

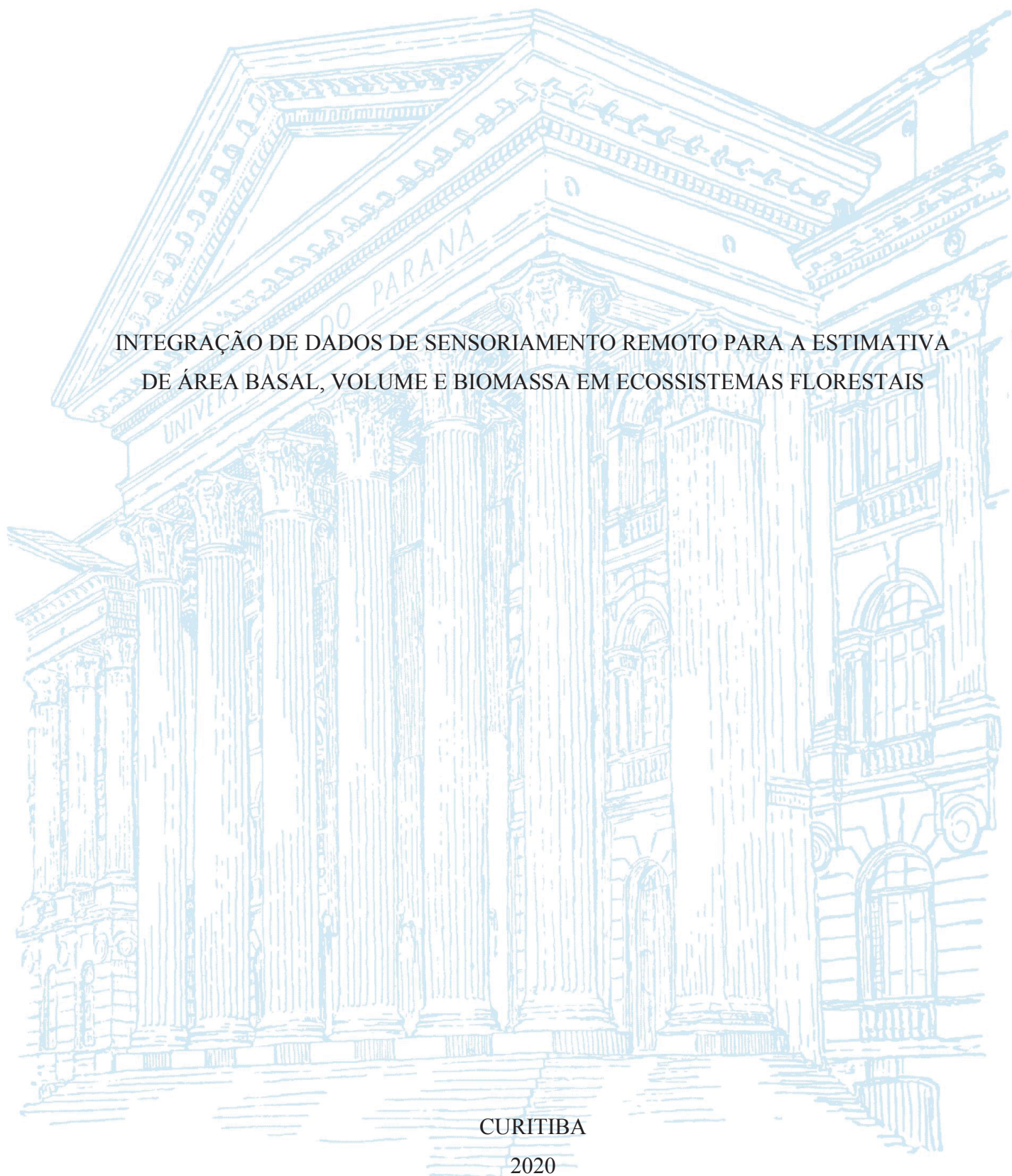
UNIVERSIDADE FEDERAL DO PARANÁ

ALINE BERNARDA DEBASTIANI

INTEGRAÇÃO DE DADOS DE SENSORIAMENTO REMOTO PARA A ESTIMATIVA
DE ÁREA BASAL, VOLUME E BIOMASSA EM ECOSISTEMAS FLORESTAIS

CURITIBA

2020



ALINE BERNARDA DEBASTIANI

INTEGRAÇÃO DE DADOS DE SENSORIAMENTO REMOTO PARA A ESTIMATIVA
DE ÁREA BASAL, VOLUME E BIOMASSA EM ECOSISTEMAS FLORESTAIS

Tese apresentada ao curso de Pós-Graduação em Engenharia Florestal, Área de Concentração em Manejo Florestal, Departamento de Ciências Florestais, Setor de Ciências Agrárias, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Doutora em Engenharia Florestal.

Orientador: Prof. Dr. Carlos Roberto Sanquetta

Coorientadoras: Prof (a). Dr (a). Ana Paula Dalla Corte
Dra. Naiara Pinto

CURITIBA

2020

Ficha catalográfica elaborada pela
Biblioteca de Ciências Florestais e da Madeira - UFPR

Debastiani, Aline Bernarda

Integração de dados de sensoriamento remoto para a estimativa de área basal, volume e biomassa em ecossistemas florestais / Aline Bernarda Debastiani. - Curitiba, 2020.

125 f. : il.

Orientador: Prof. Dr. Carlos Roberto Sanquetta

Coorientadoras: Profa. Dra. Ana Paula Dalla Corte

Dra. Naiara Pinto

Tese (Doutorado) - Universidade Federal do Paraná, Setor de Ciências Agrárias, Programa de Pós-Graduação em Engenharia Florestal. Defesa: Curitiba, 03/04/2020. Área de concentração: Manejo Florestal.

1. Sensoriamento remoto. 2. Dendrometria. 3. Biomassa florestal. 4. Eucalipto. 5. Satélites artificiais em sensoriamento remoto. 6. Aprendizado do computador. 7. Teses. I. Sanquetta, Carlos Roberto. II. Dalla Corte, Ana Paula. III. Pinto, Naiara. IV. Universidade Federal do Paraná, Setor de Ciências Agrárias. V. Título.

CDD – 634.9

CDU – 528.8

Bibliotecária: Berenice Rodrigues Ferreira – CRB 9/1160



MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIÊNCIAS AGRÁRIAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO ENGENHARIA
FLORESTAL - 40001016015P0

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ENGENHARIA FLORESTAL da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **ALINE BERNARDA DEBASTIANI** intitulada: **INTEGRAÇÃO DE DADOS DE SENSORIAMENTO REMOTO PARA A ESTIMATIVA DE ÁREA BASAL, VOLUME E BIOMASSA EM ECOSISTEMAS FLORESTAIS**, sob orientação do Prof. Dr. CARLOS ROBERTO SANQUETTA, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa. A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 03 de Abril de 2020.

Assinatura Eletrônica

18/05/2020 13:34:01.0

CARLOS ROBERTO SANQUETTA
Presidente da Banca Examinadora

Assinatura Eletrônica

18/05/2020 12:56:53.0

VERALDO LIESENBERG

Avaliador Externo (UNIVERSIDADE DO ESTADO DE SANTA
CATARINA)

Assinatura Eletrônica

18/05/2020 12:58:11.0

FABIO GUIMARÃES GONÇALVES

Avaliador Externo (CANOPY REMOTE SENSING SOLUTIONS LTDA.)

Assinatura Eletrônica

18/05/2020 11:34:53.0

JAIME WOJCIECHOWSKI

Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

18/05/2020 13:56:13.0

ALEXANDRE BEHLING

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Avenida Lothário Meissner, 632 - CURITIBA - Paraná - Brasil

CEP 80210-170 - Tel: (41) 3360-4212 - E-mail: pgfloresta@gmail.com

Documento assinado eletronicamente de acordo com o disposto na legislação federal Decreto 8539 de 08 de outubro de 2015.

Gerado e autenticado pelo SIGA-UFPR, com a seguinte identificação única: 41642

Para autenticar este documento/assinatura, acesse <https://www.prppg.ufpr.br/siga/visitante/autenticacaoassinaturas.jsp>
e insira o código 41642

AGRADECIMENTOS

Gostaria de prestar meus sinceros agradecimentos a todos que de alguma forma contribuíram para meu egresso no doutorado e para a realização da tese.

Aos meus pais Ilceu e Nelci, irmãos Douglas e Francine, estes que me apoiaram e deram suporte na minha escolha de ingressar na pós-graduação e morar longe.

Ao meu comitê de orientação, Carlos R. Sanquetta, Ana Paula Dalla Corte e Naiara Pinto, pelas contribuições e paciência durante todo esse processo de aprendizado, sem vocês não seria possível. Em especial ao meu orientador o qual tenho profunda admiração, desde meu primeiro contato, aceitou me orientar no doutorado, sempre com conselhos acadêmicos, profissionais e para a vida.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudos concedida.

Ao Programa de Pós-Graduação em Engenharia Florestal da Universidade Federal do Paraná pela oportunidade de ingressar no programa e desenvolver esta pesquisa. Aos professores do PPGEF por serem competente, inspiradores e compartilharem seus conhecimentos.

A empresa Klabin S.A. que quando achei que estava tudo perdido, se dispôs a realizar essa parceria, fomentando com dados, enviados em um curto período, isso possibilitou a realização do segundo capítulo. Que empresa hein, sensacional.

Aos inúmeros pesquisadores de radar os quais em algum momento precisei de auxílio e responderam meus e-mails e ligações, todos muito prestativos, usaram de seu tempo para auxiliar alguém que não conheciam. Alguns dos nomes: Fábio Gonçalves, Henrique Cassol, Fábio Gama, Juliano Sambatti, Rafael Rosa e outros funcionários da empresa Visiona.

Aos colegas do BIOFIX pelo companheirismo e troca de conhecimentos diariamente, sentirei saudades. Ao grupo de amigos que a pós-graduação me deu, “Entrevero”, companheiros dos rolês e tudo mais. A minha amiga da vida, Karina Guollo, que mesmo longe esteve sempre tão perto.

Eu não poderia de deixar de agradecer ao prof. Dr. Mosar Faria Botelho, foi o primeiro professor que acreditou em mim, instigou meu interesse para a pesquisa, para o sensoriamento remoto e para a pós-graduação.

A todos que de alguma forma contribuíram para a conclusão deste trabalho, o meu muito obrigada.

RESUMO

Para o adequado manejo das florestas, faz-se necessária a quantificação de métricas que a descrevam, tais como a área basal, volume e biomassa. Tendo essa necessidade, o objetivo do presente estudo é analisar o potencial em estimar de forma indireta variáveis dendrométricas a partir de dados obtidos por sensoriamento remoto (Radar Polarimétrico de Abertura Sintética (PolSar) e multiespectral), bem como, o potencial de cada variável e sua integração. Este trabalho foi constituído de dois estudos de caso. No primeiro estudo de caso, tendo os dados de 192 parcelas ajustou-se modelos de regressão linear e aprendizado de máquina para estimar a área basal, volume e biomassa aérea de plantações de eucalipto com diferentes idades. Como variáveis independentes utilizou-se da idade do povoamento, dados PolSar em banda C (satélite SENTINEL-1) e L (ALOS/PALSAR-2), assim como as derivações que originaram a textura e índices polarimétricos, imagens multiespectrais (SENTINEL-2) e índices de vegetação. O segundo estudo de caso consistiu em estimar a biomassa florestal de parte da floresta amazônica. Como variável dependente da modelagem utilizou-se do valor de biomassa aérea de 286 parcelas, cuja biomassa foi calculada a partir de métricas obtidas por LiDAR. Como variáveis de entrada utilizou-se da integração de dados de livre acesso, sendo eles PolSar de banda C oriundo do satélite SENTINEL-1, banda L oriunda do mosaico global do satélite ALOS/PALSAR-2, textura e índices polarimétricos obtidos das bandas C e L, imagens multiespectrais (SENTINEL-2) e índices de vegetação derivados. A modelagem com aprendizado de máquina para as duas áreas de estudo foi realizada com cinco modelos, sendo: *Model Tree*, *Random Forest*, *Weighted K-Nearest Neighbor*, *Artificial Neural Network* e *Support Vector Regression*. Em ambos os estudos de caso, buscou-se avaliar a influência da composição de diferentes conjuntos de dados de entrada (tratamentos) para a modelagem. Para os povoamentos de eucalipto verificou-se que a integração das bandas C e L constituíram modelos mais precisos do que a sua utilização separadamente e essa combinação PolSar foi superior ao modelar a área basal. Os erros foram maiores ao estimar o volume e biomassa aérea, mesmo assim foi necessário o uso de todas as informações PolSar e imagens multiespectrais para a sua modelagem. Utilizou-se de todo o conjunto de dados para estimar as variáveis dendrométricas por aprendizado de máquina e verificou-se que o modelo de melhor desempenho foi o *Random Forest* ($R^2 > 0,79$ e $S_{YX} < 25\%$), com desempenho superior ao da regressão linear em aproximadamente 8% de erro padrão para a estimativa da área basal, 13% para o volume total e 15% para a biomassa aérea. O erro padrão das estimativas pelo *Random Forest* foi menor e aproximadamente 10% para a estimativa das variáveis dendrométricas. Na estimativa da biomassa da floresta amazônica, a utilização de dados PolSar resultou em RMSE de 20%, com a integração da textura proporcionando leve melhoria no desempenho. A integração dos dados multiespectrais e índices de vegetação aos dados PolSar proporcionou perceptível melhora aos modelos de aprendizado de máquina (RMSE de 16%). O algoritmo *Model Tree* foi o mais adequado para as estimativas pois foi capaz de se aproximar dos valores extremos da base de dados. As estimativas das variáveis dendrométricas para as florestas em estudo com a utilização de PolSar foram satisfatórias e precisas, e a integração de dados de diferentes fontes proporcionou melhora no desempenho dos modelos.

Palavras-chave: Sensoriamento remoto. Aprendizado de máquina. Eucalipto. Amazônia. Variáveis dendrométricas.

ABSTRACT

For the proper management of forests, it is necessary to quantify metrics that describe it, such as the basal area, volume and biomass. With this need, the aim of the present study is to analyze the potential to indirectly estimate dendrometric variables from data obtained by remote sensing (Polarimetric Synthetic Aperture Radar (PolSar) and multispectral), as well as the potential of each variable and their integration. This work consisted of two case studies. In the first case study, using data from 192 plots, linear regression models and machine learning were adjusted to estimate the basal area, volume and aboveground biomass of eucalyptus plantations with different ages. As independent variables, population age was used, PolSar data in band C (satellite SENTINEL-1) and L (ALOS/PALSAR-2), as well as the derivations that originated the texture and polarimetric indices, multispectral images (SENTINEL-2) and derived vegetation indices. The second case study consisted of estimating the forest biomass of part of the Amazon rainforest. As a model-dependent variable, the aerial biomass value of 286 plots was used, whose biomass was calculated from metrics obtained with LiDAR. As input variables we used the integration of free access data, being PolSar of band C from the satellite SENTINEL-1, band L from the global mosaic of the satellite ALOS/PALSAR-2, texture and polarimetric indices obtained from bands C and L, multispectral images (SENTINEL-2) and derived vegetation indices. Modeling with machine learning for the two study areas was carried out using five models: Model Tree, Random Forest, Weighted K-Nearest Neighbor, Artificial Neural Network and Support Vector Regression. In both case studies, we sought to evaluate the influence of the composition of different sets of input data (treatments) for modeling. For the eucalyptus stands, it was found that the integration of the C and L bands constituted more accurate models than their use separately, and this PolSar combination was superior when modeling the basal area. The errors were greater when estimating the volume and aerial biomass, even so it was necessary to use all PolSar information and multispectral images for its modeling. The entire data set was used to estimate the dendrometric variables by machine learning and it was found that the best performing model was the Random Forest ($R^2 > 0.79$ and $S_{YX} < 25\%$), with a higher performance than the linear regression at approximately 8% of standard error for baseline area estimation, 13% for total volume and 15% for aerial biomass. The standard error of the estimates by the Random Forest was smaller and approximately 10% for the estimation of the dendrometric variables. In estimating the biomass of the Amazon rainforest, the use of PolSar data resulted in an RMSE of 20%, with the integration of the texture providing a slight improvement in performance. The integration of multispectral data and vegetation indexes with PolSar data provided a noticeable improvement to machine learning models (RMSE of 16%). The Model Tree algorithm was the most suitable for the estimates because it was able to approximate the extreme values of the database. The estimates of the dendrometric variables for the forests under study with the use of PolSar were satisfactory and accurate, and the integration of data from different sources provided an improvement in the performance of the models.

