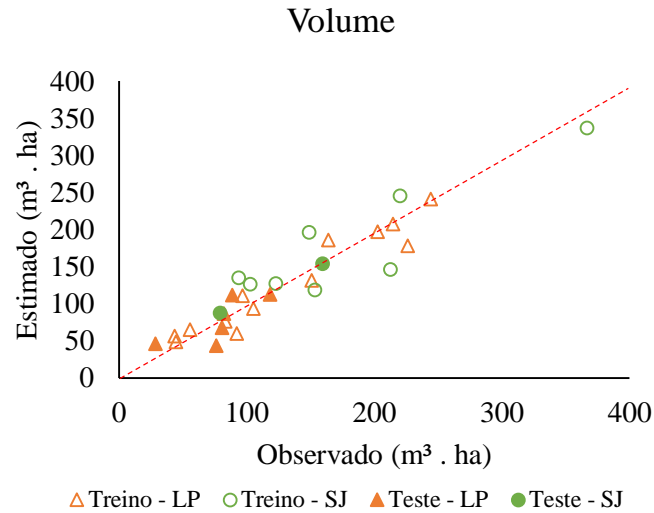


FIGURA 18: Análise gráfica da relação entre valores de referência com valores estimados de volume, de biomassa acima do solo e de carbono acima do solo através do ajuste da equação exponencial e do *Random Forest*. Em que SJ = Projeto São José e LP = Projeto Lagoa do Piauí.

Ao analisar o teste F de Graybill para as parcelas destinadas ao treinamento e validação, observamos p-valor maior que a probabilidade analisada (p-valor = 0,05). Para a estimativa de volume (p-valor = 0,14), de biomassa acima do solo (p-valor = 0,32) e para carbono acima do solo (p-valor = 0,32) temos que os métodos estimados são diferentes entre si (p-valor > 0,05). Quando estimou para todas as parcelas dos projetos, o teste F de Graybill demonstrou, que as estimativas mapeadas por meio dos dois métodos são diferentes entre si (p < 0,05), para volume, biomassa acima do solo e carbono acima do solo. Ao plotar os valores estimados da regressão e do Random Forest, observamos que para as parcelas mensuradas, as estimativas tendem a ter uma correlação linear entre elas. Porém, quando estimamos pelos métodos para toda a área dos dois projetos, observamos que para os três atributos, o Random Forest tende a estabilizar num determinado valor estimado (FIGURA 19). Isso ocorre porque ele só consegue estimar os valores dentro dos valores que foi inserido no treinamento. A equação exponencial apresentou valores superiores enquanto que as estimativas via aprendizado por máquina tornam constante a um determinado valor. Em todas as estimativas o p-valor (<< 0,00) foi inferior ao valor da probabilidade. O que recusa a hipótese H0 de que as estimativas dos modelos seriam iguais para os três atributos biofísicos.

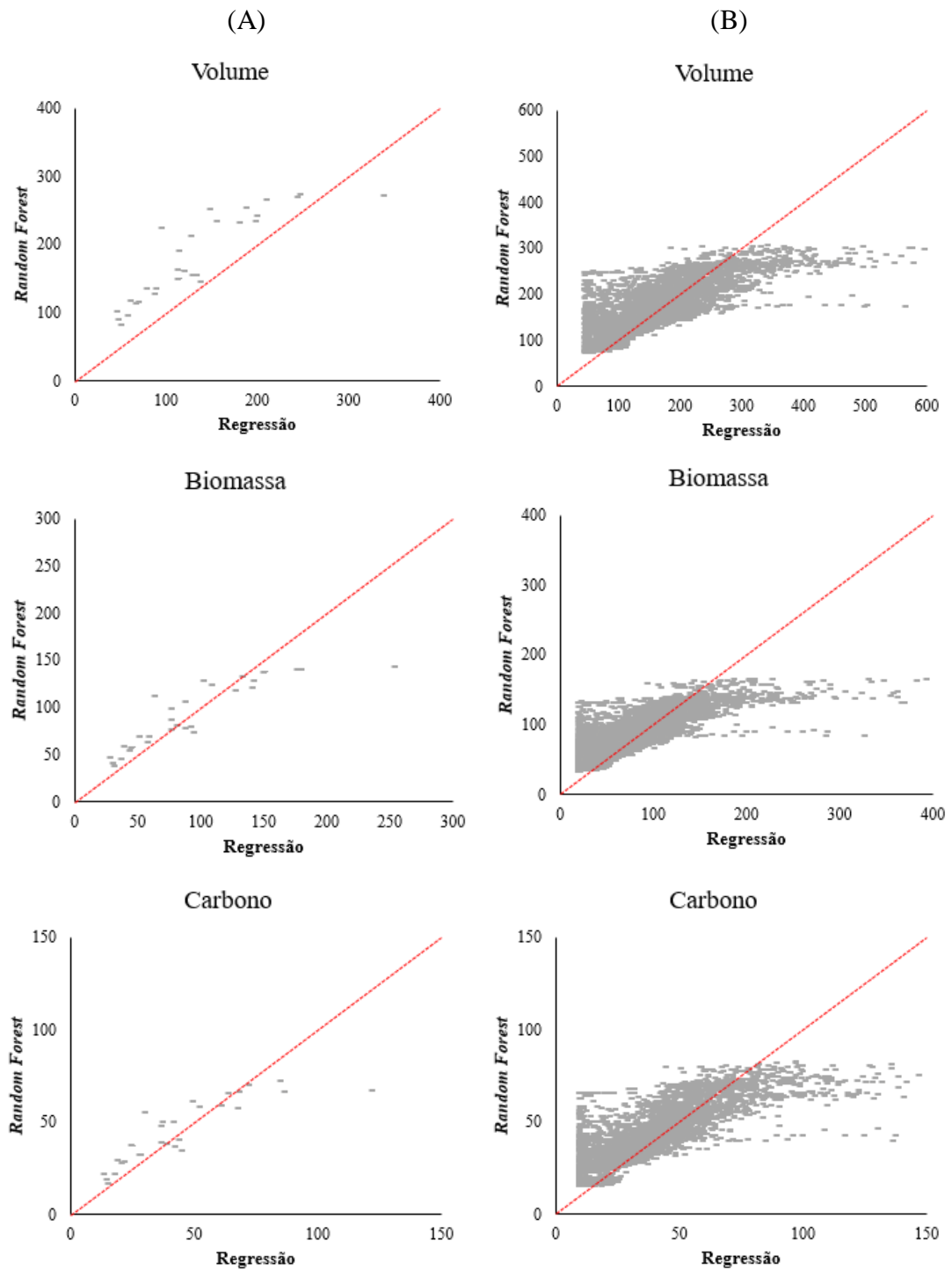


FIGURA 19: Plotagem da relação entre valores de referência (regressão) e os valores estimados (*Random Forest*) das 29 parcelas (A); e para as 6.121 parcelas (B) para os três atributos analisados.