Keywords: Remote sensing. Forest management. Eucalipto. Amazon. Dendrometric variables.

LISTA DE FIGURAS

FIGURA 1 - DISTRIBUIÇÃO ANUAL DA QUANTIDADE DE ARTIGOS PUBLICADAS SOBRE O TEMA E CITAÇÕES	23
FIGURA 2 - PRINCIPAIS PESQUISADORES DA ÁREA, INCLUÍDOS COMO AUTOR PRINCIPAL OU COAUTOR DOS ARTIGOS CIENTÍFICOS.....	23
FIGURA 3 - REPRESENTAÇÃO GRÁFICA DO RELACIONAMENTO ENTRE AUTORES EM PESQUISA SOBRE ESTIMATIVA DE BIOMASSA FLORESTAL UTILIZANDO SAR	25
FIGURA 4 - NÚMERO DE ARTIGOS E PAÍSES QUE TIVERAM SUA ÁREAS ESTUDADAS PELOS PAÍSES DE MAIOR PRODUÇÃO CIENTÍFICA..	27
FIGURA 5 - MAPEAMENTO DA COAUTORIA DE INSTITUIÇÕES DE PESQUISA	28
FIGURA 6 - MAPEAMENTO DA COCORRÊNCIA DOS 35 TERMOS MAIS CITADAS NOS ARTIGOS.....	29
FIGURA 7 - NÚMERO DE ARTIGOS CIENTÍFICOS PUBLICADOS COM CADA BANDA DO ESPECTRO ELETROMAGNÉTICO.....	30
FIGURA 8 – LOCALIZAÇÃO ESPACIAL DAS PARCELAS UTILIZADAS NESTE ESTUDO	41
FIGURA 9 – PRECIPITAÇÃO DIÁRIA EM MM DOS SETE DIAS ANTERIORES A AQUISIÇÃO DAS CENAS	48
FIGURA 10 - FLUXOGRAMA METODOLÓGICO	54
FIGURA 11 - DISPERSÃO DO RETROESPALHAMENTO DA IMAGEM DUAL-POL DO SATÉLITE SENTINEL-1 SOBRE OS VALORES DE BIOMASSA AÉREA	59
FIGURA 12 - DISPERSÃO DO RETROESPALHAMENTO DA IMAGEM FULL-POLARIMÉTRICA DO SATÉLITE ALOS/PALSAR-2 SOBRE OS VALORES DE BIOMASSA AÉREA	59
FIGURA 13 - DISTRIBUIÇÃO DOS RESÍDUOS PARA A ESTIMATIVA DE ÁREA BASAL ($m^2 \cdot ha^{-1}$) POR REGRESSÃO LINEAR	61
FIGURA 14 - DISTRIBUIÇÃO DOS RESÍDUOS PARA A ESTIMATIVA DE VOLUME TOTAL ($m^3 \cdot ha^{-1}$) POR REGRESSÃO LINEAR	64
FIGURA 15 - DISTRIBUIÇÃO DOS RESÍDUOS PARA A ESTIMATIVA DE BIOMASSA AÉREA ($Mg \cdot ha^{-1}$) POR REGRESSÃO LINEAR.....	67

FIGURA 16 - DISTRIBUIÇÃO DOS RESÍDUOS PARA A ESTIMATIVA DE ÁREA BASAL ($m^2 \cdot ha^{-1}$) POR APRENDIZADO DE MÁQUINA	70
FIGURA 17 - DISTRIBUIÇÃO DOS RESÍDUOS PARA A ESTIMATIVA DE VOLUME TOTAL ($m^3 \cdot ha^{-1}$) POR APRENDIZADO DE MÁQUINA	72
FIGURA 18 - DISTRIBUIÇÃO DOS RESÍDUOS PARA A ESTIMATIVA DE BIOMASSA AÉREA ($Mg \cdot ha^{-1}$) POR APRENDIZADO DE MÁQUINA.....	74
FIGURA 19 - DISTRIBUIÇÃO DOS RESÍDUOS PARA A ESTIMATIVA REGRESSÃO LINEAR E DO ALGORITMO DE MELHOR DESEMPENHO (RF) PARA A ESTIMATIVA DA ÁREA BASAL (G), VOLUME TOTAL (V) E BIOMASSA AÉREA (AGB)	76
FIGURA 20 - ERRO DE AMOSTRAGEM ENTRE OS VALORES ORIUNDOS DO INVENTÁRIO E DAS ESTIMATIVAS COM MODELAGEM POR REGRESSÃO LINEAR E RANDOM FOREST	77
FIGURA 21 - COMPARAÇÃO ENTRE AS MÉDIAS E INTERVALO DE CONFIANÇA POR CLASSE DE IDADE PARA AS MÉTRICAS ORIUNDAS DO INVENTÁRIO (Obs) E ORIUNDAS DAS ESTIMATIVAS POLSAR	78
FIGURA 22 - AGB ($Mg \cdot ha^{-1}$) OBTIDA POR MEIO DE DADOS LASER AEROTRANSPORTADO E LOCALIZAÇÃO DA ÁREA DE ESTUDO SOBRE O BRASIL	89
FIGURA 23 - FLUXO DE TRABALHO REALIZADO COM OS DADOS DISPONÍVEIS NESTE ESTUDO	97
FIGURA 24 - DISTRIBUIÇÃO DOS VALORES DE AGB REPRESENTADOS SOBRE UM HISTOGRAMA DE FREQUÊNCIA E BOXPLOT	98
FIGURA 25 - BOXPLOT DOS VALORES OBSERVADOS E ESTIMADOS EM CADA UM DOS TRATAMENTOS	102
FIGURA 26 - DISPERSÃO GRÁFICA DOS RESÍDUOS GERADOS PELOS MODELOS DE APRENDIZADO DE MÁQUINA COM MELHOR DESEMPENHO, SOBRE O EIXO X QUE REPRESENTA OS VALORES ESTIMADOS ..	104

LISTA DE TABELAS

TABELA 1 - COLEÇÕES DISPONÍVEIS DA BASE DE DADOS WOS E ANO DE INÍCIO	21
TABELA 2 - DISTRIBUIÇÃO ANUAL DO NÚMERO DE ARTIGOS PUBLICADOS E DAS CITAÇÕES DESTES ARTIGOS NAS SEIS PRINCIPAIS REVISTAS	26
TABELA 3 – ESPECIFICAÇÕES DAS IMAGENS DOS SATÉLITES SENTINEL-1 E SENTINEL-2	44
TABELA 4 – ÍNDICES DE VEGETAÇÃO APLICADOS AS BANDAS DO SATÉLITE SENTINEL-2 A	45
TABELA 5 - CARACTERÍSTICAS DO SISTEMA ALOS/PALSAR-2	47
TABELA 6 - PARÂMETROS DAS CENAS UTILIZADAS E ADQUIRIDAS PELO SATÉLITE ALOS-2 SENSOR PALSAR-2	47
TABELA 7 - MODELOS DE APRENDIZADO DE MÁQUINA UTILIZADOS PARA A MODELAGEM DA AGB DA FLORESTA TROPICAL	52
TABELA 8 - ESTATÍSTICA DESCRITIVA DAS VARIÁVEIS DENDROMÉTRICAS E ÍNDICES DE VEGETAÇÃO GERADOS A PARTIR DAS IMAGENS DO SATÉLITE SENTINEL-2	55
TABELA 9 - ESTATÍSTICA DESCRITIVA DAS POLARIZAÇÕES DO SATÉLITE SENTINEL-1, ALOS/PALSAR-2 E ÍNDICES POLARIMÉTRICOS DO ALOS/PALSAR-2	56
TABELA 10 - CORRELAÇÃO DE SPEARMAN (ρ) ENTRE AS VARIÁVEIS DENDROMÉTRICAS E AS VARIÁVEIS SELECIONADAS PARA CADA TRATAMENTO	58
TABELA 11 - ESTATÍSTICAS DE AJUSTE E DESEMPENHO DA REGRESSÃO LINEAR PARA A ESTIMATIVA DE ÁREA BASAL ($m^2 \cdot ha^{-1}$)	60
TABELA 12 - EQUAÇÕES AJUSTADAS PARA a ESTIMATIVA DA ÁREA BASAL (G) ($m^2 \cdot ha^{-1}$) E O P-VALOR DO TESTE DE NORMALIDADE E HOMOGENEIDADE	62
TABELA 13 - ESTATÍSTICAS DE AJUSTE E DESEMPENHO DA REGRESSÃO LINEAR PARA A ESTIMATIVA DE VOLUME TOTAL ($m^3 \cdot ha^{-1}$)	63

TABELA 14 - EQUAÇÕES AJUSTADAS PARA ESTIMATIVA DO VOLUME TOTAL (V) ($\text{m}^3 \cdot \text{ha}^{-1}$) E O P-VALOR DO TESTE DE NORMALIDADE E HOMOGENEIDADE.....	65
TABELA 15 - ESTATÍSTICAS DE AJUSTE E DESEMPENHO DA REGRESSÃO LINEAR PARA A ESTIMATIVA DE BIOMASSA AÉREA ($\text{Mg} \cdot \text{ha}^{-1}$).....	66
TABELA 16 - EQUAÇÕES AJUSTADAS PARA ESTIMATIVA DA BIOMASSA AÉREA (AGB) ($\text{Mg} \cdot \text{ha}^{-1}$) E O P-VALOR DO TESTE DE NORMALIDADE E HOMOGENEIDADE.....	68
TABELA 17 - ESTATÍSTICAS DE AJUSTE E DESEMPENHO DOS MODELOS DE APRENDIZADO DE MÁQUINA PARA A ESTIMATIVA DE ÁREA BASAL ($\text{m}^2 \cdot \text{ha}^{-1}$).....	69
TABELA 18 - ESTATÍSTICAS DE AJUSTE E DESEMPENHO DOS MODELOS DE APRENDIZADO DE MÁQUINA PARA A ESTIMATIVA DO VOLUME TOTAL ($\text{m}^3 \cdot \text{ha}^{-1}$).....	71
TABELA 19 - ESTATÍSTICAS DE AJUSTE E DESEMPENHO DOS MODELOS DE APRENDIZADO DE MÁQUINA PARA A ESTIMATIVA DA BIOMASSA AÉREA ($\text{Mg} \cdot \text{ha}^{-1}$).....	73
TABELA 20 - ESTATÍSTICAS DE AJUSTE E DESEMPENHO DA REGRESSÃO LINEAR E DO ALGORITMO RANDOM FOREST (RF) PARA A ESTIMATIVA DA ÁREA BASAL (G), VOLUME TOTAL (V) E BIOMASSA AÉREA (AGB).....	75
TABELA 21 - RESUMO DAS ESPECIFICAÇÕES DOS SATÉLITES SENTINEL-1 E SENTINEL-2.....	90
TABELA 22 - ESPECIFICAÇÕES DO AEROLEVANTAMENTO COM LIDAR.....	93
TABELA 23 - MODELOS DE APRENDIZADO DE MÁQUINA UTILIZADOS PARA A MODELAGEM DA AGB DA FLORESTA TROPICAL.....	96
TABELA 24 - CORRELAÇÃO DE SPEARMAN (ρ) ENTRE AGB E AS VARIÁVEIS PREDITORAS SELECIONADAS PELO ALGORITMO CFS.....	99
TABELA 25 - DESEMPENHO DOS ALGORITMOS FRENTE ÀS VARIÁVEIS PREDITORAS DO TRATAMENTO 1, UTILIZAÇÃO DOS VALORES DE RETROESPALHAMENTO DAS BANDAS C E L.....	100
TABELA 26 - DESEMPENHO DOS ALGORITMOS FRENTE ÀS VARIÁVEIS PREDITORAS DO TRATAMENTO 2, ADIÇÃO DE VARIÁVEIS DE TEXTURA E ÍNDICES POLARIMÉTRICOS.....	101

TABELA 27 - DESEMPENHO DOS ALGORITMOS FRENTE ÀS VARIÁVEIS PREDITORAS DO TRATAMENTO 3. USO DO RETROESPALHAMENTO, TEXTURA COM A ADIÇÃO DAS BANDAS E SUBPRODUTOS DO SATÉLITE SENTINEL-2.....	101
--	-----

LISTA DE ABREVIATURAS OU SIGLAS

AGB	- <i>Above Ground Biomass</i>
ALOS	- <i>Advanced Land Observing Satellite</i>
BMI	- Biomassa florestal
CSI	- Estrutura do dossel
EUA	- Estados Unidos da América
GNSS	- <i>Global Navigation Satellite System</i>
HyP3	- <i>Hybrid Pluggable Processing Pipeline</i>
IAF	- Índice de área foliar
JERS	- <i>Japanese Earth Resource Satellite</i>
NDVI	- Índice de Vegetação Normalizada
PALSAR	- <i>Phased Array L-band Synthetic Aperture Radar</i>
POLSAR	- Radar Polarimétrico de Abertura Sintética
RADAR	- <i>Radio Detection and Ranging</i>
RFDI	- <i>Radar Forest Degradation Index</i>
RVI	- Índice de vegetação
SAR	- Radar de Abertura Sintética
SAVI	- Índice de Vegetação Ajustado ao Solo
SNAP	- <i>Sentinel Application Platform</i>
UTM	- <i>Universal Transversa de Mercator</i>
VSI	- Espalhamento volumétrico

SUMÁRIO

INTRODUÇÃO GERAL	15
OBJETIVOS	17
CAPÍTULO 1: REVISÃO BIBLIOMÉTRICA SOBRE ESTIMATIVA DE BIOMASSA FLORESTAL UTILIZANDO RADAR DE ABERTURA SINTÉTICA.....	19
1.1 INTRODUÇÃO.....	19
1.2 OBJETIVOS.....	20
1.3 MATERIAL E MÉTODOS.....	21
1.4 RESULTADOS	22
1.5 DISCUSSÃO.....	30
1.5.1 Perspectivas para o Futuro.....	33
1.6 CONCLUSÃO.....	34
CAPÍTULO 2: ESTIMATIVA DE VARIÁVEIS DENDROMÉTRICAS AO NÍVEL DE PARCELA OBTIDA PELA INTEGRAÇÃO DE SENSORIAMENTO REMOTO ATIVO E PASSIVO EM POVOAMENTOS DE <i>EUCALYPTUS</i> SPP.	36
2.1 INTRODUÇÃO.....	36
2.1.1 Aprendizado de máquina.....	38
2.2 OBJETIVOS.....	40
2.3 MATERIAL E MÉTODOS.....	41
2.3.1 Área de estudo	41
2.3.2 Características e mensuração das parcelas amostrais	42
2.3.2.1 Processamento das variáveis dendrométricas.....	43
2.3.3 Aquisição e processamento de dados espaciais.....	43
2.3.3.1 Dados SENTINEL-1 e SENTINEL-2	44
2.3.3.2 Dados ALOS/PALSAR-2.....	46
2.3.3.2.1 Processamento dos dados ALOS/PALSAR-2.....	48
2.3.4 Modelagem para a estimativa de variáveis dendrométricas	50
2.3.4.1 Regressão linear múltipla	50
2.3.4.2 Aprendizado de máquina.....	51
2.3.5 Avaliação dos modelos.....	53
2.3.6 Fluxograma Metodológico	54
2.4 RESULTADOS	55
2.4.1 Análise dos dados	55

2.4.2 Seleção de variáveis e ajuste dos modelos	60
2.4.2.1 Regressão linear múltipla	60
2.4.2.2 Aprendizado de máquina	68
2.4.3 Regressão linear x aprendizado de máquina.....	75
2.4.4 Erros oriundos do inventário e da modelagem	77
2.5 DISCUSSÃO	78
2.6 CONCLUSÃO.....	84
CAPÍTULO 3: ESTIMATIVA DE BIOMASSA EM FLORESTA TROPICAL POR MEIO DE APRENDIZAGEM DE MÁQUINA E DADOS MULTISENSOR OBTIDO POR SENSORIAMENTO REMOTO DE LIVRE ACESSO.....	86
3.1 INTRODUÇÃO.....	86
3.2 OBJETIVOS.....	89
3.3 MATERIAL E MÉTODOS.....	89
3.3.1 Área de estudo	89
3.3.2 Aquisição e processamento de dados SENTINEL	90
3.3.3 Aquisição e processamento do mosaico ALOS/PALSAR-2.....	91
3.3.4 Dados LiDAR e Processamento	92
3.3.5 Modelagem da biomassa a partir das imagens SENTINEL	94
3.4 RESULTADOS	97
3.4.1 Banco de dados e seleção de atributos	97
3.4.2 Modelagem da biomassa	100
3.5 DISCUSSÃO	105
3.5.1 Análise do banco de dados	105
3.5.2 Modelagem da biomassa	106
3.6 CONCLUSÃO.....	109
DISCUSSÃO GERAL	110
CONCLUSÃO GERAL	113
REFERÊNCIAS	114

INTRODUÇÃO GERAL

Tanto no setor de florestas plantadas quanto em florestas naturais, se faz necessária a mensuração de variáveis dendrométricas para um adequado manejo e aproveitamento dos recursos. Para tal, as técnicas de sensoriamento remoto têm muitas vantagens na estimativa de variáveis dendrométricas em relação aos métodos tradicionais de medição em campo e tem potencial para estimativas em grandes escalas, com destaque para os radares.

O termo Radar provém da sigla do termo *Radio Detection and Ranging* e se refere à utilização de ondas microondas, no caso dos radares imageadores, para detecção e posicionamento de objetos (BALDAUF; GARCIA, 2016). Existem vários tipos de sistemas de RADAR, porém na área florestal normalmente se utilizam os sensores RADAR polarimétricos de abertura sintética, conhecidos como PolSar (*Polarimetric Synthetic Aperture RADAR*) (BALDAUF; GARCIA, 2016; HYYPPÄ et al., 2015).

Esses sensores mostram grande potencial devido à capacidade de atuarem tanto de dia quanto a noite, e mesmo em condições climáticas adversas (KÖHL; MAGNUSSEN; MARCHETTI, 2006). O princípio de funcionamento dos sensores PolSar é o de um sistema imageador, que envia energia (sensor ativo) e coleta a energia retransmitida sendo que esta é chamada de retroespalhamento ou *backscattering* (KÖHL; MAGNUSSEN; MARCHETTI, 2006).

A informação de retroespalhamento baseada em dados PolSar é amplamente conhecida para o mapeamento de parâmetros biofísicos da floresta, como o volume e a *Above-Ground Biomass* (AGB) que se refere a biomassa acima do solo. Consiste na mensuração da energia recebida pelo sensor após a transmissão, conhecida como *backscatter*, em que posteriormente essa medida é relacionada com as medições da floresta em campo. O *backscatter* é dependente do comprimento de onda, da polarização e é sensível à quantidade de AGB, este valor é diretamente proporcional ao aumento da AGB, até certo ponto em que ocorre a saturação, no qual a sensibilidade do *backscatter* à AGB fica estagnada (GHASEMI; SAHEBI; MOHAMMADZADEH, 2011).

Os dados PolSar podem ser adquiridos em bandas com diferentes frequências (X, C, S, L e P), diferentes polarizações (HH, VV, VH, HV), resolução de alcance e azimute. Cada banda com suas peculiaridades em relação ao estande florestal. A banda X interage com as folhas e a superfície de cobertura do dossel, sendo, portanto, adequada para a informação da camada superficial da copa das árvores. A banda C penetra por meio das folhas e o sinal é espalhado por pequenos ramos e características subjacentes. A banda L penetra pelas camadas

superficiais e é espalhada pelo tronco e ramos principais. Com maior capacidade de penetração, a banda P adentra na cobertura da copa, tendo a maior parte do retroespalhamento dada pelo tronco e as interações entre tronco e terra (RANSON; SUN, 1994; PONZONI; SHIMABUKURO; KUPLICH, 2012).

A interação entre a energia e o estande florestal depende do conteúdo de umidade do solo, já que influencia nas propriedades dielétricas dos alvos e consequente retroespalhamento (WOODHOUSE, 2006).

Além do comprimento de onda, outro parâmetro importante dos dados PolSar é a polarização dos sinais transmitidos e recebidos. A polarização é a direção do campo elétrico nas ondas eletromagnéticas e é o principal fator nas interações entre os sinais e os refletores. Os sensores de microondas podem emitir e receber sinais em polarizações horizontais (H) ou verticais (V). Tem-se quatro combinações de polarizações de dados PolSar: HH: os sinais emitidos e recebidos tem polarização horizontal; HV: o sinal emitido tem polarização horizontal e o sinal recebido tem polarização vertical; VH: o sinal emitido tem polarização vertical e o sinal recebido possui polarização horizontal; VV: ambos os sinais emitidos e recebidos tem polarização vertical (GHASEMI; SAHEBI; MOHAMMADZADEH, 2011).

Para cada característica da vegetação e local de área de estudo tem-se um PolSar mais apropriado, porém a indisponibilidade de alguns dados, de banda P por exemplo, deixa um déficit de informações. Para isso, necessita-se de metodologias alternativas para buscar suprir essa lacuna de conhecimento. Este estudo tem como contribuição buscar responder algumas perguntas, tais como: i) Qual a capacidade que o PolSar de banda C possui na representação de variáveis dendrométricas?; ii) Qual a capacidade que o PolSar de banda L possui na representação de variáveis dendrométricas?; iii) Qual a capacidade que a integração do PolSar de banda C e L possuem na representação de variáveis dendrométricas?; iv) Qual a capacidade que as informações oriundas de satélite óptico possuem na representação de variáveis dendrométricas?; v) Qual a capacidade que a integração de dados PolSar e ópticos possuem na representação de variáveis dendrométricas?.

Com o desafio da integração de dados de diferentes fontes, surgem outros desafios, como a alta dimensionalidade dos dados, a redundância de algumas métricas e a necessidade da seleção do modelo de predição mais adequado.

Modelos de regressão linear são comumente usados para estimar variáveis dendrométricas a partir de dados de sensoriamento remoto, devido a sua facilidade no ajuste e interpretabilidade (FASSNACHT et al., 2014). Já as técnicas não paramétricas, a exemplo o aprendizado de máquina, são mais versáteis que a regressão linear na identificação de relações

não lineares complexas e no tratamento de alta dimensionalidade dos dados (ALMEIDA et al., 2019). Tais técnicas podem fornecer estimativas de variáveis dendrométricas mais precisas do que os modelos de regressão linear, especialmente quando dados *multisource* são usados (LU et al., 2014).

A hipótese principal que guia este estudo é que é possível estimar variáveis dendrométricas de plantações de eucalipto e de floresta tropical de forma indireta, utilizando apenas de informações obtidas por sensoriamento remoto.

Buscando atender a hipótese, esta tese foi dividida em três capítulos, sendo que nestes são abordados:

Capítulo 1: Panorama bibliométrico dos estudos que estimaram biomassa florestal utilizando Radar de Abertura Sintética;

Capítulo 2: Aborda um estudo de caso em que buscou-se obter de forma indireta a área basal, o volume e a biomassa aérea de eucalipto de diferentes idades com apenas o uso de dados de sensoriamento remoto. Tem como intuito fornecer modelos para o macroplanejamento florestal em casos de expansão da indústria, no qual se faz necessária uma busca por áreas produtivas em um certo raio de abrangência, ou como suporte para especuladores, auditores ou qualquer pessoa interessada em conhecer a floresta para fins comerciais ou de fiscalização sem, no entanto, ter acesso a nenhuma das métricas da floresta adquiridas por inventário florestal tradicional, tais como, medidas de diâmetros e alturas médias.

Capítulo 3: Aborda um estudo de caso para obtenção de forma indireta da biomassa aérea para floresta amazônica.

A partir deste estudo espera-se encontrar resultados que deem suporte para a utilização de PolSar em mensuração florestal em área de plantações de eucalipto e floresta amazônica.

OBJETIVOS

O objetivo geral foi analisar o potencial em estimar de forma indireta variáveis dendrométricas a partir de dados PolSar. Sendo assim, os objetivos específicos são atendidos em três capítulos:

- 1) Descrever um panorama bibliométrico das publicações que versam sobre a estimativa de biomassa florestal utilizando Radar de Abertura Sintética;

- 2) Analisar a possibilidade de estimar variáveis dendrométricas (área basal, volume e biomassa aérea) de povoamentos de eucalipto com diferentes idades, a partir de dados PolSar;

3) Analisar a potencialidade do uso de imagens de livre acesso em estimar a AGB de floresta tropical.

CAPÍTULO 1: REVISÃO BIBLIOMÉTRICA SOBRE ESTIMATIVA DE BIOMASSA FLORESTAL UTILIZANDO RADAR DE ABERTURA SINTÉTICA

Resumo: A principal forma terrestre de fixação de carbono é pelas florestas, e para que seja possível executar políticas de redução das emissões de carbono, são necessárias metodologias eficientes para a quantificação da biomassa florestal. Sabendo que as estimativas por sensoriamento remoto ativo possuem várias vantagens, dentre elas a possibilidade de quantificação para larga escala, muitos estudos foram realizados. Sendo assim, o objetivo do presente estudo é apresentar um panorama das publicações referentes às estimativas de biomassa florestal utilizando dados de Radar de Abertura Sintética (SAR). Para tal, foi realizado um estudo bibliométrico por meio da pesquisa dos termos “*estima**”, “*forest biomass*” e “*SAR*”, nas bases de dados indexadoras *Scopus* e *Web of Science*. Foram considerados todos os artigos publicados em língua inglesa, independente do ano da sua publicação, resultando em 159 artigos para a análise. Constatou-se que o mais antigo artigo relacionado ao tema nas bases de dados analisadas, foi publicado em 1991 e que desde então o assunto obteve crescimento. Entre todas as revistas, a que mais publicou o assunto e a mais citada é a *Remote Sensing of Environment*. Considerando a afiliação do primeiro autor, o país que mais publica e é detentor das áreas em pesquisa é os Estados Unidos da América, dentre elas a instituição Caltech, e dentre as métricas SAR, a banda L é a mais empregada nos estudos.

Palavras-chave: bibliometria; quantificação da floresta; carbono; sensoriamento remoto.

1.1 INTRODUÇÃO

A preocupação internacional sobre as mudanças climáticas causadas pelo aumento das emissões de carbono nas últimas décadas, resultaram em políticas e medidas, como o Acordo de Paris, que visam frear este cenário pela conservação dos remanescentes florestais, uma vez que a maior parte de sequestro de carbono terrestre é feito por florestas. Para quantificar esse carbono sequestrado na biomassa florestal, existem diversas formas de mensurações e monitoramento. No entanto, ainda são necessários métodos fidedignos para aplicação em larga escala.

Indiretamente, a quantificação de biomassa florestal pode ser realizada por três métodos principais: i) inventário florestal: envolvem medições de parcelas amostrais, e podem fornecer informações confiáveis em escala local ou regional, sendo estimativa direta por método destrutivo (CHAVE et al., 2004); ii) modelos alométricos: fornecem estimativas da biomassa em escala local a global com base em variáveis de fácil mensuração, como o diâmetro (LU, 2006); iii) imagens de satélite: geralmente são combinações de medições remotas e as obtidas pelo inventário em campo, permitem estimativas em escala global (STICKLER et al., 2009).

Dentre os dados obtidos por sensoriamento remoto, o Radar de Abertura Sintética (SAR) é um sensor ativo que permite o imageamento mesmo durante a noite ou com cobertura de nuvens, sendo amplamente estudado para quantificação da biomassa florestal (CLARK; KELLNER, 2012). Dado o potencial do SAR e a necessidade de estimativa de biomassa florestal em larga escala, surge a necessidade de analisar a tendência dos estudos já publicados com esse tema, os desafios a serem alcançados e as oportunidades para futuros estudos, buscando sempre otimizar o desempenho das estimativas.

O grande quantitativo de literatura científica disponível sobre o assunto, torna desafiador o trabalho de pesquisadores e profissionais para terem uma visão geral e estruturada das informações realmente relevantes (ZHOU; GOH; LI, 2015). Do ponto de vista acadêmico é importante que sejam classificados os materiais publicados, de forma a se acompanhar todos os estudos, avanços e tendências sobre a estimativa de biomassa florestal, um exemplo disso é o aumento na utilização de geotecnologias que permitem o mapeamento em grandes escalas. Uma maneira de realizar tal classificação da literatura existente, é através da bibliometria, que permite direcionar os interessados pelo assunto para estudos mais influentes e lacunas do conhecimento (GODIN, 2006).

A bibliometria é um método estatístico que permite uma visão macroscópica de uma grande quantidade de publicações, sendo eficaz em resumir, quantificar e analisar sistematicamente os estudos publicados em determinado assunto. Pode ser usado para avaliar padrões de pesquisa por autores, periódicos, países e instituições, e ainda para identificar redes de cooperação entre eles (LI; ZHAO, 2015). Com esse método, também é possível identificar os pontos-chave e tendências das pesquisas científicas sobre a estimativa de biomassa florestal utilizando dados SAR nas últimas décadas (MAO et al., 2018; LI; HALE, 2015). Portanto, pode ser utilizada para apoiar formuladores de políticas e agências de financiamento para a pesquisa (UGOLINI et al., 2015). Uma parte fundamental da pesquisa é revisar a literatura acadêmica existente em cada tema (CROOM, 2009).

1.2 OBJETIVOS

Este capítulo inicial da tese tem como objetivo geral descrever um panorama bibliométrico das publicações que versam sobre a estimativa de biomassa florestal utilizando SAR. De forma específica, identificar os principais autores sobre o tema, as revistas científicas que publicam, onde são suas áreas de estudo, quais as bandas utilizadas para o estudo e o relacionamento entre os autores desse tema.

1.3 MATERIAL E MÉTODOS

Os dados para este estudo foram obtidos nas bases de dados indexadoras Web of Science (WoS) e Scopus em 17 de junho de 2019. O portal de periódicos WoS foi escolhido como mecanismo de busca por ser o banco de dados mais aceito e utilizado para análise de publicações científicas (YANG et al., 2013). De acordo com Merigó et al. (2015) o WoS é uma plataforma bibliométrica digital reconhecida internacionalmente entre os pesquisadores por apresentar altos padrões de qualidade. O WoS possui em sua base de dados aproximadamente 12.000 periódicos e oferece a possibilidade de acesso a cinco coleções apresentadas na TABELA 1.

TABELA 1 - COLEÇÕES DISPONÍVEIS DA BASE DE DADOS WOS E ANO DE INÍCIO

Coleção	Disponibilidade de dados com início em:
<i>Science Citation Index Expanded (SCI-EXPANDED)</i>	1945
<i>Social Sciences Citation Index (SSCI)</i>	1945
<i>Arts & Humanities Citation Index (A&HCI)</i>	1975
<i>Conference Proceedings Citation Index- Science (CPCI-S)</i>	1991
<i>Conference Proceedings Citation Index - Social Science & Humanities (CPCI-SSH)</i>	1991

FONTE: A autora (2020).

A base de dados Scopus é a base referencial da Editora Elsevier, abrange as áreas de ciências biológicas, da saúde, físicas e sociais, e o período de acesso é desde o ano de 1823 até o presente momento (CAPES, 2018).

As palavras-chave foram definidas previamente visando realizar a pesquisa nos bancos de dados Scopus e WoS, sendo elas: “estima*”, “*forest biomass*” e “SAR”. Estas palavras foram buscadas no título, resumo e/ou palavras-chave. O asterisco no termo “estima” foi incluindo para garantir a eficácia da pesquisa, uma vez que permite variações desta, tais como, *estimation*, *estimated*. A pesquisa foi limitada para documentos publicados apenas em inglês na categoria de artigo científico, sendo ignorados os demais tipos de documento. Não se limitou o ano inicial de publicação, para que fosse possível analisar toda a evolução temporal até o final do ano 2018.

Todos os artigos filtrados nas bases de dados foram exportados em formato Research Information Systems (.ris) para o software EndNote X9 com todos os metadados, incluindo

autores, ano, resumo, palavras-chave, referências, número de citações, instituições, países, revista, ano de publicação entre outras informações complementares.