O teste t pareado, calculado para as parcelas mensuráveis e as parcelas totais dos projetos proporciona uma outra abordagem para as estatísticas descritivas (TABELA 10). Para as parcelas mensuráveis e para as parcelas totais, os três atributos entre os

valores do intervalo de t tabelado, todos tendo apresentaram evidências para rejeitar a hipótese de que as médias fossem iguais.

TABELA 10: Estimativas dos valores t pareado para os métodos de regressão e Random Forest nas duas amostragens.

	T pareado (mensuradas)	T pareado (totais)
Volume	-0,22906	-0,0005
Biomassa	0,00354	-0,0005
Carbono	-,02298	-0,0017
T tabelado	-2,048 a 2,048	-1,645 a 1,645

Para o projeto Lagoa do Piau, os modelos se mostraram bem similares com alguns pontos em que o ajuste do modelo exponencial apresentou maiores valores do que o *Random Forest* pela coloração marrom avermelhada dos mapas. Enquanto que para o projeto São José, as estimativas do *Random Forest* foram superiores ao do ajuste da regressão não linear tornando a coloração do mapa azulada. Esse comportamento pode ser visualizado para volume (FIGURA 20), para biomassa acima do solo (FIGURA 21) e para carbono acima do solo (FIGURA 22).

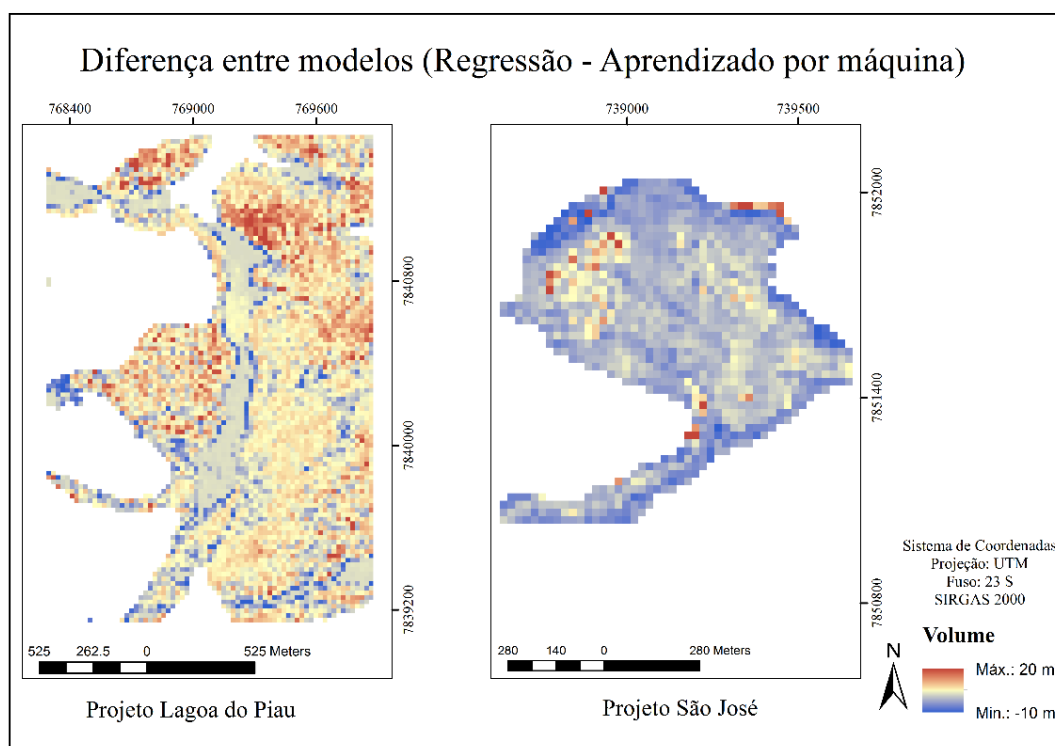


FIGURA 20: Diferença das estimativas de volume (m^3) entre os modelos ajustados destinados aos projetos da empresa CENIBRA S.A., localizada a leste do estado de Minas Gerais, Brasil.

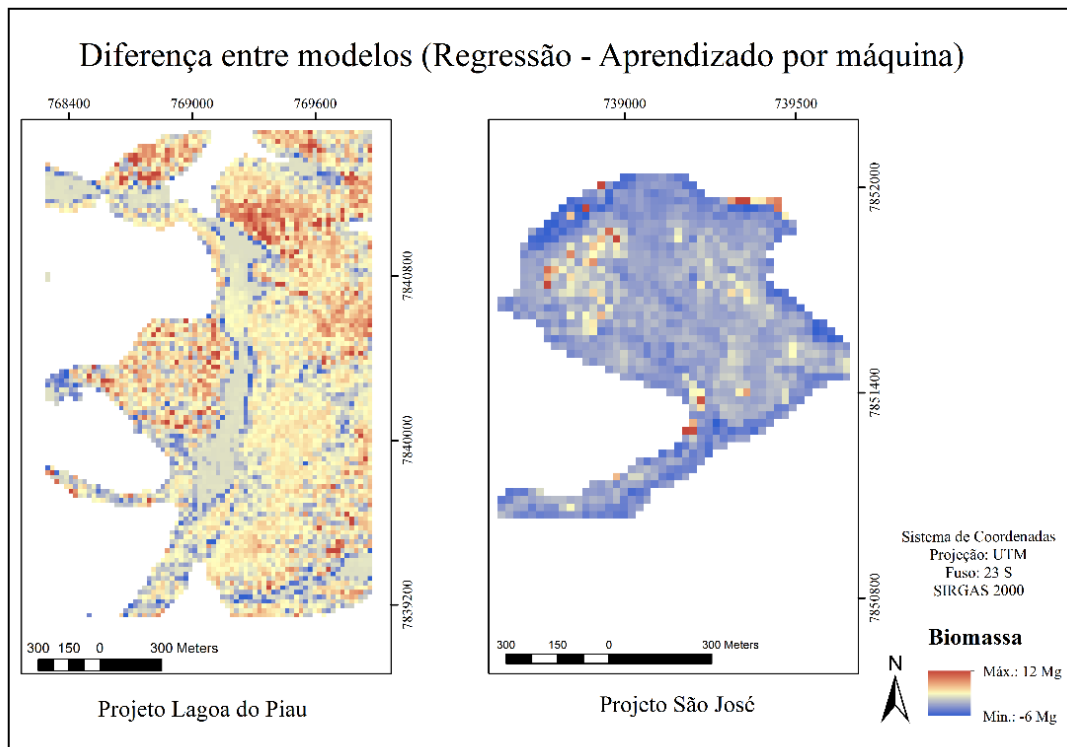


FIGURA 21: Diferença das estimativas de biomassa acima do solo (Mg) entre os modelos ajustados destinados aos projetos da empresa CENIBRA S.A., localizada a leste do estado de Minas Gerais, Brasil.