Ao realizar a busca pelas palavras-chave, foram encontrados 108 artigos na base de dados Scopus e 190 artigos na WoS. Do total de 298 artigos da pesquisa que foram exportados, 139 artigos foram excluídos por serem iguais (nome, autores e palavras-chave). Posteriormente realizou-se uma filtragem refinada no texto de cada artigo, a fim de excluir todos não relacionados com o tema, ou os que realizavam apenas a revisão ou a classificação do uso da terra, restando 159 artigos. A listagem dos 159 artigos pode ser encontrada no endereço: <https://url.gratis/IUiJA>.

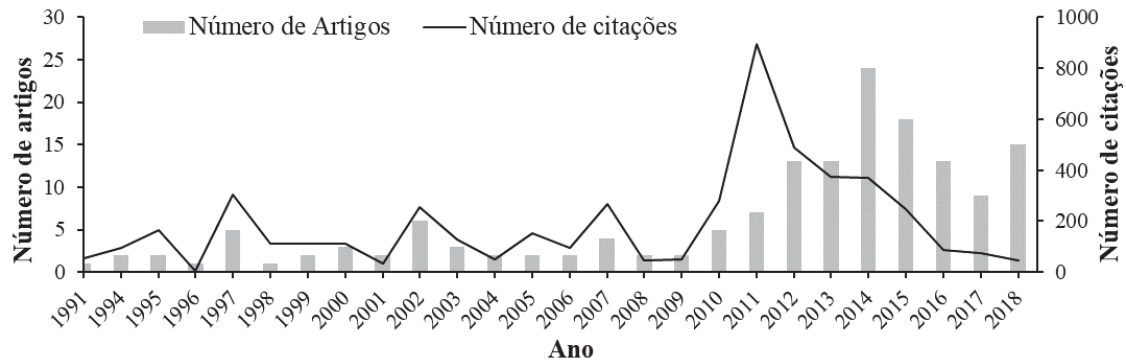
Os 159 registros associados aos termos da pesquisa, foram analisados por meio da análise bibliométrica, sendo analisados os seguintes tópicos: (1) produção de publicações, citações e tendência de crescimento; (2) principais autores/coautores; (3) principais revistas e sua evolução; (4) o país de filiação do primeiro autor e os principais países que estão sendo pesquisados; (5) principais bandas do espectro eletromagnético que estão sendo utilizadas (6) mapeamento da coautoria utilizando a base WoS; (7) mapeamento da coocorrência de termos utilizando a base WoS.

O mapeamento é um dos procedimentos principais da análise bibliométrica, consiste na representação espacial de como diferentes autores se relacionam (SMALL, 1999). A coocorrência de palavras-chave considera as palavras mais importantes de documentos para estudar a estrutura conceitual de um campo de pesquisa (CALLON et al., 1983). Com esse intuito, foi realizado o mapeamento da coautoria dos artigos e coocorrência de palavras-chave da base de dados da WoS com o auxílio do software VOSviewer (VAN ECK; WALTMAN, 2010). Os gráficos são representados por uma rede de círculos com tamanhos distintos, dependendo da importância do elemento, as conexões indicam as proximidades dos elementos e as cores representam o agrupamento destes em clusters.

1.4 RESULTADOS

O número de publicações em periódicos científicos revisado por pares é um indicador importante para mensurar o desenvolvimento de determinado assunto, sendo assim, observa-se uma evolução no número de artigos científicos que tratam sobre a estimativa de biomassa florestal utilizando dados SAR (FIGURA 1).

FIGURA 1 - DISTRIBUIÇÃO ANUAL DA QUANTIDADE DE ARTIGOS PUBLICADAS SOBRE O TEMA E CITAÇÕES



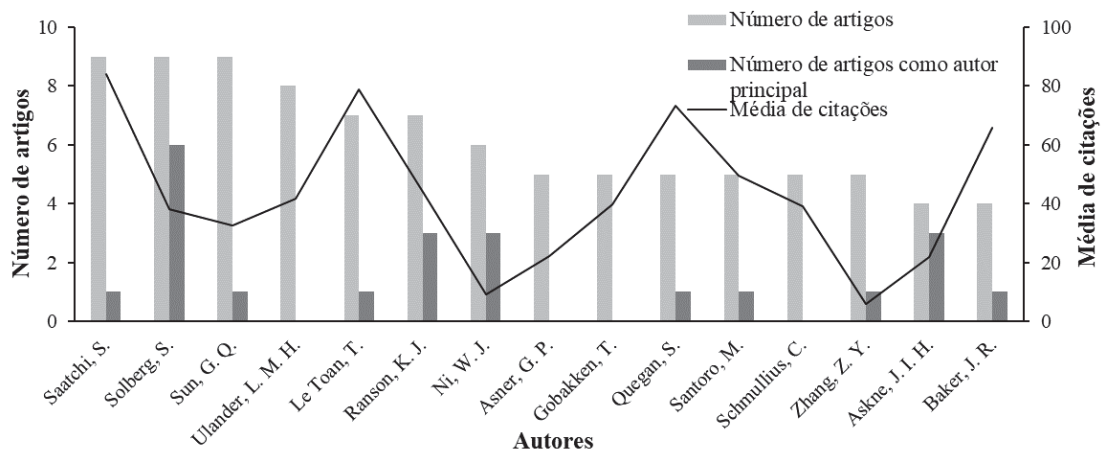
FONTE: A autora (2020).

Um pico de publicação é alcançado em 2014 ($n = 24$), após o qual ocorre uma tendência decrescente, com menor número de artigos em 2017 ($n = 9$).

Em relação ao número de citações, ocorre um pico em 2011 ($n = 893$) em que foram publicados apenas 7 artigos com este tema, indicando que este grande número de citações foi por artigos antigos ou de outros temas, e o mesmo se observa para o pico de publicação (2014) em que o número de citações dos artigos para aquele ano foi de apenas 368.

A classificação dos autores mais produtivos foi baseada no número total de artigos, desconsiderando a ordem de autoria, a listagem dos 15 autores que mais desenvolveram pesquisas sobre o tema está representada na FIGURA 2.

FIGURA 2 - PRINCIPAIS PESQUISADORES DA ÁREA, INCLUÍDOS COMO AUTOR PRINCIPAL OU COAUTOR DOS ARTIGOS CIENTÍFICOS



FONTE: A autora (2020).

Os 159 artigos foram escritos por 476 autores e coautores, em média três autores por artigo, destes, apenas três artigos tinham autoria única (1,89%), 23 artigos apresentam dois autores (14,47%), cerca de 19% dos artigos creditados por três, quatro ou cinco autores, e 27% dos artigos com autoria de seis ou mais autores (máximo de 16 autores).

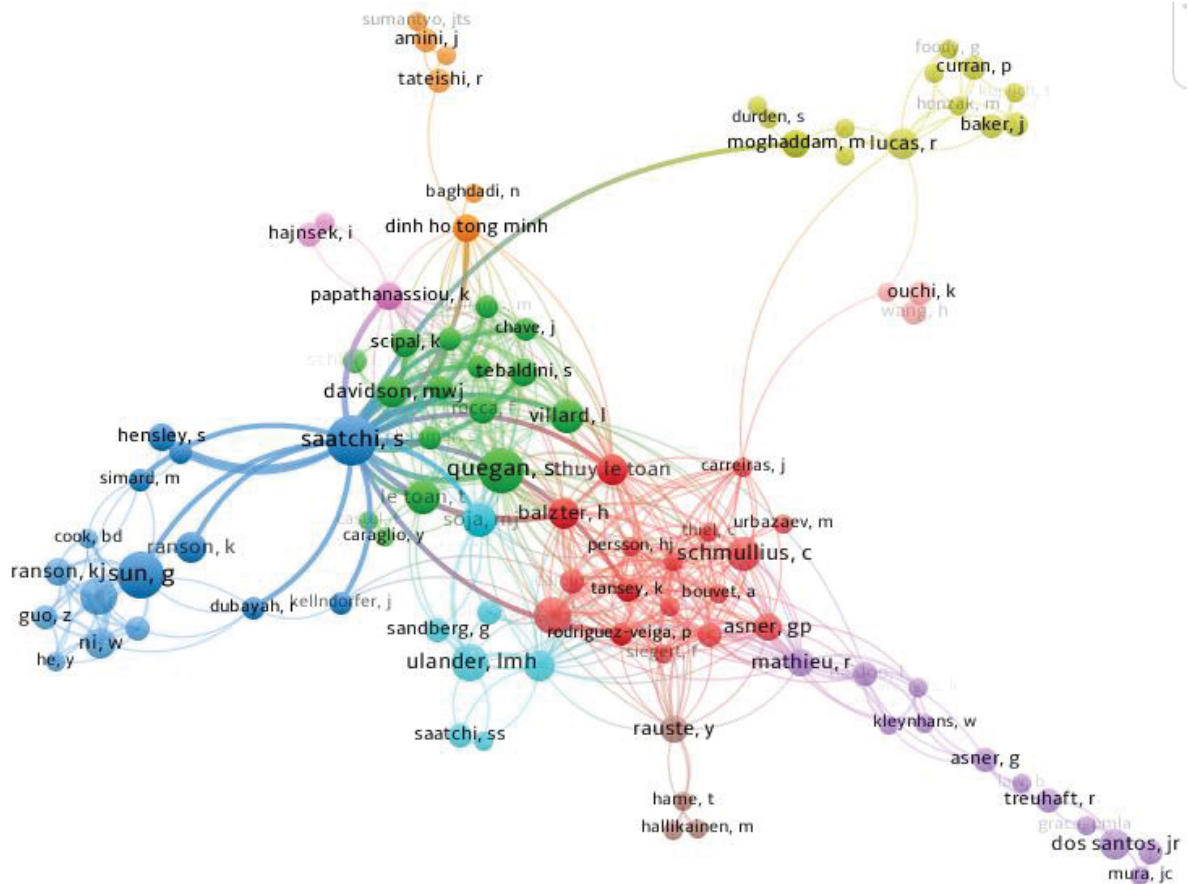
A maior parte dos autores (n = 342 ou 71,85%) participou da elaboração de apenas um artigo, 16,39% dos autores (n = 78) foram creditados por dois artigos, e os 11,76% restantes participaram da autoria de 3 a 9 artigos, sendo considerados como os mais influentes no tema.

Entre os autores/coautores mais produtivos sobre o tema, destacam-se o Norueguês S. Solberg e os Norte-americanos G. Sun e S. Saatchi, cada um destes responsáveis por 9 artigos. Dentre os autores mais produtivos, foi observado o Norueguês S. Solberg como autor principal em seis artigos, em média seus artigos foram citados 38 vezes. O segundo autor principal de maior produção é o Norte-americano R. N. Treuhft, responsável pela coautoria de quatro artigos, obtendo em média 32 citações.

No total dos 15 autores/coautores mais produtivos, o autor mais citado é o Norte Americano S. Saatchi com citação média de 84 vezes, seus estudos utilizaram principalmente as bandas L e P, e ocorreram nos Estados Unidos da América (EUA), Malásia, Suécia e México. Mas o artigo mais citado é da Francesa T. Le Toan, o qual foi publicado em 2011 e foi citado mais de 316 vezes, tendo como tema principal a missão BIOMASS, sendo que a metodologia aplicada pode ser realizada para qualquer outra região ou país. Dentre as publicações de Le Toan, a mesma utilizou as bandas X, C, L e P. Os estudos desta autora foram principalmente na França e Suécia.

Foram observados 154 autores com artigos disponíveis na base de dados indexadora WoS, com participação mínima em dois artigos publicados, sendo que 96 desses se conectam com os demais. Os autores foram agrupados em 10 clusters, destes, o maior conjunto de autores agrupados consiste em 35 nomes. O Norte-Americano S. Saatchi é autor/coautor de 9 artigos e apresenta a maior rede de cooperação, sendo formada de três grupos de autores com alta produtividade (FIGURA 3).

FIGURA 3 - REPRESENTAÇÃO GRÁFICA DO RELACIONAMENTO ENTRE AUTORES EM PESQUISA SOBRE ESTIMATIVA DE BIOMASSA FLORESTAL UTILIZANDO SAR



FONTE: A autora (2020).

No total, os 159 artigos relacionados ao tema foram publicados em 45 revistas científicas, dentre estas, as que publicaram o maior número foram: *Remote Sensing of Environment* (fator de impacto = 6,45), *IEEE Transactions on Geoscience and Remote Sensing* (fator de impacto = 4,66), *Remote Sensing* (fator de impacto = 3,41), *Internacional Journal of Remote Sensing* (fator de impacto = 1,78), *IEEE Geoscience and Remote Sensing Letters* (fator de impacto = 2,89), e *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (fator de impacto = 2,77), respectivamente (TABELA 2).

TABELA 2 - DISTRIBUIÇÃO ANUAL DO NÚMERO DE ARTIGOS PUBLICADOS E DAS CITAÇÕES DESTES ARTIGOS NAS SEIS PRINCIPAIS REVISTAS

	Remote Sensing of Environment		IEEE Transactions on Geoscience and Remote Sensing		Remote Sensing		International Journal of Remote Sensing		IEEE Geoscience and Remote Sensing Letters		IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing	
	Pub.	Cit.	Pub.	Cit.	Pub.	Cit.	Pub.	Cit.	Pub.	Cit.	Pub.	Cit.
1991			1	53								
1994	1	28					1	68				
1995			2	166								
1996												
1997	2	194					1	46				
1998	1	110										
1999			1	93								
2000			1	85								
2001			1	33								
2002	2	103					2	88				
2003												
2004			2	49								
2005	1	75					1	78				
2006	1	87							1	9		
2007	1	98	1	124			1	28	1	16		
2008			1	17			1	28				
2009			1	26			1	23				
2010	1	64			1	46			1	5	1	136
2011	5	734										
2012	1	75	2	123	1	10	1	24				
2013	4	181	3	67	1	55	1	25			1	12
2014	2	84	6	75	2	21			2	8	1	6
2015	2	62	3	48	1	7			3	58	2	19
2016	1	18	2	22	2	7						
2017	5	44			1	10					1	7
2018	3	23	1	0	4	12	1	0			2	2
Σ	33	1980	28	981	13	168	11	408	8	96	8	182

FONTE: A autora (2020).

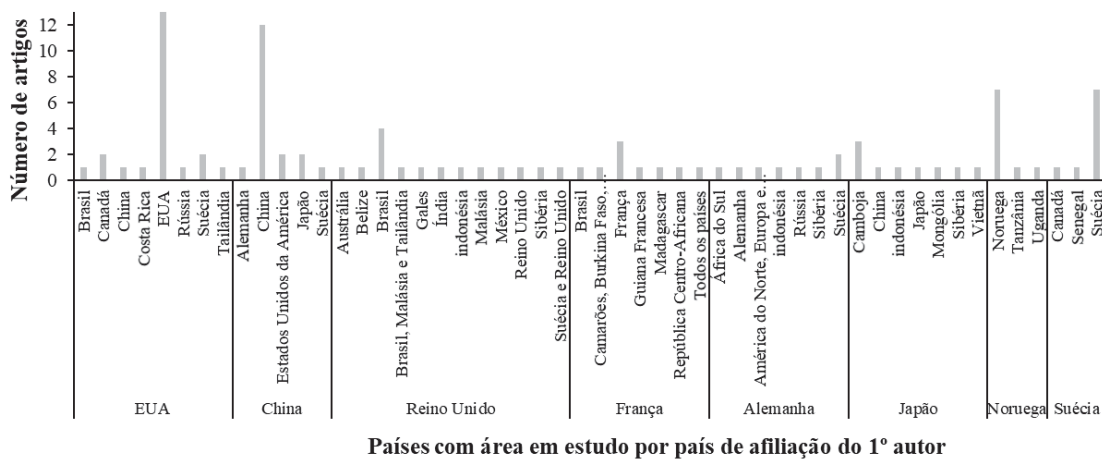
LEGENDA: Pub.: número de publicações; Cit.: número de citações.

A revista *Remote Sensing of Environment* com maior número de artigos aceitos em seus editoriais (n = 33), iniciou suas publicações com o tema em 1994, com picos nos anos de 2011, 2013 e 2017, já as citações apresentaram pico em 2011, próximo de 734. Um exemplo da popularidade desta revista, é que da lista completa dos artigos relacionados ao tema, o mais citado foi publicado nesta mesma revista em 2011, se trata do artigo do Francês Le Toan et al. (2011) que foi citado mais de 316 vezes, mesmo que não expressivo, em revistas florestais.

Esse artigo expõe as características e definições da missão BIOMASS, projeto em fase de implantação que tem como meta o mapeamento global da biomassa com a banda P, o que possibilitará melhor entendimento no ciclo e fixação de carbono terrestre.

Nota-se que a maioria dos autores tem como área de estudo, locais do próprio país, com exceção do Reino Unido que pesquisa principalmente áreas no Brasil, e o Japão que pesquisa mais o país de Camboja (FIGURA 4).

FIGURA 4 - NÚMERO DE ARTIGOS E PAÍSES QUE TIVERAM SUAS ÁREAS ESTUDADAS PELOS PAÍSES DE MAIOR PRODUÇÃO CIENTÍFICA



FONTE: A autora (2020).

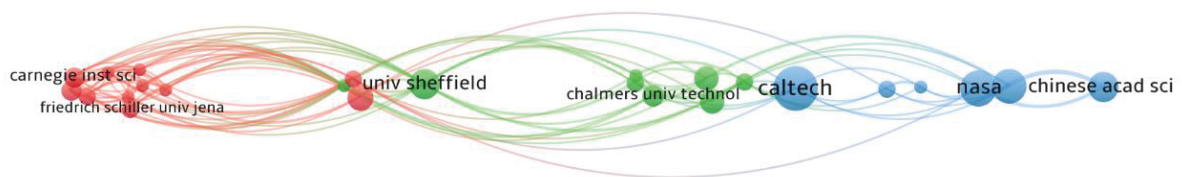
Os principais países detentores das áreas em estudo são os EUA (áreas com floresta boreal, decíduas, pinheiros), China (pesquisas realizadas em áreas de pradarias, floresta de folha larga, floresta de coníferas, mangue), Suécia (áreas ocupadas por floresta boreal e hemi-boreal), Noruega (floresta boreal e floresta com abeto) e Brasil (áreas com povoamentos de eucalipto e floresta tropical). Constatou-se que uma minoria dos estudos é realizada em florestas plantadas.

No total, 25 países são associados a autores que pesquisam sobre estimativa de biomassa florestal utilizando SAR. Destes, os principais países com maior produção de artigos e as influências sobre o tema são respectivamente: EUA (n = 27), China (n = 18), Reino Unido (n = 17), França (n = 10), Alemanha (n = 9), Noruega (n = 9), Japão (n = 9), e Suécia (n = 9).

O Brasil aparece em 11º lugar no *ranking* de produção, com cinco artigos publicados, sendo dos seguintes brasileiros: Santos et al. (2002), Kuplich, Curran e Atkinson (2005), Gama, Santos e Mura (2010), Santos, Gama e Conceição Bispo (2014) e Gama, Santos e Mura (2016). Estes autores pesquisaram áreas no próprio país, ocupadas por povoamento de eucalipto, floresta tropical e savana.

Na FIGURA 5 é possível observar a relação de coautoria das instituições de pesquisa, conforme listado na base de dados indexadora WoS e analisado no VOSviewer. Os dados corroboram com o representado na FIGURA 4, em que os EUA é o país de maior produtividade no assunto, sendo representado principalmente pelo *California Institute of Technology* (Caltech), a *National Aeronautics and Space Administration* (NASA) e *University of Maryland*. Em segundo lugar (China), sendo destacada pela Academia China de Ciências (*Chinese AcadSci*), a mais promissora em tal categoria.

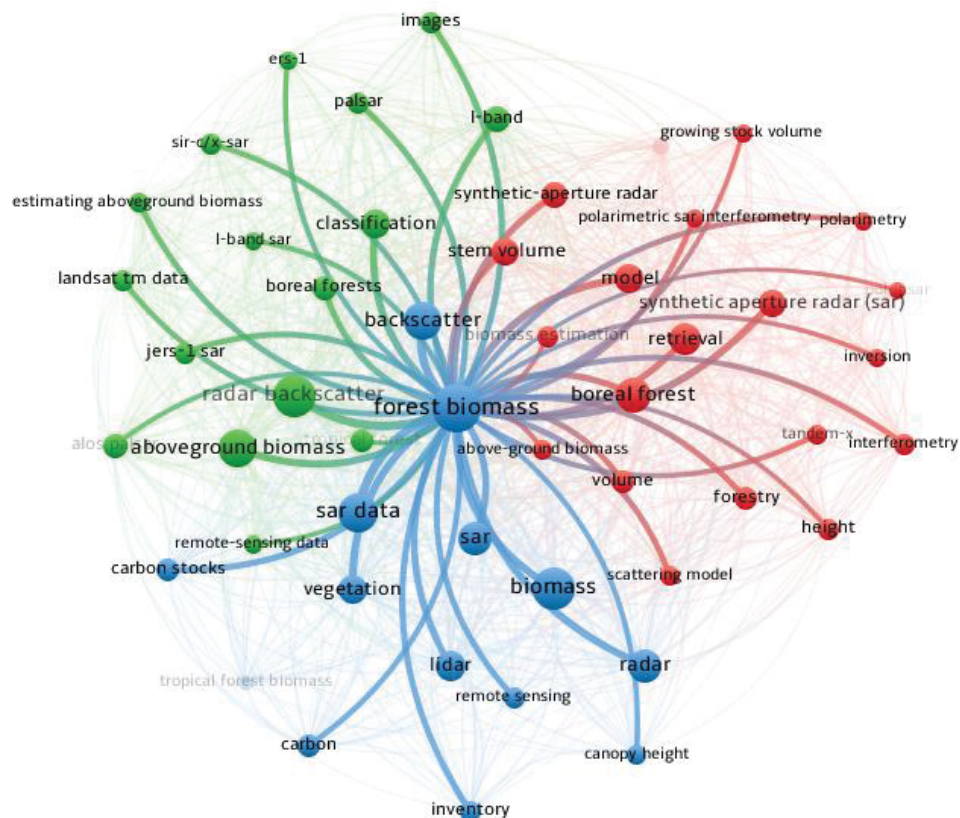
FIGURA 5 - MAPEAMENTO DA COAUTORIA DE INSTITUIÇÕES DE PESQUISA



FONTE: A autora (2020).

Foi utilizado o VOSviewer para analisar os termos mais usados no título e resumo das publicações sobre o tema, o resultado da análise das 35 palavras mais utilizadas está representado na FIGURA 6. O tamanho dos círculos está relacionado com a quantidade de ocorrência do termo, isto é, quanto mais utilizado o termo, maiores são os círculos. As cores representam os clusters e quanto menor a distância de sua ligação, mais forte é a relação entre os termos do cluster. A relação de termos é obtida pela contagem do número de vezes com que os termos ocorrem juntos nos títulos e resumos (RODRIGUES et al., 2014).

FIGURA 6 - MAPEAMENTO DA COCORRÊNCIA DOS 35 TERMOS MAIS CITADOS NOS ARTIGOS

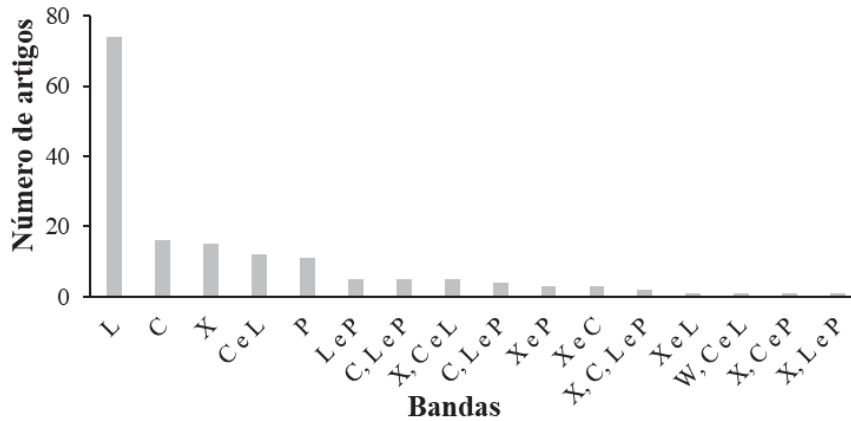


FONTE: A autora (2020).

Com o mapeamento da coocorrência dos termos, percebe-se que a palavra central é biomassa florestal, palavra utilizada na busca dos artigos sobre o tema. A partir desta tem várias ramificações em três principais clusters. Os ramos mais próximos com a palavra central apresentam relação mais forte com o tema, como o termo radar e retroespalhamento. Mais distantes do termo principal e, portanto, com menor relação ao tema, aparecem termos relacionados com a fisionomia florestal (*rain-forest*, *tropical forest*, *boreal forest*), ao satélite dos dados SAR, a utilização de dados *Light Detection and Ranging* (LiDAR) para validar as estimativas e dados auxiliares para a modelagem.

Fez-se um mapeamento das bandas SAR utilizadas em cada artigo e sua representação resumida pode ser observada na FIGURA 7.

FIGURA 7 - NÚMERO DE ARTIGOS CIENTÍFICOS PUBLICADOS COM CADA BANDA DO ESPECTRO ELETROMAGNÉTICO



FONTE: A autora (2020).

Observa-se que em aproximadamente 50% dos artigos, os autores utilizaram dados da banda L, podendo ser de forma isolada ou combinada com outras, com dados oriundos principalmente dos satélites ALOS/PALSAR e JERS-1. A segunda banda mais utilizada foi a frequência C (22%), seguida da banda P (14,3%) e banda X (13,9%). No total, 71% dos artigos utilizaram apenas uma frequência. Estudos baseados em SAR, adquiridos em duas ou três frequências foram relatadas em 14,7% e 10,4% dos trabalhos encontrados nas bases de dados selecionadas, respectivamente. Dados SAR de quatro frequência (bandas X, C, L e P), foram utilizados em dois artigos, correspondendo a 1,2%.

1.5 DISCUSSÃO

O comportamento da produção anual de artigos pode ser justificado pela lei de Price (PRICE, 1963) que descreve a existência de quatro fases que explicam o comportamento da pesquisa em determinado tema, sendo: i) fase de pioneiros, em que um pequeno grupo de cientistas começa a publicar um novo tema; ii) um número crescente de cientistas é atraído pelo tema e as lacunas, em virtude disto há um crescimento exponencial em determinado tema; iii) elucidação das lacunas do conhecimento; e ix) ocorre uma diminuição do número de publicações sobre o tema devido a elucidação do assunto. No entanto, é questionável que já tenha ocorrido a elucidação das lacunas envolvendo a utilização de SAR para a estimativa de biomassa florestal, já que essa é uma tecnologia que está em constante atualização e que necessita de quantificação e monitoramento para escala global.

O artigo mais antigo nas bases de dados analisadas, que se trata sobre estimativa de biomassa florestal utilizando SAR foi publicado no ano de 1991 por Hussin, Reich e Hoffer (1991), um pesquisador Norte-Americano que utilizou dados SAR de banda L para desenvolver um sistema de equações para predição da biomassa média do povoamento em função do *backscatter*, utilizando para o estudo de 55 parcelas de *Pinus sp.*, localizadas na Flórida. O sistema de equações desenvolvido pelos autores foi responsável por 97% da variabilidade observada na biomassa média do povoamento. Neste sentido, ao aplicarem as equações para um conjunto de dados independente, o desvio médio foi menor que 1% e aproximadamente 3% para o erro padrão.

Desde a publicação do primeiro artigo nas bases de dados estudadas (1991), o número de publicações apresentou tendência crescente até 2014. Este padrão de publicações pode ser justificado pelo lançamento de novos satélites e disponibilidade de dados SAR, como por exemplo com a operação dos sensores ALOS/PALSAR de 2006 a 2011, em que o número de estudos com a Banda L cresceu e teve pico em 2014, sendo utilizada em 17 de 24 artigos. Acredita-se que há uma correlação forte com a linha do tempo das diferentes missões, como por exemplo decréscimo na utilização de SAR com a popularização do LiDAR orbital e aerotransportado, da fotogrametria e acesso livre a imagens dos satélites LANDSAT e SENTINEL.

A maioria dos artigos foi escrita por mais de dois autores (83,6%), sendo possível observar que há um crescente desenvolvimento de pesquisas em parceria sobre o tema em questão. Um elevado número de publicações em coautoria indica uma relação mais próxima entre os autores dentro do mesmo tema e uma oportunidade para colaboração futura (WANG et al., 2014).

Existe um conceito de que o número de citações reflete a influência e notoriedade de uma publicação, conseqüentemente sua qualidade para a ciência (UGOLINI et al., 2015). No entanto, existem pesquisadores como Ho (2012) que afirmam que o número de citações indica apenas qual a visibilidade do estudo. Neste aspecto, os autores mais visíveis sobre o tema são Le Toan e Saatchi.

A revista mais popular sobre o tema é *Remote Sensing of Environment* e aparentemente ainda está na segunda fase da lei de Price, e com ascensão de artigos que tratam sobre estimativas de biomassa florestal utilizando dados SAR.

Os países estão investindo cada vez mais em pesquisa, com base na premissa de que essa atividade promove o desenvolvimento econômico e o crescimento de um país (BECKER, 2015).

Corroborando com os resultados dos principais países em número de artigos publicados, Liu et al. (2012) e Yang et al. (2013) afirmam que o padrão de domínio das sete economias mais avançadas do mundo de acordo com o Fundo Monetário Internacional (FMI), denominado G7, composto por EUA, Japão, Itália, Alemanha, Reino Unido, Canadá e França, tem ocorrido na maioria dos campos científicos e reflete na alta atividade econômica e o nível acadêmico dos países do grupo. Gaviria-Marin, Merigó e Baier-Fluentes (2019) ao realizarem uma revisão bibliométrica sobre gestão do conhecimento, constataram que os EUA lideravam o *ranking*, seguidos pelo Reino Unido. Mao et al. (2018) ao realizarem também um estudo bibliométrico sobre energia de biomassa e meio ambiente, evidenciaram que os EUA é o país mais influente, seguido pela China.

Um fato que pode justificar a lista de países de maior produção científica é que estes apresentam as primeiras colocações no *ranking* do Índice Global de Inovação (IGI) de 2018, sendo por ordem: Suécia (3^a posição no *ranking*), Reino Unido (4^a), EUA (6^a), Japão (13^a), França (16^a), China (17^a) e Noruega (19^a). O IGI leva em consideração 80 indicadores, que variam desde registro de propriedade intelectual, criação de aplicativos, gastos com educação e publicações científicas e técnicas.

Também é possível associar a produtividade científica dos países com o relatório das emissões de CO₂ com o uso de combustível fóssil e processos industriais de 2013, em que estes países estão na lista dos que mais emitiram CO₂ na seguinte ordem: China (1^a posição no *ranking*), EUA (2^a posição no *ranking*), Japão (6^a posição no *ranking*), Reino Unido (14^a posição no *ranking*) e França (20^a posição no *ranking*). Acredita-se que os primeiros colocados no ranking estejam realizando pesquisas para atender as medidas mitigadoras do Acordo de Paris.

Para o tema em questão, o Brasil aparece em 11^a posição no ranking em número de artigos, em outro tema, energia de biomassa e meio ambiente, o Brasil ficou na 8^a posição (MAO et al., 2018).

Ao agrupar os estudos de acordo com o conjunto de frequências em que os dados SAR foram adquiridos, ficou evidente que o comprimento de onda da banda L foi preferido, seguido pela banda C. Corroborando com os resultados deste estudo, Santoro e Cartus (2018) declararam que essa banda teve representatividade em 71% dos artigos listados em uma revisão bibliográfica sobre recuperação de AGB utilizando SAR. Acreditamos que esse fato esteja associado a disponibilidade gratuita de algumas imagens nessas frequências, tais como SENTINEL-1 e ALOS/PALSAR que esteve disponível por um período.

O radar de maior comprimento de onda apresenta maior capacidade de penetrar no dossel da floresta, capturando mais informações sobre a estrutura vertical. Estudos anteriores apontaram que os dados das bandas L e P são mais sensíveis à estimativa da biomassa do que os dados da banda C, por exemplo (GHASEMI et al., 2011). Isso ocorre porque a banda C apresenta comprimento de onda menor, o qual interage principalmente com elementos do dossel, sendo apropriada para locais de baixa biomassa. Em contraste, a banda L ou P de comprimento de onda longo pode interagir com ramos, tronco e solo sob o dossel da floresta, e é adequada para densidade de biomassa relativamente alta (SAATCHI et al., 2017).