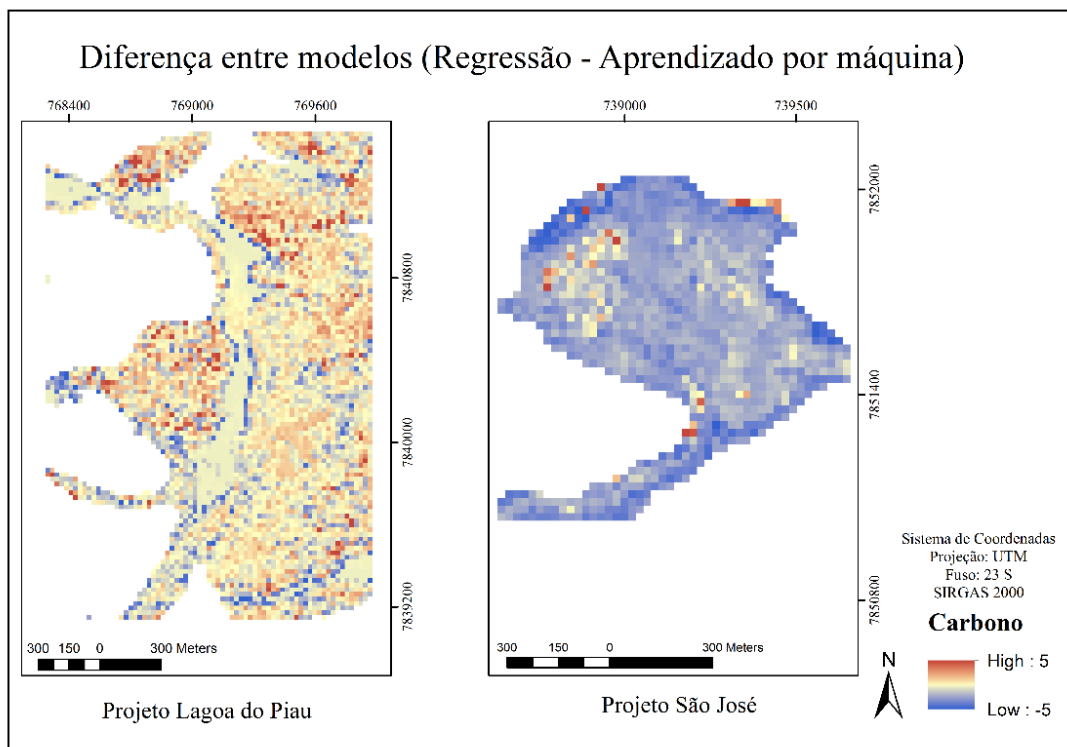


FIGURA 22: Diferença das estimativas de volume carbono acima do solo (Mg) entre os modelos ajustados destinados aos projetos da empresa CENIBRA S.A., localizada a leste do estado de Minas Gerais, Brasil.

Os valores das estimativas por hectare nos três atributos biofísicos tiveram valores variando para superestimação e subestimação quando tem por base o valor da estimativa em campo. Para o projeto Lago do Piau, o método por *Random Forest* apresentou superestimação em todos os atributos. No projeto São José, apenas o volume estimado pelo *Random Forest* apresentou um valor superior dos demais, enquanto que para biomassa e carbono nos dois métodos não houve diferença significativa (TABELA 11).

TABELA 11: Estimativa dos parâmetros biofísicos para toda a área dos projetos São José e Lagoa do Piau, pertencente a empresa CENIBRA S.A. localizada a leste do estado de Minas Gerais, Brasil. Exp.: Equação Exponencial; RF: *Random Forest* e; campo: parcelas de campo

Atributos Biofísicos	Projeto Lagoa do Piau			Projeto São José		
	Campo	Exp.	RF	Campo	Exp.	RF
Volume (m ³ . ha ⁻¹)	122,07	117,99	172,19	332,32	335,44	413,06
Biomassa acima do solo (Mg/ha)	79,55	80,62	86,99	250,12	237,04	210,94
Carbono acima do solo (Mg/ha)	38,18	38,69	43,93	120,06	113,78	99,80

6. DISCUSSÃO

As estimativas de campo da biomassa acima do solo calculada a partir de volume obtiveram valores semelhantes com trabalhos desenvolvidos em Floresta Estacional Semidecidual (FES). Souza et al. (2011) trabalhando para a mesma área de estudo (região Vale do Rio Doce) encontrou valores de biomassa acima do solo variando entre 73,39 ton/ha a 152,35 ton/ha. Amaro (2010) encontrou valores de 197,73 ton/ha de biomassa acima do solo e raízes e 90,01 ton/ha para carbono acima do solo e raízes em um fragmento da FES localizado em Viçosa, Minas Gerais. No mesmo fragmento de

Viçosa, Figueiredo et al. (2015) encontraram uma média no estoque de carbono referente a 61,81 ton/ha.

O uso da tecnologia LiDAR possibilita a visualização estrutural das florestas, mesmo aquelas que apresentam vegetação densa, pois trata-se de dados a laser que conseguem capturar essas informações no interior da floresta através de aplicativos dos softwares e equações que geram essas informações. Como Gorgens et al. (2016) que caracterizaram e discriminaram diferentes florestas brasileiras através do perfilamento a laser, mostraram o comportamento de curva sigmoide das camadas relacionando com as nuvens de pontos, assim como observado nesse estudo.

Com o aumento na amplitude das estimativas em alturas no decorrer dos percentis, ficou demonstrado que, para o percentil de elevação (H99) existe uma variabilidade estrutural em relação as alturas máximas de cada parcela.

O gráfico da distribuição da altura pelo seu percentil com sua imagem observa-se na curva que quando essa apresenta uma inflexão acentuada ocorre predominância de indivíduos altos e baixos. Porém, quando a parcela apresenta homogeneidade nas alturas dos indivíduos, a curva tende a ser mais retilínea nos percentis medianos, sem a ocorrência de alturas discrepantes entre os indivíduos.

A métrica que mais correlacionou com os parâmetros florestais biofísicos foi a da estimativa de altura no percentil 40. Quando essas estimativas da altura no percentil 40 foram plotadas em relação a cada atributo biofísico, a curva de tendência que mais aproximou dos dados foi a exponencial em que obteve coeficiente de determinação $R^2 > 0,73$ para volume e $R^2 > 0,77$ para biomassa acima do solo e carbono acima do solo, em que o uso de métricas LiDAR sozinhas conseguem prever satisfatoriamente variáveis de povoamento florestal (XU et al., 2018).

No ajuste da equação do modelo exponencial a partir das parcelas de treinamento foram obtidas medidas estatísticas com $R^2 > 0,88$ para todos os atributos biofísicos, obtendo valores médios de biomassa acima do solo referente a 50 Mg para estimativas via regressão e de 82 Mg/ha via aprendizado por máquinas para fragmento de Mata Atlântica.

Cao et al. (2016) estimaram a biomassa acima do solo utilizando regressão múltipla para uma floresta subtropical no sudeste da China, obtendo equações de biomassa com R^2 variando entre 0,74 a 0,79, com RMSE de 16,94 a 15,21% a relação entre biomassa

estimada em campo com a biomassa estimada do LiDAR para os anos de 2007 e 2013, respectivamente.

Yang et al. (2018) estimaram o avanço da seca na região amazônica utilizando dados LiDAR orbital GLAS (*Geoscience Laser Altimeter System*) e imagens de satélite, e através de equações alométricas e a métrica que tiveram o melhor desempenho foram a altura do percentil 30 e 90 pois explicaram com maior precisão o comportamento padrão da estrutura florestal. Já Réjou-Méchain et al. (2015) estudando a variação da biomassa em florestas Neotropicais obtiveram como variável independente a altura estimada no percentil 50 em que geraram modelo com o ajuste da equação. Utilizando regressão linear linearizada, Means et al. (2000) modelaram volume para uma floresta em Oregon, em que a altura no percentil 20 e 80, juntamente com a densidade de pontos na copa foram as variáveis independentes do modelo.

As medidas de validação para o ajuste da equação do modelo exponencial considerando apenas uma variável independente ao modelo obtiveram $R^2 > 0,91$ para volume, biomassa acima do solo e carbono acima do solo. Isso mostra que a métrica referente a altura da vegetação tem forte correlação com os atributos biofísicos, corroborando com a realidade biológica, uma vez que para ajuste de modelos volumétricos a altura é uma das variáveis independentes.

Os dados de coeficiente de determinação para o treinamento quanto para a validação foram satisfatórios, uma vez que Means et al. (2000) encontraram valores de R^2 de 0,95 a 0,97 para volume, usando métricas de altura no percentil 80 e densidade de copa como variáveis independentes em florestas no Oregon.

Shi et al. (2018) utilizaram o algoritmo *Random Forest* como separador em amostras de treinamento e de validação para posteriormente realizar a modelagem de parâmetros florestais. Chen et al. (2011), por sua vez, modelaram através de regressão linear a biomassa e o carbono empregando regressão dos dados para treinamento e 10% para validação com coeficiente de determinação superior a 80%.

Dentre as sete métricas selecionadas para o modelo *Random Forest*, de maior valor de importância foi a moda da elevação, o que explica a maior repetibilidade e permanência dos pontos em sua plotagem e distribuição. Assim a maior quantidade de material florestal é interceptada pelo laser concentrando nessa região a maior quantidade de pontos. A elevação máxima foi outra métrica selecionada muito importante para o ajuste do modelo ótimo, pois ela retorna o ponto mais alto da

floresta. A altura estimada no percentil 1, também foi uma métrica selecionada para compor o modelo, que demonstra a presença de sub-bosque na floresta, pois como trata-se de floresta inequidâneas, considera a presença de muito ou pouco indivíduos da regeneração florestal.

As métricas referentes a elevação dos momentos 3 e 4 juntamente com a métrica interquartil retratam a proporção de pontos de uma camada para outra na nuvem de pontos. E a métrica selecionada referente a porcentagem entre o primeiro retorno abaixo de 2,00 m retrata uma medida proporcional dos pontos em relação aos seus pontos de retorno. Uma vez que as métricas provenientes da nuvem de pontos LiDAR que calculam a porcentagem e a relação entre os retornos e a sua proporção entre elas são importantes para descrever a distribuição vertical da floresta (PALACE et al., 2015; PEARSE et al., 2017; XU et al., 2018).

Utilizando regressão linear múltipla, Xu et al. (2018) utilizaram as métricas derivadas do LiDAR para compor sua modelagem, em que as candidatas para o modelo foram: 4th retorno acima de 0,5 metros; Hqav (média quadrática da altura); I_L4; ((todos os retornos acima de 3 metros) / (todos os primeiros retornos)) *100 para estimar o volume em parcelas de povoamentos florestais na Nova Zelândia.

As estimativas dos atributos biofísicos com o modelo *Random Forest* (RF) apresentaram $R^2 > 0,95$, Ghosh et al. (2018) utilizaram árvores de decisão (*Random Forest*) para estimativas da biomassa em florestas tropicais densas do nordeste da Índia apresentando coeficiente de determinação superior a 71%. Mostrando que o algoritmo *Random Forest* aliado a dados LiDAR prediz o volume considerando o coeficiente de determinação de 0,83 (XU et al., 2018).