1.5.1 Perspectivas para o Futuro

Com a análise bibliométrica foi possível mensurar os estudos já realizados, dando suporte para a identificação de lacunas do conhecimento. Para futuros estudos sobre a estimativa de biomassa florestal e carbono utilizando de dados SAR, ressalta-se a necessidade de explorar mais o potencial da banda P e de metodologias que visem a produção de resultados em escala global com menor incerteza sobre essas estimativas. Com a revisão, constatou-se que a banda P tem sido pouco empregada, motivado pelo seu alto custo e inexistência de imageamento por satélite, problema a ser resolvido em 2022 com o lançamento da missão BIOMASS (LE TOAN et al., 2011).

Dados de banda L eram disponibilizados gratuitamente pelo satélite ALOS/PALSAR, porém com o fim de sua vida útil, houve uma interrupção na disponibilidade de dados nessa frequência. Problema a ser resolvido com o lançamento do satélite NISAR previsto para o final de 2021, este será de dupla frequência (Banda S e L), revisita a cada 12 dias, quad-pol (HH, HV, VH, VV) e disponibilidade gratuita (DUNCANSON et al., 2020). Com a disponibilidade desses novos dados, necessita-se que as pesquisas continuem explorando os potenciais da banda L.

A maioria dos estudos utilizou de apenas uma frequência da faixa do espectro eletromagnético. Demonstrando a necessidade de se aliar dados SAR em diferentes comprimentos de onda a outras fontes, tais como imagens multiespectrais, hiperespectrais e LiDAR obtidos por satélite ou veículos aéreo não tripulados.

Embora este estudo tenha sido baseado em SAR, há outras opções baseadas em sensoriamento remoto que apresentam grande potencial para a estimativa da AGB e ainda são novidades com grande potencial, tais como o GEDI, MOLI e ICESat-2. Um dos produtos do ICESat-2 fornece estimativas de altura do terreno e do dossel (ATL08), os quais facilitarão as

avaliações e monitoramento da biomassa e carbono florestal em escalas globais (NEUENSCHWANDER; PITTS, 2019). O GEDI e MOLI realizam imageamento com um altímetro a laser, os quais prometem contribuir muito com as pesquisas.

Além disso, para as missões BIOMASS e NISAR, a ESA e a NASA estão em parceria no desenvolvimento de uma Plataforma de Análise Multi-Missão a qual permitirá o processamento e análise on-line de dados SAR (ALBINET et al., 2018). Algumas plataformas já estão disponíveis atualmente, como a Pipeline de processamento híbrido conectável da ASF (HyP3), Plataforma de exploração temática da ESA (TEPs) e o Google Earth Engine (GEE). Com essas plataformas os usuários de SAR ficaram livres de alguns pré-processamentos, menos propensos a erros e a desperdiçar recursos fora da área geográfica de interesse, assim como, facilidade de processamento de dados em larga escala.

Por fim, o SAR fornece recursos exclusivos para o monitoramento florestal, complementando os recursos fornecidos por sistemas ópticos de sensoriamento remoto, e, portanto, o setor florestal não deve considerar o uso do SAR como uma solução a ser considerada isoladamente. Desafios e oportunidades são abundantes na combinação de dados SAR com outras técnicas de sensoriamento remoto e baseados em dados de inventário para aplicações florestais e estimativa de biomassa.

1.6 CONCLUSÃO

As publicações sobre estimativa de biomassa florestal utilizando SAR foram iniciadas em 1991, com crescimento nos últimos anos, somando 159 artigos até o fim de 2018 que estão disponíveis nas bases de dados indexadoras *Web of Science* e Scopus, com soma de 4880 citações ao todo.

O país líder em produtividade é os EUA, sendo os estudos desenvolvidos principalmente no instituto Caltech, em segundo lugar está a China, representada principalmente pela Academia Chinesa de Ciências. Os autores de maior produtividade são Solberg, Saatchi e Sun. Apesar disso, Le Toan é o autor do artigo mais citado sobre o tema. O país que tem suas áreas mais estudadas é a China, a revista mais procurada para publicação é a *Remote Sensing of Environment* e a banda do espectro eletromagnético mais empregada nos estudos é a L.

Os países que mais realizam pesquisas sobre estimativas de biomassa florestal utilizando SAR são os países mais desenvolvidos e os que mais poluem. Observou-se que a maior parte dos estudos utiliza apenas uma banda, a qual é ineficiente dependendo da cobertura

florestal. A integração de dados de diferentes fontes deve ser uma alternativa que melhora a precisão das estimativas.

Para estudos futuros recomenda-se que sejam explorados mais características técnicas dos artigos, tais como: tipo de técnica SAR (PolSAR, InSAR, PolInSAR), evolução das técnicas de SAR para quantificação de variáveis na floresta, tipo de floresta analisada, métodos estatísticos empregados, uso de fusão de dados, variáveis explicativas que apresentaram o melhor desempenho, ocorrência ou não de saturação e em quais valores de biomassa a saturação foi observada.

CAPÍTULO 2: ESTIMATIVA DE VARIÁVEIS DENDROMÉTRICAS AO NÍVEL DE PARCELA OBTIDA PELA INTEGRAÇÃO DE SENSORIAMENTO REMOTO ATIVO E PASSIVO EM POVOAMENTOS DE *EUCALYPTUS* SPP.

Resumo: Tendo em vista a grande importância que o setor de florestas plantadas desempenha sobre a economia do Brasil e a influência que estas plantações podem exercer sobre a mitigação das mudanças climáticas, se faz necessária a quantificação dos estoques da floresta. A quantificação das métricas da floresta por sensoriamento remoto pode dar suporte ao adequado manejo dos recursos fornecidos pela floresta. O objetivo para este capítulo é modelar a área basal (G), volume total (V) e biomassa aérea a partir de dados de Radar Polarimétrico de Abertura Sintética (PolSar) e dados ópticos multiespectrais, comparando-se o desempenho de técnicas como a regressão linear e modelos de aprendizado de máquina. O estudo foi realizado com informações dendrométricas obtidas de 192 parcelas de *Eucalyptus* spp., localizados no interior do estado do Paraná. Utilizou-se de dados multiespectrais e índices de vegetação gerados a partir do satélite SENTINEL-2. Para compor o conjunto de dados PolSar, utilizou-se da banda C oriunda do satélite SENTINEL-1 e a banda L do satélite ALOS/PALSAR-2, assim como a textura obtida dessas bandas e índices polarimétricos obtidos com a banda L. Ajustou-se modelos de regressão linear para estimar a área basal, volume e biomassa aérea, tendo para isso cinco diferentes conjuntos de variáveis independentes (tratamentos), para o melhor tratamento, aplicou-se de técnicas de aprendizado de máquina para gerar as mesmas estimativas. Para a estimativa de área basal, a integração das bandas C e L produziram as melhores estimativas. Para o volume e biomassa aérea, em geral os erros foram maiores, sendo necessária a integração de todas as informações PolSar e multiespectral para a modelagem. Dentre as técnicas de aprendizado de máquina, o modelo *Random Forest* de sobressaiu sobre os demais. Ao comparar as estimativas do *Random Forest* com a regressão linear, o ganho foi de 8, 12 e 15% ao estimar respectivamente, a área basal, volume e biomassa. Pode-se concluir que com as informações e metodologias utilizadas é possível obter estimativas precisas, auxiliando assim a reduzir os custos com o inventário florestal e consequentemente o manejo da floresta.

Palavras-chave: SENTINEL, ALOS/PALSAR-2, aprendizado de máquina, regressão linear.

2.1 INTRODUÇÃO

Entre os mais evoluídos do mundo, o melhoramento genético brasileiro garantiu a alta produção das florestas plantadas de eucalipto. A área total de árvores plantadas no Brasil em 2018, era de 9,9 milhões de hectares, em que deste total, 7,5 milhões de hectares correspondiam a plantações de eucalipto (IBGE, 2019).

As plantações florestais são suprimentos indispensáveis para a redução da pressão sobre o desmatamento ilegal das florestas nativas, colaboração nos ciclos biogeoquímicos e geofísicos em que estão inseridos (DAI et al., 2017). Por apresentarem destaque no desenvolvimento de baixo carbono (IBÁ, 2019) e alta taxa de crescimento, o setor de florestas plantadas estoca uma grande quantidade de carbono em seus tecidos vegetais, cerca de 1,7

bilhão de toneladas de CO₂ eq¹ em 2018, tornando essas plantações elegíveis para participar de programas para mitigação das mudanças climáticas, proteger a biodiversidade e os serviços ecossistêmicos, como exemplo, temos o Programa de Redução de Emissões do Desmatamento e Degradação florestal (REDD+) (DUDLEY, 2009), e o Mecanismo de Desenvolvimento Limpo (MDL).

As mudanças climáticas proporcionam riscos, mas também oportunidades para o setor de florestas plantadas devido a sua relevância (IBÁ, 2019). Tanto para a viabilidade dos projetos de mitigação das mudanças climáticas, como para o adequado manejo e gerenciamento das plantações florestais, se faz necessária a quantificação dos estoques da floresta.

Os inventários baseados em campo constituem a abordagem tradicionalmente utilizada para a obtenção das métricas da floresta e conseqüentemente a sua quantificação. O inventário florestal é tido como a forma mais precisa, no entanto, esse método é demorado e apresenta limites operacionais em grandes áreas (BOISVENUE et al., 2016). Como ferramenta de suporte para os inventários florestais, o uso de sensoriamento remoto ajuda no mapeamento e extrapolação de dados para grandes áreas.

Englhart, Keuck e Siegert (2011) afirmam que os produtos obtidos por sensoriamento remoto, quando aliados aos dados de inventário, fornecem estimativas precisas de parâmetros da floresta adequadas ao manejo florestal. Correspondente ao sensoriamento remoto passivo, a aplicação de imagens multiespectrais foi a primeira tecnologia amplamente utilizada para o mapeamento da biomassa florestal, atribuindo-se a sua utilização devido a disponibilidade temporal e em grande parte gratuita (DOMINGUES et al., 2020). Apesar dos benefícios, estas também apresentam limitações, como a cobertura de nuvens, aerossóis, neblina e saturação em altos níveis de biomassa florestal (CAMPBELL; WYNNE, 2011). Estas limitações podem ser atenuadas com a sua integração com informações obtidas por outros sensores.

Os dados de Radar Polarimétrico de Abertura Sintética (PolSAR) são uma fonte de dados ideal para os mapeamentos dos atributos da floresta, pois consistem em: i) sistemas de sensoriamento remoto ativo, correspondendo à faixa do microondas do espectro eletromagnético, na qual o sinal do radar penetra no dossel da floresta; ii) por ser um sistema ativo, é capaz de capturar dados mesmo sem a energia solar, mesmo com a cobertura de nuvens, neblina, fumaça e poeira (MITCHARD et al., 2011).

O PolSar opera com a premissa de que as microondas interagem com as propriedades estruturais e dielétricas dos seus alvos, em que a energia emitida pelo sensor se retroespalha nos alvos, de acordo com sua rugosidade, retornando ao sensor, formando uma imagem com ruídos ou *speckle* (LEE; POTTIER, 2009). Quanto maior o comprimento de onda do radar

(banda L e P, aproximadamente 24 e 68 cm, respectivamente), maior a capacidade de penetrar no dossel da vegetação, e saturação apenas com elevada densidade de biomassa e, conseqüentemente, mais bem-sucedidos para a estimativa de variáveis dendrométricas (LE TOAN et al., 1992; MITCHARD et al., 2011).

Comprimentos de onda mais curtos (banda C, ~5,6 cm), tendem a interagir mais com pequenos galhos e folhas nas camadas superiores da floresta, isso dependendo do fechamento do dossel da floresta (LUCAS; MOGHADDAM; CRONIN, 2004). Supõem-se que para plantações florestais a utilização dessa banda traria ganhos para a modelagem até certa idade, e que sua integração com outras informações espaciais, agregariam informações valiosas. A insistência na utilização da banda C apesar de suas limitações, é que até o momento esse é o único conjunto de dados PolSar disponíveis gratuitamente e coletados sistematicamente.

Os dados obtidos na região do microondas e multiespectrais se complementam, os quais, quando combinados, podem levar a melhorias na modelagem de atributos da floresta. Os dados multiespectrais fornecem informações referentes às características químicas e biofísicas da superfície do dossel, ao passo que dados PolSar fornecem dados dielétricos, quando da existência de umidade, e estruturais da camada vertical da floresta (DOMINGUES et al., 2020).

A contribuição deste estudo é na investigação da utilização de cada banda separadamente e também com o seu uso integrado, buscando entender a contribuição que cada uma proporciona na estimativa das variáveis dendrométricas. Acredita-se que a integração de dados obtidos por sensoriamento remoto oriundos de diferentes fontes juntamente com metodologias adequadas, possibilitem ajustar modelos para estimar área basal, volume e biomassa aérea.

2.1.1 Aprendizado de máquina

São necessários mais esforços de pesquisa para explorar diferentes procedimentos estatísticos e métricas obtidas a partir de sensores remotos para a modelagem de variáveis biofísicas da floresta, principalmente quando se trata da alta complexidade das florestas tropicais. Técnicas não paramétricas de aprendizado de máquina, são versáteis na identificação de relações não lineares complexas e no tratamento de alta dimensionalidade dos dados, a exemplo podemos citar o *Weighted k-Nearest-Neighbor* (wKNN), *Support Vector Regression* (SVR), *Artificial Neural Network* (ANN), *Random Forest* (RF) e *Model Tree* (M5) (ALMEIDA et al., 2019). Tais técnicas apresentam o potencial de fornecer estimativas mais precisas de

AGB do que modelos de regressão linear, especialmente quando dados *multisource* são usados (LU et al., 2014).

O *k-Nearest-Neighbor* (KNN) representa uma das técnicas mais simples e intuitivas de estatística não paramétrica. Para a classificação de uma nova observação, esta é colocada junto com a classe da observação do conjunto de dados mais próximo a ela, com relação as covariáveis utilizadas. A determinação dessa similaridade entre o conjunto de dados é baseada em medidas de distância. Já o wKNN se baseia na ideia de que tais observações dentro do conjunto de dados, que são próximas da observação, devem ter um peso maior na decisão comparado aos vizinhos que estão distantes. Para conseguir atender esse objetivo, as distâncias, nas quais a busca pelo vizinho mais próximo se baseia no primeiro passo, devem ser transformadas em medidas de similaridade, que podem ser usadas como pesos (TAN, 2009; HECHENBICHLER; SCHLIEP, 2004).

O SVR é uma derivação de *Support Vector Machine* (SVM) aplicado a problemas de regressão, sendo uma técnica de aprendizado de máquina não paramétrica amplamente utilizada para problemas lineares. A ideia principal do SVR é traçar um hiperplano de máxima separação entre os dados, mapeando os dados de entrada e separando-os em um espaço de recurso de alta dimensionalidade a partir da semelhança das observações. Para cada problema tem-se uma função que melhor se adapta, dentre elas, a função de base radial tem demonstrado ser muito eficaz. Além da função, é necessário definir no ajuste de uma SVR, o parâmetro custo que controla a complexidade do limite entre os vetores e o sigma que é um parâmetro de suavização (JAMES et al., 2013).

A ANN é uma rede de camadas interconectadas que buscam simular a estrutura e funcionamento do cérebro humano em um computador. É constituída por três camadas, camada de entrada em que o número de neurônio depende do número de variáveis de entrada para o modelo, uma ou mais camadas ocultas que não tem número de neurônios fixo e camada de saída que corresponde ao valor que se deseja estimar. Os sinais de entrada são transferidos da camada de entrada para a camada de saída através da camada oculta com ajuda da função de ativação, peso, e dependem da taxa de aprendizado. Em redes do tipo *multi layer perceptron* é possível utilizar do algoritmo de aprendizado *backpropagation* para ajustar os pesos das conexões e chegar a um resultado mais próximo do esperado. Isso é possível quando na camada de saída é comparado o valor estimado com o observado, se não atender ao limite de erro estabelecido, o erro é retropropagado para as camadas ocultas pelo qual é feito o ajuste dos pesos sinápticos (MCCULLOCH; PITTS, 1943; ROSENBLATT, 1958; MINSKY; PAPERT, 1969; WITTEN; FRANK; HALL, 2011). Os precursores do uso das ANN na área florestal no Brasil são de

autoria de Gorgens et al. (2009), Binoti (2010), Sanquetta et al. (2013), Binoti et al. (2014), Gorgens et al. (2014), Binoti et al. (2015) e Sanquetta et al. (2015).

O RF é um método de aprendizado de máquina conjunto que combina previsões (por média) de múltiplas árvores de classificação ou regressão, dependendo do problema abordado. Cada árvore é criada independentemente a partir de uma amostra de inicialização dos dados originais (uma abordagem de empacotamento). O princípio dessa abordagem é construir uma infinidade de árvores (*n-tree*) e cada uma representa um processo de aprendizado. Cada nó da árvore é dividido usando um número específico de recursos selecionados aleatoriamente (*m-try*). Cada árvore pode selecionar aleatoriamente recursos e amostras. A previsão final é obtida calculando a média das previsões das árvores (BREIMAN, 2001). A técnica RF é reconhecida por ser precisa mesmo quando várias covariáveis são usadas, principalmente porque ela não produz estimativas tendenciosas ou leva ao sobreajuste (QUINLAN, 1993; HENGL et al., 2018).

O M5 é utilizado para induzir uma árvore modelo. O critério de divisão para o algoritmo M5 é baseado em tratar o desvio padrão dos valores da classe que atingem um nó como uma medida do erro nesse nó. O cálculo da redução esperada para esse erro é como um resultado de teste de cada atributo desse nó. Depois de avaliar todos os grupos possíveis, o algoritmo escolhe aquele que minimiza o erro previsto (QUINLAN, 1992). A principal vantagem da árvore modelo sobre a árvore de regressão é que as árvores modelos são muito menores do que as de regressão, a decisão é clara e as funções de regressão normalmente não envolvem muitas variáveis (QUINLAN, 1992).

2.2 OBJETIVOS

O objetivo geral desse capítulo é analisar o desempenho em estimar variáveis dendrométricas (área basal, volume e biomassa aérea) de povoamentos de eucalipto com diferentes idades, a partir de dados PolSar. De maneira específica, os objetivos são:

- 1) Avaliar o desempenho das estimativas das variáveis dendrométricas utilizando-se de representações da banda C, da banda L e da integração dessas;
- 2) Verificar se a integração da textura e índices polarimétricos contribuem para a modelagem;
- 3) Avaliar o desempenho das estimativas ao integrar diferentes fontes (PolSar e multiespectral) para a estimativa das variáveis dendrométricas;

4) Avaliar o desempenho das estimativas utilizando apenas as variações de representação de dados multiespectrais;

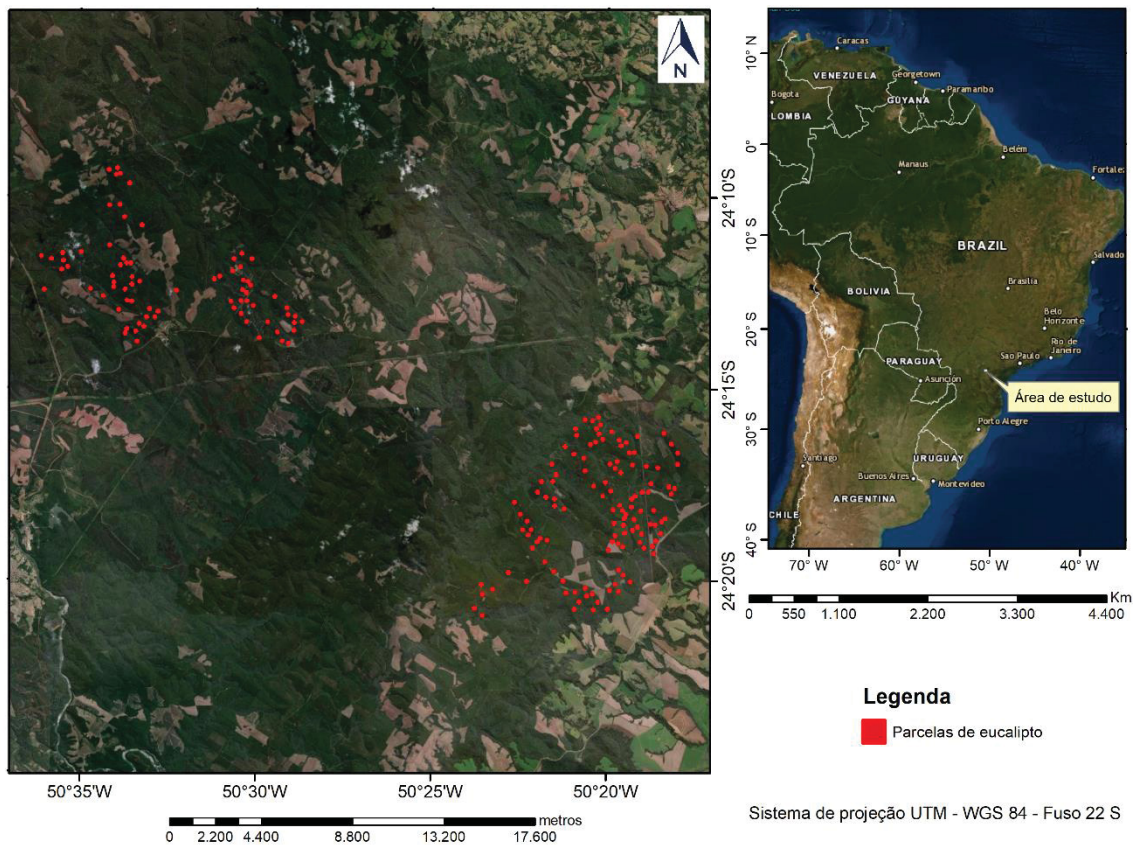
5) Comparar as estatísticas de desempenho dos modelos de regressão linear e aprendizado de máquina.

2.3 MATERIAL E MÉTODOS

2.3.1 Área de estudo

Para este estudo foram consideradas 192 parcelas de eucalipto pertencentes à empresa florestal Klabin S.A., localizadas no município de Telêmaco Borba – PR (FIGURA 8).

FIGURA 8 – LOCALIZAÇÃO ESPACIAL DAS PARCELAS UTILIZADAS NESTE ESTUDO



FONTE: A autora (2020).

A mesorregião em que está inserida a área de estudo é caracterizada pela presença de solos pouco férteis e declividades acentuadas, favorecendo a expansão de florestas com consequente diminuição de áreas agrícolas (IPARDES, 2004). Na porção onde está localizada

as parcelas, o solo é classificado como Argissolo Vermelho-Amarelo distrófico e Latossolo Vermelho Distrófico, com declividade média de 36° (IPARDES, 2004).

O clima do município é caracterizado como uma área de transição climática entre Cfa (clima subtropical) e Cfb (clima temperado). O tipo Cfa segundo a classificação de Köppen, é úmido, mesotérmico e sem estação seca definida. Os verões são quentes e com precipitação concentrada em alguns locais, e no inverno ocorre com pouca frequência geadas. O clima Cfb, caracterizado por verões frescos e sem estação definida, ocorre devido às áreas de alta declividade existentes nas extremidades do município (ITCG, 2008).

Com maior proximidade à área de estudo e de responsabilidade da entidade Águas Paraná, está instalada a estação pluviométrica Ventania (2450034) localizada aos 24°30'39''S e 50°13'29''W, com distância aproximada de 14 a 31 km. Na outra extremidade da área de estudo, a estação pluviométrica Ortigueira (2450008) localizada aos 24°12'34'' S e 50°55'28'' W, com distância aproximada entre 40 e 60 km, dependendo do agrupamento de parcelas analisados.

2.3.2 Características e mensuração das parcelas amostrais

As 192 parcelas consideradas nesse estudo são parcelas de inventário florestal contínuo de *Eucalyptus saligna* Smith (material genético: CL2864) e *Eucalyptus urograndis* (cruzamento de *Eucalyptus grandis* W. Hill ex. Maiden com *Eucalyptus urophylla* S.T. Blake, material genético: CLI 042, CLI 144, CLI 224), com idade variando de 1 a 10 anos.

O inventário florestal foi realizado pela empresa detentora das plantações, nos meses de fevereiro, março e abril de 2019, no qual mensuraram o diâmetro à altura do peito (DAP) de todos os indivíduos da parcela e a altura total de aproximadamente 25% das árvores com maior diâmetro. As parcelas apresentam espaçamento variável entre árvores, por isso o inventário foi constituído da alocação de parcelas também com tamanho variável, seguindo a orientação de seis linhas de plantio por 10 árvores, com área da parcela variando de 348 à 885 m².

Além da mensuração das árvores, a empresa forneceu as coordenadas do centro da parcela, as quais foram coletadas com receptor GNSS (*Global Navigation Satellite System*) geodésico, este receptor apresenta precisão centimétrica. A partir da coordenada precisa, das dimensões de cada parcela e com imagens de satélite de alta resolução, foi possível desenhar a parcela retangular e orientá-la sobre a plantação para posteriormente extrair as informações obtidas por sensoriamento remoto.

2.3.2.1 Processamento das variáveis dendrométricas

A empresa detentora dos povoamentos forneceu a base de dados processada por parcela contendo as informações de área basal, altura média, volume com e sem casca, densidade básica da madeira e outras informações complementares sobre a plantação. A altura total de todos os indivíduos foi obtida a partir de equação ajustada com os indivíduos mensurados na respectiva parcela. As variáveis de volumes e densidade básica da madeira de todos os indivíduos foram obtidas por equações previamente ajustadas com banco de dados para a mesma espécie e idade. Os modelos empregados para o ajuste da altura, volume e densidade básica da madeira não foram fornecidos por motivos de confidencialidade da empresa.

A partir do Volume total com casca (V ($\text{m}^3 \cdot \text{ha}^{-1}$)), densidade básica da madeira (MEB ($\text{g} \cdot \text{cm}^{-3}$)), e Fator de Expansão de Biomassa (FEB), pode-se obter a biomassa acima do solo (AGB) (Equação 1). Considerou-se o FEB como sendo 1,05 para o gênero *Eucalyptus* (CORTE et al., 2015), valor único para todas as idades devido a indisponibilidade de outras equações de maior detalhamento, fato este considerado como uma limitação na obtenção da AGB.

$$AGB = V * MEB * FEB \quad (1)$$

2.3.3 Aquisição e processamento de dados espaciais

Além das variáveis dendrométricas das parcelas de *Eucalyptus* spp. utilizou-se de dados obtidos a partir de imagens de satélite (SENTINEL-1, SENTINEL-2, e ALOS/PALSAR-2), para compor informações espectrais com comprimento de onda na região do visível (490 – 665 nm), infravermelho (842 – 1375 nm) e microondas (5 – 24 cm).

Considerou-se parcelas com as mesmas dimensões e localização constantes no inventário, em que para esse tamanho de parcela foram extraídas as informações dos pixels das imagens correspondentes e, para esse conjunto de valores por parcela, obtidos somente a média e todos os pixel apresentaram o mesmo peso. Levou-se em conta apenas a média dos valores e mesmo peso na representação da parcela por se tratar de plantações homogêneas, em que ocorre pouca variância das variáveis dendrométricas e conseqüentemente da resposta dos alvos obtido por sensoriamento remoto.

2.3.3.1 Dados SENTINEL-1 e SENTINEL-2

Foram utilizadas imagens obtidas pelo satélite SENTINEL-1 (PolSar) e SENTINEL-2 (multiespectral), com as especificações constantes na TABELA 3.

TABELA 3 – ESPECIFICAÇÕES DAS IMAGENS DOS SATÉLITES SENTINEL-1 E SENTINEL-2

SENTINEL-1			SENTINEL-2		
Data de aquisição	26-04-2019	20-05-2019	Data de aquisição	18-04-2019	
Satélite	SENTINEL-1 B	SENTINEL-1 B	Satélite	SENTINEL-2 A	
Modo do instrumento	IW	IW	Instrumento	MSI	
Órbita	24	24	Número da órbita	19955	
Ponto	671	671	Identificador	2015-028A	
Órbita relativa	15975	16325	Direção do satélite	Descendente	
Polarização	VV, VH	VV, VH	Nível de processamento	1C	
Direção do satélite	Descendente	Descendente			
Tipo de produto	GRD	GRD			

FONTE: A autora (2020).

As imagens do satélite SENTINEL-2 A foram adquiridas pelo Centro de Dados Científicos da agência Copernicus Sentinels (ESA, 2016), com nível de processamento 1C, com correção radiométrica e geométrica, restando a correção atmosférica que foi realizada no software Qgis 2.18.16, método DOS1 (*Dark Object Subtraction*). Das 13 bandas totais, 4 delas correspondem à resolução espacial de 10 m, 6 bandas com resolução de 20 m, e 3 bandas de 60 m. A partir da imagem multiespectral, gerou-se índices a fim de ressaltar a vegetação, sendo utilizados os apresentados na TABELA 4.

TABELA 4 – ÍNDICES DE VEGETAÇÃO APLICADOS AS BANDAS DO SATÉLITE SENTINEL-2 A

Índice	Formulação	Autores
Índice de Vegetação (RVI)	$RVI = \frac{\rho IV}{\rho V}$	Allen; Tasumi; Trezza (2002)
Índice de Vegetação Normalizada (NDVI)	$NDVI = \frac{(\rho IV - \rho V)}{(\rho IV + \rho V)}$	Allen; Tasumi; Trezza (2002)
Índice de Vegetação Ajustado ao Solo (SAVI)	$SAVI = \frac{(1 + L)(\rho IV - \rho V)}{(\rho IV + \rho V + L)}$	Huete (1988)
Índice de área foliar (IAF)	$IAF = -\frac{\ln\left(\frac{0,69 - SAVI}{0,59}\right)}{0,91}$	Allen; Tasumi; Trezza (2002)

FONTE: A autora (2020).

LEGENDA: ρIV : reflectância na região do infravermelho próximo, banda 8 do satélite SENTINEL-2 (10 m de resolução espacial); ρV : reflectância na região do vermelho, banda 4 do satélite SENTINEL-2 (10 m de resolução espacial); L: ajuste na linha do solo considerado como 0,5.

O índice de área foliar descrito por Trezza (2002) não se trata do mesmo valor de IAF que seria obtido no interior da floresta já que o índice foi desenvolvido a partir de dados de culturas agrícolas. As bandas utilizadas para o cálculo dos índices de vegetação apresentam mesma resolução espacial e para as demais bandas com diferentes resoluções, utilizou-se a média dos pixels correspondente a cada parcela alocada em campo.

Sabendo-se que as medidas oriundas do radar dependem da estrutura, propriedades dielétricas dos componentes da vegetação e da superfície subjacente do solo, os quais são influenciados fortemente pela precipitação (WOODHOUSE, 2006), as cenas PolSar foram escolhidas devido a inexistência de precipitação nos sete dias anteriores a passagem do satélite. Estas foram adquiridas na plataforma *Hybrid Pluggable Processing Pipeline* (HyP3), as quais são fornecidas com correção radiométrica do terreno (RTC - GAMMA) (HOGENSON et al., 2016).

O algoritmo RTC usa o software GAMMA para criar produtos de imagens PolSar corrigidos, tanto geometricamente quanto radiometricamente. O algoritmo usa um Modelo Digital de Elevação (DEM) que cobre as imagens, para criar uma imagem de radar simulada. Essa imagem simulada é então combinada com a imagem de radar real para criar um mapeamento preciso do espaço radar para o espaço DEM (nesse caso, projeção UTM – Universal Transversa de Mercator). Esse mapeamento é usado para mover todos os pixels para um produto geocodificado. Após a geocodificação, uma correção radiométrica é aplicada usando a abordagem de integração da área de pixels (SMALL, 2011).

A calibração radiométrica do terreno visa uma interpretação correta das assinaturas de retroespalhamento, corrigir os efeitos do ângulo de incidência local devido à topografia e normalização da área real de pixels (SMALL, 2011).

Nas imagens obtidas pelo HyP3, foi aplicado o filtro polarimétrico *Lee Filter* com janela de 5 x 5, para a minimização do efeito *speckle*. Ainda com esse objetivo, foi realizada a média entre as mesmas polarizações (VV e VH) das duas cenas, ambos os processamentos realizados no software Envi 5.3.

Como informação adicional ao retroespalhamento, gerou-se imagens de textura baseadas na Matriz de Co-ocorrência de Nível de Cinza (GLCM) para cada polarização (VV e VH) sendo as seguintes texturas: contraste, dissimilaridade, homogeneidade, segundo momento angular (ASM), entropia, média, variância e correlação (HARALICK, 1979). As texturas foram geradas com tamanho de janela de 5 x 5 e calculadas em ambas as direções.