As estimativas oriundas dos mapeamentos apresentaram valores mais elevados com o uso da regressão não linear, porém nos mapas observa-se que em áreas que ocorrem maiores valores com a regressão também ocorrem valores altos com o *Random Forest*. Assim, espacialmente é possível observar que esses dois métodos estimam na mesma proporção, porém com valores distintos. Esse comportamento é evidente no projeto Lagoa do Piau para os três atributos biofísicos estimados. Já no projeto São José, os modelos não seguem uma padronização da estimativa. Nesse projeto, o *Random Forest* tende a superestimar os valores do que a estimativa por regressão. Isso pode estar associado a menor densidade amostral oriunda dessa área. Longo et al. (2016), por sua vez, estimou carbono acima do solo em florestas intactas e degradadas da Amazônia

utilizando métricas LiDAR com modelos de regressão do *Random Forest* mostrando ser promissora essa tecnologia para geração de estimativas em larga escala. De maneira geral, os métodos são semelhantes, porém o Random Forest apresenta estimativas superiores de forma geral e a equação exponencial de forma pontual.

Para validar as estimativas do mapeamento foi calculado o Intervalo de Confiança (IC) da média em que os valores estimados via mapeamento se enquadraram dentro dos limites do IC. Gerou um mapa da diferença entre os modelos demonstrando o quanto similares suas estimativas. No projeto Lagoa do Piau, as estimativas foram semelhantes entre eles por apresentarem valores próximos de zero. Para o projeto São José, o RF estimou valores mais altos do que a regressão, como já observado no mapa das estimativas de modelos.

A comparação estatística entre os modelos foi realizada considerando o teste F de Graybill, que foi comprovado pelo p-valor ($< 0,001$) que os modelos não têm igualdade entre eles.

Hiil et al. (2014) criaram mapas de volume para a região oeste da Suíça, separados em classes de predição pela acurácia obtendo um coeficiente de determinação (R^2) de 64% utilizando regressão linear simples e maior entendimento das áreas estudadas, enquanto que nesse estudo o R^2 foi superior a 97% de predição do volume no mapeamento utilizando regressão não linear de modelo exponencial e o uso de aprendizado por máquina (*Random Forest*). Yao et al (2018) mapearam as florestas chinesas através do armazenamento da biomassa florestal representado pelo carbono nela inserido através de um modelo semi-empírico presente nas parcelas, considerando o R^2 e o RMSE como critério estatístico para selecionar o melhor modelo. Corroborando a esse estudo, as estimativas do mapeamento para esse trabalho atenderam aos limites atribuídos ao Intervalo de Confiança destinado as parcelas de validação.

Assim, de forma geral, foi calculado o estoque total em hectare das estimativas dos atributos biofísicos (volume, biomassa acima do solo e carbono acima do solo) para toda a área dos projetos da empresa CENIBRA S.A. utilizando dados de regressão não linear e aprendizado por máquina aliados aos dados de sensoriamento remoto em consonância com os dados provenientes de campo.

7. CONCLUSÃO

Este estudo permite concluir que a descrição da estrutura e a estimação dos atributos biofísicos da Floresta Estacional Semidecidual podem ser feitas com eficiência utilizando dados obtidos com LiDAR aerotransportado. Apenas uma métrica é suficiente para estimar o volume, a biomassa e o carbono com o emprego de regressão exponencial. O estudo permite concluir, também, que poucas métricas são necessárias para emprego do *Random Forest*. Os níveis de precisão obtidos ao empregar a regressão e o *Random Forest* podem ser considerados como satisfatórios, com menor viés ao utilizar a regressão.

8. REFERÊNCIAS BIBLIOGRÁFICAS

AMARO, M.A. **Quantificação do estoque volumétrico, de biomassa e de carbono em uma Floresta Estacional Semidecidual no município de Viçosa – MG**. Tese (Doutorado em Ciência Florestal). Universidade Federal de Viçosa, MG. 2010. 183 p.

BABCOCK, C.; FINLEY, A. O.; COOK, B. D.; WEISKITTED, A.; WOODALL, C. W. Modeling forest biomass and growth: Coupling long-term inventory and LiDAR data. **Remote Sensing of Environment**, 182, p. 1 – 12, 2016.

BECKNELL, J.M.; KELLER, M.; PIOTTO, D.; LONGO, M.; dos-SANTOS, M.N.; SCARANELLO, M.A.; CAVALCANTE, R.B.O.; PORDER, S. Landscape-scale lidar analysis of aboveground biomass distribution in secondary Brazilian Atlantic Forest. **BIOTROPICA**, v.0, doi: 10.1111/btp.12538, p 1 – 11, 2018.

CAO, L.; COOPS, N. C.; INNES, J. L.; SHEPPARD, S. R. J.; FU, L.; RUAN, H.; SHE, G. Estimation os forest biomass dynamics in subtropical forests using multi-temporal airborne LiDAR data. **Remote Sensing of Environment**, 178, p 158 – 171, 2016.

CERRI, C.E.P.; SPAROVEK,G.; BERNOUX, M.; EASTERLING, W.E.; MELILLO, J.M.; CERRI, C.C. Tropical agriculture and global warming: impacts and mitigation options. **Scientia Agricola**, v.64, n.1, p 83 – 99, 2007.

CHAVE, J.; ANDALO, C.; BROWN, S.; CAIRNS,M.A.; CHAMBERS, J.Q.; EAMUS, E.; FOLSTER, H.; FROMARD, F.; HIGUCHI, N.; KIRA, T.; LESCURE, J.P.; NELSON, B.W.; OGAWA, H.; PUIG, H.; PIERA, B.; YAMAKURA, T. Tree

allometry and improved estimation of carbon stocks and balance in tropical forests. **Oecologia**, v. 145, p 87 – 99, 2005.

CHEN, X.; LIU, S.; ZHU, Z.; VOGELMANN, J.; LI, Z.; OHLEN, D. Estimating aboveground forest biomass carbon and fire consumption in the U.S. Utah High Plateaus using data from the Forest Inventory and Analysis Program, Landsat, and LANDFIRE. **Ecological Indicators**, v. 11, p 140 – 148, 2011.

CONNEL, J.H. On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. Advanced Study Institute on Dynamics of numbers in populations. **Proceedings**. Oosterbeek, Centre for Agricultural Publication and Documentation - The Netherlands, pag 298 – 312, Setembro de 1970.

FAO – Organização das Nações Unidas para Alimentação e Agricultura. **Voluntary Guidelines on National Forest Monitoring**, Roma, Itália, 2017. Acesso em 02 de dezembro de 2018: < <http://www.fao.org/3/a-I6767e.pdf>>.

FENG, Y.; LU, D.; CHEN, Q.; KELLER, M.; MORAN, E.; DOS-SANTOS, M.N.; BOLFE, E.L.; BATISTELAA, M. 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**, v. 10, n. 10, p 996 - 1016, 2017.

FIGUEIREDO, L. T. M.; SOARES, C.P.B.; SOUSA, A.L.; LEITE, H.G.; SILVA, G.F. Carbon stock dynamics of tree stems in a tropical semideciduous forest. **CERNE**, v. 21, n. 1, p 161–167, mar. 2015.

FIGUEIREDO, E.O.; d'OLIVEIRA, M.V.N.; BRAZ, E.M.; PAPA, D.A.; FEARNSSIDE, P.M. LIDAR-based estimation of bole biomass for precision management of an Amazonian forest: Comparisons of ground-based and remotely sensed estimates. **Remote Sensing of Environment**, v. 187, p 281 – 293, 2016.

FILHO, E.I.F. Collection of functions to fit models with emphasis in land use and soil mapping. **Labgeo R package version 0.3.8.2**. n.d. Acesso em: <<http://github.com/elpidiofilho/labgeo>>, 2018.

FUNDAÇÃO INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Manual técnico da vegetação brasileira**. Rio de Janeiro: Fundação Instituto Brasileiro de Geografia e Estatística, 62 p. (Série: Manuais técnicos em geociências; n.2), 1992.

FUNDAÇÃO SOS MATA ATLÂNTICA. **Atlas dos remanescentes florestais da Mata Atlântica: Período 2012 - 2013**. Disponível em: < https://www.sosma.org.br/wp-content/uploads/2014/05/atlas_2012-2013_relatorio_tecnico_20141.pdf> 2014. Acesso em outubro de 2018

GARCÍA M.; SAATCHI S.; USTIN S.; BALZTER, H. Modelling forest canopy height by integrating airborne LiDAR samples with satellite Radar and multispectral imagery. **International Journal Applied Earth Observation Geoinformation**. v. 66 p. 159 – 173, 2018.

GARCIA-GUTIERREZ, J.; GONZALEZ-FERREIRO, E.; RIQUELME-SANTOS, J.C.; MIRANDA, D.; DEGUEZ-ARANDA, U.; NAVARRO-CERRILLO, R. Evolutionary feature selection to estimate forest stand variables using LiDAR. **International Journal of Applied Earth Observation and Geoinformation**, v. 26, p. 119–131, 2014.

GHOSH, S.M.; BEHERA, M.D. Aboveground biomass estimation using multi-sensor data synergy and Machine learning algorithms in a dense tropical forest. **Applied Geography**, v.96, p. 29 – 40, 2018.

GIONGO, M.; KOEHLER, H.S.; MACHADO, S.A.; KIRCHNER, F.F.; MARCHETTI, M. LiDAR: princípios e aplicações florestais. **Pesquisa Florestal Brasileira**, Colombo, v. 30, n.63, p. 231 – 244, 2010.

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

GORGENS, E. B.; SILVA, A. G. P.; RODRIGUEZ, L. C. E. **LIDAR: Aplicações florestais**. 1. ed - Curitiba, PR: Editora CRV, 2014. 132p.

GORGENS, E.B.; SOARES, C.P.B.; NUNES, M.H.; RODRIGUEZ, L.C.E. Characterization of Brazilian forestas types utilizing canopy height profiles derived from airborne laser scanning. **Applied Vegetation Science**. Doi: 10.1111/avsc.12224, 2016.

GRAYBILL, F. A. **Theory and application of the linear model**. Belmont: Duxbury Press, 1976. 704p.

HAYKIN, S. **Neural Networks – A Comprehensive Foundation**. Prentice – Hall, New Jersey, 2 ed. 1999.

HILL, A.; BRESCHAN, J.; MANDALLAZ, D. Accuracy Assessment of Timber Volume Maps Using Forest Inventory Data and LiDAR canopy height models. **Forests**, v. 5, p. 2253 – 2275, 2014.

HUDAK, A.T.; STRAND, E.K.; VIERLING, L.A.; BYRNE, J.C.; EITEL, J.U.H.; MARTINUZZI, S.; FALKOWSKI, M.J. Quantifying aboveground forest carbono pools and fluxes from repeat LiDAR surveys. **Remote Sensing of Environment**, v. 123, p. 25 – 40, 2012.

HIGUCHI, N.; CARVALHO JR.; J.A. Fitomassa, conteúdo e carbono de espécies arbóreas da Amazônia. In: SEMINÁRIO EMISSÃO X SEQÜESTRO DE CO₂ – Uma nova oportunidade de negócios para o Brasil, 1994, Rio de Janeiro. **Anais...** Rio de Janeiro: Companhia Vale do Rio Doce, 1994.

HUMAGAIN, K.; PORTILLO-QUINTERO, C.; COX, R.D.; CAIN, J.W. Estimating forest canopy cover dynamics in Valles Caldera National Preserve, New Mexico, using LiDAR and Landsat data. **Applied Geography**, v. 99, p. 120 – 132, 2018.

HUXLEY, J.S.; TEISSER, G. Terminology of relative growth. **Nature**. v. 137, p. 780-781, 1936.

IM, J.; LU, Z.; RHEE, J.; QUACKENBUSH, L.J. Impervious surface quantification using a synthesis of artificial immune networks and decision/regression trees from multi-sensor data. **Remote Sensing of Environment**, v. 117, p. 102 – 113, 2012.

INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE. **Climate change 2007: Mitigation of Climate Change** – Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. 2007

JANSEN, D.H. Herbivores and the number of tree species in tropical forests. **The American Naturalist**, v. 104, n. 940, p. 501 – 528, 1970.

JENSEN, J.R. **Sensoriamento Remoto do ambiente: uma perspectiva em recursos terrestres**. 2 ed. São José dos Campos: Parêntese, 2009. 598 p.

KETTERINGS, Q.M.; COE, R.; NOORDWIJK, M.V.; AMBAGAU, Y.; PALM, C.A. Reducing uncertainty in the use of allometric biomass equations for predicting above-ground tree biomass in mixed secondary forests. **Forest Ecology and Management**, v. 146, p. 199 – 209, 2001.

KUHN, M.; WING, J.; WESTON, S.; WILLIAMS, A.; KEEFER, C.; ENGELHARDT, A.; COOPER, T.; MAYER, Z.; KENKEL, B.; BENESTY, M.; LESCARBEAU, R.; ZIEM, A.; SCRUCICA, L.; TANG, Y. CANDAN, C.; HUNT, T. **Caret: Classification and Regression Training**. R package version 6.0-81., 2018.

LASTools. Disponível em: <https://rapidlasso.com/lastools/> Acesso em: junho de 2018.

LAURIN, G. V.; PULETTI, N.; CHEN, Q.; CORONA, P.; PAPALE, D.; VALENTINI, R. Aboveground biomass and tree species richness estimation with airborne lidar in tropical Ghana forests. **International Journal of Applied Earth Observation and Geoinformation**, 52, p. 371 – 379. 2016.

LEFSKY, M.A.; COHEN, W.B.; PARKER, G.G.; HARDING, D.J. LiDAR remote sensing for ecosystem studies. **Bioscience**, v. 52, n.1, p. 19 – 30, 2002.

LEITOLD, V.; KELLER, M.; MORTON, D.C.; COOK, B.D.; SHIMABUKURO, Y.E. Airborne lidar-based estimates of tropical forest structure in complex terrain: opportunities and trade-offs for REDD+. **Carbon Balance and Management**. DOI 10.1186/s13021-015-0013-x. 2015.

LI, M.; IM, J.; QUACKENBUSH, L.J.; LIU, T. Forest Biomass and Carbon Stock Quantification Using Airborne LiDAR Data : A Case Study Over Huntington Wildlife Forest in the Adirondack Park. **IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING**, v. 7, n. 7, p. 3143–3156, 2014

LILLESAND, T.; KIEFER, R.W.; CHIPMAN, J. Remote Sensing and Image Interpretation. **7 ed. Wiley**, 2015. 768p.

LONGO, M.; KELLER, M.; DOS-SANTOS, M. N.; LEITOLD, V.; PINAGÉ, E. R.; BACCINI, A.; SAATCHI, S.; NOGUEIRA, E. M.; BATISTELLA, M.; MORTON, D. C. Aboveground biomass variability across intact and degraded forests in the Brazilian Amazon. **Global Biogeochemical Cycles**, v. 30, n. 11, p. 1639 – 1660, 2016.

LU, D.; CHEN, Q.; WANG, G.; LIU, L.; MORAN, E. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. **International Journal of Digital Earth**, v. 9, n. 1, p. 63–105, 2016.

LUO, S.; WANG, C.; XI, X.; PAN, F.; PENG, D.; ZOU, J.; NIE, S.; QIN, H. Fusion of airborne LiDAR data hyperspectral imagery for aboveground and belowground forest biomass estimation. **Ecological Indicators**, v. 73, p. 378 – 387. 2017.

MALTAMO, M.; MEHTÄTALO, L.; VAUHKONEN, J.; PACKALÉN, P. Predicting and calibrating tree attributes by means of airborne laser scanning and field measurements. **Canadian Journal of Forests Research**, Toronto, v. 43, p. 1896 – 1907, 2012.

McGAUGHEY, R. J. **FUSION/LDV**: Software for Lidar Data Analysis and Visualization. Washington, DC: United States Department of Agriculture. Forest Service. Pacific Northwest Research Station, 2018. 214p. Disponível em: <<http://forsys.cfr.washington.edu/fusion/fusionlatest.html>>. Acesso em: maio 2018.

MEANS, J.E.; ACKER, S.A.; FITT, B.J.; RENSLOW, M.; EMERSON, L.; ABSTRACT, C.J.H. Predicting Forest Stand Characteristics with Airborne Scanning LiDAR. **Photogrammetric Engineering & Remote Sensing**, v.66, n. 11, p. 1367 – 1371, 2000.

MÉCHAIN, M.R.; TYMEN, B.; BLANC, L.; FAUSET, S.; FELDPAUSCH, T.R.; MONTEAGUDO, A; PHILLIPS, O.L.; RICHARD, H.; CHAVE, J. Using repeated

small-footprint LiDAR acquisitions to infer spatial and temporal variations of a high-biomass Neotropical forest. **Remote Sensing of Environment**, v. 169, p. 93 – 101, 2015.

MELO, E.A.S.C. Desafios e oportunidades para a Silvicultura de Precisão: uma síntese do congresso brasileiro de agricultura de precisão de 2014. 50ª reunião Técnico-Científico do Programa Cooperativo sobre Silvicultura e Manejo. **Anais da Série Técnica**. IPEF, v. 24, n. 45, 2016.

MILENKOVIC, M.; SCHNELL, S.; HOLMGREN, J.; RESSL, C.; LINDBERG, E.; HOLLAUS, M.; PFEIFER, N.; OLSSON, H. Influence of footprint size and geolocation error on the precision of forest biomass estimates from space-borne waveform LiDAR. **Remote Sensing of Environment**, 200, p. 74 – 88. 2017

MONTAÑO, R.A.N.R. **Aplicação de Técnicas de Aprendizado de Máquinas na Mensuração Florestal**. Tese de Doutorado em Ciências Exatas na Universidade Federal do Paraná. 102p. Curitiba, 2016.

MYERS, N.; MITTERMEIER, R.A.; MITTERMEIER, C.G.; FONSECA, G.A.B. da; KENT, J. Biodiversity hotspot for conservation priorities. **Nature**, v.403, p. 853 – 858, 2000.

NÆSSET, E. Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in Scandinavia. **Scandinavian Journal of Forest Research**, v. 22, n.5, p. 433 – 442, 2007.

NÆSSET, E., E GOBAKKEN, T. Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. **Remote Sensing of Environment**, v. 112, p. 3079–3090,2008.

NEVES, D.M; DEXTER, K.G.; PENNINGTON, R.T; VALENTE, A.S.M.; BUENO, M.L.; EISENLOHR, P.V.; FONTES, M.A.L.; MIRANDA, P.L.S.; MOREIRA, S.N.; REZENDE, V.L.; SAITER, F.Z.; OLIVEIRA-FILHO, A.T. Dissecting a biodiversity hotspot: The importance of environmentally marginal habitats in the Atlantic Forest Domain of South America. **Diversity and Distributions**, v. 23, p. 898–909, 2017.

PALACE, M.W.; SULLIVAN, F.B.; DUCEY, M.J.; TREUHAFT, R.N.; HERRICK, C.; SHIMBO, J.Z.; MOTA-E-SILVA, J. Estimating forest structure in a tropical forest using field measurements, a synthetic model and discrete return lidar data. **Remote Sensing of Environment**, v. 161, p. 1 – 11, 2015.

PEARSE, G.D.; MORGENROTH, J.; WATT, M.S.; DASH, J.P. Optimising prediction of forest leaf area index from discrete airborne lidar. **Remote Sensing of Environment**, v.200, p. 220 – 239, 2017.

PICARD, N.; GAMARRA, J.G.P.; BIRIGAZZI, L.; BRANTHOMME, A. Plot-level variability in biomass for tropical forest inventory designs. **Forest Ecology and Management**, v. 430, p. 10 – 20, 2018.

POWELL, S.L.; COHEN, W.B.; HEALEY, S.P.; KENNEDY, R.E.; MOISEN, G.G.; PIERCE, K.B. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. **Remote Sensing of Environment**, v.114, p. 1053 – 1068, 2010.

R Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL <<https://www.R-project.org/>>, 2017.

SAATCHI, S.S.; HARRUS, N.L.; BROWN, S.; LEFSKY, M.; MITCHARD, E.T.A.; SALAS, W.; ZUTTA, B.R.; BUERMANN, W.; LEWIS, S.L.; HAGEN, S.; PETROVA, S.; WHITE, L.; SILMAN, M.; MOREL, A. Benchmark map of forest carbon stocks in tropical regions across three continents. **Environmental Science**, v.108, n. 24, p. 9899 – 9904, 2011.

SABOL, J.; PROCHÁZKA, D.; PATOCKA, Z. Development of models for forest variable estimations from airborne laser scanning data using an area-based approach at a plot level. **Journal of Forest Science**, v. 62, p. 137 – 142, 2016.

SALVIN, M. **Multicolinearidade**. Universidade de São Paulo – Escola Superior de Agricultura “Luiz de Queiroz”. Piracicaba – SP. 2016

SCHNEIDER, P. R. **Análise de regressão aplicada à engenharia florestal**. Santa Maria: UFSM, 1997. 217p.

SHI, Y.; WANG, T.; SKIDMORE, A.K.; HEURICH, M. Important LiDAR metrics for discriminating forest tree species in Central Europe. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 137, p. 163 – 174, 2018.

SILVEIRA, P.; KOEHLER, H. S.; SANQUETTA, C. R.; ARCE, J. E. O estado da arte na estimativa de biomassa e carbono em formações florestais. **Revista Floresta**, v. 38, n.1, p. 185 – 206. 2008.

SILVA, C. A. **Carbono na parte aérea de plantios de Eucalyptus spp. – em nível de árvore por amostragem destrutiva e para talhões inteiros após o ajuste de métricas LiDAR**. 2013. Dissertação (Mestrado em Ciência: Recursos Florestais) – Universidade de São Paulo, Escola Superior de Agricultura “Luiz de Queiroz”. Piracicaba, São Paulo, 2013, 153p.

SOARES, C. P. B.; NETO, F. P.; SOUZA, A. L. **Dendrometria e Inventário Florestal**. 2 ed. Viçosa: Editora UFV, 2011.

SOUTO, M.C.P.; LORENA, A.C.; DELBEM, A.C.B.; CARVALHO, A.C.P.L.F. Técnicas de Aprendizado de Máquinas para problemas de Biologia Molecular. **Anais XXIII Congresso da Sociedade Brasileira de Computação** – Minicurso de Inteligência Artificial, Jornada de Atualização Científica em Inteligência Artificial, p. 103 – 152, 2003.

SOUZA, A.L.; BOINA, A.; SOARES, C.P.B.; VITAL, B.R.; GASPAR, R.O.; LANA, J.M. Estoque e crescimento em volume, biomassa, carbon e dióxido de carbono em Floresta Estacional Semidecidual. **Revista Árvore**, v.35, n.6, p. 1277 – 1285, 2011.

VIERLING, L.A.; MARTINUZZI, S.; ASNER, G.P.; STOKER, J.; JOHNSON, B.R. LiDAR: providing structure. **Frontiers in Ecology and the Environment**, Washington, v.9, p. 261 – 262, 2011.

YANG, Y.; SAATCHI, S.S.; XU, L.; YU, Y.; CHOI, S.; PHILLIPS, N.; KENNEDY, R.; KELLER, M.; KNYAZIKHIN, Y.; MYNENI, R.B. Post-drought decline of the Amazon carbon sink. **Nature Communications**, 9:3172, doi: 10.1038/s41467-018-05668-6, 2018.

YAO, Y.; PIAO, S.; WANG, T. Future biomass carbono sequestration capacity of Chinese forests. **Science Bulletin**, v. 63, p. 1108 – 1117, 2018.

YU, C.M. **Sequestro florestal do carbono no Brasil – Dimensões políticas, socioeconômicas e ecológicas**. Tese (Doutorado em Meio Ambiente e Desenvolvimento). Universidade Federal do Paraná, PR. 2004 279 p.

XIANG, W.H.; ZHOU, J.; OUYANG, S.; ZHANG, S.L.; LEI, P.F.; LI, J.X.; DENG, X.W.; FANG, X.; FORRESTER, D.I. Species-specific and general allometric equations for estimating tree biomass components of subtropical forests in southern China. **European Journal of Forest Research**, v. 135, p. 963 – 979, 2016.

XU, C.; MANLEY, B.; MORGENROTH, J. Evaluation of modelling approaches in predicting forest volume and stand age for small-scale plantation forestes in New Zealand with RapidEye and LiDAR. **International Journal Applied Earth Observation Geoinformation**, v. 73, p. 386 – 396, 2018.

WALTON, J.T. Subpixel urban land cover estimation: Comparing cubista, Random forests, and support vector regression. **Photogrammetric Engineering and Remote Sensing**, v.74, p. 1213 – 1222, 2008.

WATZLAWICK, L. F. **Estimativa de biomassa e carbono em floresta ombrófila mista e plantações florestais a partir de dados de imagens do satélite Ikonos II**. 2003. 138 f. Tese (Doutorado em Engenharia Florestal) - Universidade Federal do Paraná, Curitiba, 2003.

WENDLING, J. L. Ç. **Modelos matemáticos de crescimento e de produção em biomassa para árvores de *Euterpe edulis* MART. plantadas a céu aberto.** 1998. 185 f. Dissertação (Mestrado em Engenharia Florestal) - Universidade Federal do Paraná, Curitiba, 1998.

ZHANG, Z.; CAO, L.; SHE, G. Estimating forest structural parameters using Canopy metrics derived from airborne LiDAR data subtropical forests. **Remote Sensing**, v.9, n.9, p. 940 - 966, 2017.

ZOLKOS, S.G.; GOETZ, S.J.; DUBAYAH, R. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. **Remote Sensing Environment**, v.128, p. 289–298 2013.