Os valores de retroespalhamento para as polarizações VV e VH foram convertidos para decibéis (dB) a partir da Equação 2.

$$\sigma^0 = 10 * \log_{10}(DN^2) - CF \quad (2)$$

Em que: σ^0 : é o coeficiente de retroespalhamento sigma 0 em dB; DN: é a amplitude do sinal em cada polarização; CF: fator de calibração, o qual depende do ângulo de incidência, sendo 0 para imagens geradas a partir do satélite SENTINEL-1.

2.3.3.2 Dados ALOS/PALSAR-2

O sistema sensor *Phased Array L-band Synthetic Aperture Radar-2* (PALSAR-2) a bordo do satélite de observação terrestre *Advanced Land Observing Satellite-2* (ALOS-2) foi desenvolvido pela Agência Espacial Japonesa e lançado em 2014. O ALOS/PALSAR-2 representa a continuidade de um programa de observação da Terra com o uso de radares de abertura sintética em banda L, iniciado pelo *Japanese Earth Resource Satellite* (JERS-1) que esteve em funcionamento de 1992 a 1998, e o ALOS/PALSAR de 2006 a 2011 (JAXA, 2015).

Os modos de observação do ALOS/PALSAR-2 apresentam resolução espacial de 1 x 3 (*Ultrafine*) a 100 m (Nominal), faixa de imageamento de 25 (*Spotlight*) a 490 km (*Wide*), diferentes modos de polarização e resolução temporal de 14 dias (TABELA 5).

TABELA 5 - CARACTERÍSTICAS DO SISTEMA ALOS/PALSAR-2

Modo de observação	Resolução	Faixa de imageamento	Largura de banda	Polarização	Ângulo de incidência	
Strip map	Spotlight (SPT)	1 m (Az) x 3 m (Rg)	25 x 25 km (Rg)	84 Mhz	SP	8-70°
	Ultrafine (SM1)	3 m	50 km	84 Mhz	SP/DP	8-70°
	High Sensitive (SM2)	6 m	70 km	42 Mhz	SP/DP/FP	8-70°
	Fine (SM3)	10 m	70 km	28 Mhz	SP/DP/FP	8-70°
Scan SAR	Nominal (WD1)	100 m (3 looks)	350 km	14-28 Mhz	SP/DP	8-70°
	Wide (WD2)	60 m (1,5 looks)	490 km	14 Mhz	SP/DP	8-70°

FONTE: A autora (2020).

LEGENDA: SP (*single polarization*): HH, VV ou HV; DP (*dual-polarization*): HH+HV ou VV+VH; FP (*full-polarization*): HH, HV, VH e VV.

Para este estudo foram adquiridas três cenas, no modo de observação *High Sensitive* com resolução espacial de 6 m, *full-polarimétricas* (HH, HV, VH e VV), no formato CEOSAR e nível de processamento 1.1 (TABELA 6).

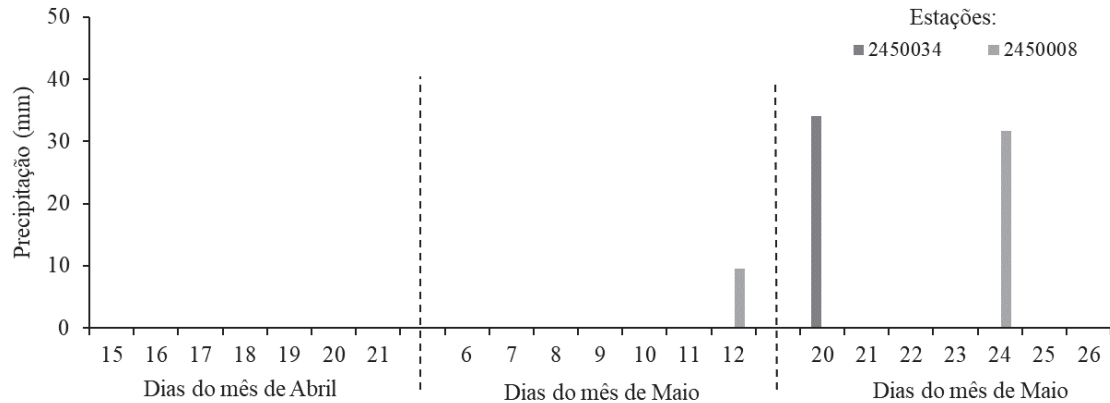
TABELA 6 - PARÂMETROS DAS CENAS UTILIZADAS E ADQUIRIDAS PELO SATÉLITE ALOS-2 SENSOR PALSAR-2

Cena	ALOS2265184070-190421	ALOS2268216700-190512	ALOS2270286700-190526
Data de aquisição	21-04-2019	12-05-2019	26-05-2019
Hora de aquisição	15:23 h	03:48 h	03:48 h
Satélite/sensor	ALOS/ PALSAR-2	ALOS/ PALSAR-2	ALOS/ PALSAR-2
Órbita de aquisição	Descendente	Ascendente	Ascendente
Direção de aquisição	Esquerda	Direita	Direita
Resolução do pixel (range x azimute)	3,13 x 2,86	3,13 x 2,86	3,13 x 2,86
Órbita / Ponto	125/4070	20/6700	20/6700
Polarização	HH +HV+VH+VV	HH +HV+VH+VV	HH +HV+VH+VV
Ângulo de incidência	32,7°	25°	25°
Nível de processamento	1.1	1.1	1.1
Modo de operação	SM2	SM2	SM2

FONTE: A autora (2020).

Buscou-se por cenas com data de aquisição próximo a data do inventário florestal, com menor precipitação acumulada, considerando os últimos 7 dias a passagem do satélite (FIGURA 9).

FIGURA 9 – PRECIPITAÇÃO DIÁRIA EM MM DOS SETE DIAS ANTECEDENTES A AQUISIÇÃO DAS CENAS



FONTE: A autora (2020).

A precipitação anterior à data de aquisição das imagens do satélite ALOS/PALSAR-2 foi praticamente nula, para a estação meteorológica mais próxima à área de estudo (código: 2450034) registrou 34 mm seis dias antes da aquisição da última cena. A segunda estação registrou volume de precipitação considerável em duas das cenas, porém acredita-se que devido a sua considerável distância com a área de estudo, esse volume precipitado não afetou a constante dielétrica dos alvos.

2.3.3.2.1 Processamento dos dados ALOS/PALSAR-2

Os processamentos das cenas do ALOS/PALSAR-2 foram realizados com o software SNAP (*Sentinel Application Platform*) da ESA, sendo os procedimentos de calibração radiométrica, *multilooking*, transformação para matriz de covariância (C3), filtragem polarimétrica, geocodificação, geração de texturas, índices polarimétricos e conversão do retroespalhamento para valores em decibéis (Equação 2).

A calibração radiométrica consiste na transformação dos números digitais da imagem para valores de retroespalhamento por unidade de área no alcance do terreno. A calibração radiométrica tem como objetivo ajustar o valor dos pixels para que estes correspondam diretamente à intensidade do sinal do radar, o retroespalhamento. Com as imagens calibradas radiometricamente, foi possível comparar imagens PolSar adquiridas com sensores, modos, processadores ou datas distintas.

O *multilook* foi realizado a fim de produzir imagens com dimensões geométricas regulares e reduzir o efeito do *speckle*, o qual é característico das imagens PolSar e influenciam

as informações radiométricas. Para as cenas do ALOS/PALSAR-2 o fator de reamostragem em azimute e alcance (*range*) foi fixado em 2:1.

Após o *multilook*, transformou-se os dados em uma matriz de covariância polarimétrica (matriz C3), essa matriz conserva em sua diagonal principal os valores para cada polarização, em que C11 corresponde à coerência para a polarização HH, C22 as polarizações HV/VH e na posição C33 da matriz a polarização VV (LEE; POTTIER, 2009). Neste estudo os elementos da matriz C3 foram renomeados para corresponder a banda a qual foi gerada, banda L, sendo chamados, por exemplo, de L11, L22 e L33.

A partir da matriz covariância, aplicou-se a filtragem polarimétrica que é uma etapa fundamental no processamento e que pode influenciar na maior relação com as variáveis dendrométricas. A filtragem polarimétrica necessita de um equilíbrio entre a redução do *speckle* e a perda de informação espacial, em que filtros de maior dimensão tendem a suavizar o *speckle* e os menores preservam as informações texturais (LEE; POTTIER, 2009).

Para a filtragem polarimétrica, o filtro *Refined Lee* (LEE; GRUNES; GRANDI, 1999) foi escolhido por ser um filtro adaptativo, o qual considera normalmente para sua análise uma janela com tamanho maior, para assim reconhecer os padrões em oito direções cardeais (LEE; POTTIER, 2009). Além disso, foi o filtro com melhor desempenho no estudo de Cassol (2017), quando avaliado em imagens ALOS/PALSAR-2 em floresta tropical.

Depois da retirada do *speckle*, gerou-se os índices polarimétricos descritos a seguir:

1) Índice de Degradação de Radar (*Radar Forest Degradation Index* - RFDI) por Mitchard et al. (2012). Calcula a degradação florestal a qual é sentida pelo aumento da importância relativa do retroespalhamento na polarização HH (σ_{HH}^0).

$$RFDI = \frac{\sigma_{HH}^0 - \sigma_{HV}^0}{\sigma_{HH}^0 + \sigma_{HV}^0} \quad (3)$$

2) Índices de Pope, Rey-Benayas e Paris (1994) (Equação 4), os quais se relacionam com diferentes características da vegetação, sendo, CSI corresponde à estrutura do dossel; BMI se refere a biomassa florestal; e o VSI ao espalhamento volumétrico.

$$CSI = \frac{VV}{VV+HH} \quad BMI = \frac{VV+HH}{2}; \quad VSI = \frac{CS}{CS+BMI} \quad (4)$$

Em que: VV, HH, VH e HV = se refere as respectivas polarizações; CS = média entre as polarizações HV e VH. Os valores das polarizações devem estar na escala linear.

3) Coeficiente de retroespalhamento da floresta (NGUYEN et al., 2016) (Equação 5), que tem como vantagem não apresentar a saturação do sinal para altos valores de biomassa e ter um significado físico na resposta desse índice.

$$\sigma_{forest}^0 = \sigma_{HV+\sigma_{HH}^0}^0 * \frac{\sigma_{HH}^0}{\sigma_{HV}^0} \quad (5)$$

Além dos índices polarimétricos, gerou-se índices de textura para cada polarização (HH, VV, HV/VH) computados com a mesma dimensão de janela selecionada para a redução do *speckle* (5 x 5), sendo as texturas baseadas na matriz de Co-ocorrência de Nivel de Cinza (GLCM): média, variância, homogeneidade, contraste, dissimilaridade, entropia, segundo momento angular (ASM) e correlação (HARALICK, 1979).

O último processo consistiu na realização da geocodificação (ortorretificação) das cenas, operação também conhecida como *Range Doppler Terrain Correction* utilizando o modelo digital de elevação do *SRTM 3Sec* como base. Esse procedimento é necessário devido às distorções nas distâncias das cenas, causadas por variações topográficas e a inclinação do sensor de satélite. A geocodificação busca compensar essas distorções para que a representação geométrica das cenas seja o mais próximo possível do mundo real (SCHREIER, 1993).

Juntamente com o processo de geocodificação, realizou-se a normalização radiométrica por γ^0 (γ^0 , em dB), utilizando o ângulo de incidência local (θ) (Equação 6) (JAXA, 2015).

$$\gamma^0 = \sigma^0 / \cos \theta \quad (6)$$

Em que o coeficiente de retroespalhamento σ^0 (σ^0) para cada uma das polarizações das cenas em nível de processamento 1.1.

Existem várias representações para o coeficiente de retroespalhamento do radar, tais como, σ^0 , γ^0 e β^0 . Utilizou-se γ^0 com correção de inclinação do terreno pois pode reduzir a dependência do ângulo de incidência do retrodispersor de radar para um alvo difuso (SHIMADA, 2010), além de ser equivalente ao processamento realizado com as polarizações da banda C.

2.3.4 Modelagem para a estimativa de variáveis dendrométricas

2.3.4.1 Regressão linear múltipla

A fim de atender aos objetivos, foram pré-definidos cinco tratamentos compostos por conjuntos de variáveis de entrada para a estimativa de G, V e AGB, sendo eles:

- 1) Idade e banda C: VV, VH e texturas. Total de 17 variáveis;
- 2) Idade e banda L: matriz C3, índices polarimétricos e textura. Total de 39 variáveis;
- 3) Idade, banda C e banda L. Total de 55 variáveis;

4) Idade, tratamento PolSar de melhor desempenho, bandas do satélite SENTINEL-2 e índices de vegetação. Total de 72 variáveis;

5) Idade, bandas do satélite SENTINEL-2 e índices de vegetação. Total de 18 variáveis;

Aplicou-se a seleção das variáveis de cada tratamento a partir do processo de *stepwise* por meio das funções “stepVIF” (pacote “*Pedometrics*”, SAMUEL-ROSA et al., 2014) seguido de “stepAIC” (pacote “*MASS*”, RIPLEY et al., 2002) no software R versão 3.6.1 (R DEVELOPMENT CORE TEAM, 2019), para cada tratamento, a fim de excluir variáveis multicolineares e simplificar o modelo, sem afetar o seu desempenho.

A função stepVIF calcula o impacto sobre a variância de cada variável como consequência das correlações de outros regressores presentes, para isso, usa um modelo linear e seleciona o subconjunto de variáveis preditivas que atendem a um limite de colinearidade específico do usuário, medido pelo fator de inflação de variação (VENABLES; RIPLEY, 2002).

Já a função stepAIC escolhe o melhor modelo da classe de acordo com o critério de informação de Akaike (AIC), inicia-se com o modelo abrangendo todas as variáveis independentes e o modelo final é composto por variáveis que apresentam o menor valor de AIC. O AIC é menor de acordo com um menor número de variáveis do modelo (VENABLES; RIPLEY, 2002).

Tendo as variáveis selecionadas para cada tratamento, ajustou-se equações de regressão linear múltipla com validação cruzada por *k-folds*, tendo definido $k = 10$. A validação cruzada é empregada para avaliar a capacidade de predição de um modelo. Consiste em dividir o banco de dados em subconjuntos, sendo parte destinada para a estimação dos parâmetros do modelo e outro para a validação. A validação cruzada com particionamento dos subconjuntos em *k-folds*, consiste em dividir o banco de dados em k subconjuntos do mesmo tamanho, em que um subconjunto é utilizado para teste e os $k-1$ restantes são utilizados para a estimação dos parâmetros. Esse processo é repetido k vezes, em que no final ter-se-á k medidas da métrica de avaliação escolhida, no qual utilizou-se a média.

2.3.4.2 Aprendizado de máquina

Para o tratamento de dados de entrada da regressão linear com melhor desempenho verificado pelas estatísticas de ajuste e desempenho, aplicou-se de técnicas de aprendizado de máquina. Com o intuito de comparação com a modelagem tradicional, utilizou-se das mesmas variáveis selecionadas pelo processo de *stepwise*.

Foram selecionados cinco modelos de aprendizado de máquina, sendo eles, *Model Tree* (M5), *Random Forest* (RF), *Weighted K-Nearest Neighbor* (wKNN), *Artificial Neural Network* (ANN) e *Support Vector Regression* (SVR) (TABELA 7). Todos os algoritmos desses modelos foram implementados no pacote Caret (KUHNS, 2008) do *software* R, com o auxílio do *script* disponibilizado por Souza (2020).

TABELA 7 - MODELOS DE APRENDIZADO DE MÁQUINA UTILIZADOS PARA A MODELAGEM DA AGB DA FLORESTA TROPICAL

Modelo	Amplitude de Hiperparâmetros	Método	Pacote do R	Autor
<i>Model Tree</i> (M5)	Poda = sim; não	M5	RWeka	Hornik; Buchta; Zeileis (2009)
<i>Random Forest</i> (RF)	Mtry = (1:21); ntree = seq (10, 500,50)	rf	randomForest	Liaw; Wiener (2002)
<i>Weighted K-Nearest Neighbor</i> (wKNN)	Kmax = seq (1, 150, 2); distance = 1:3; Kernel = rectangular, biweight, cos, epanechnikov, gaussian, inv, optimal, rank, triangular, triweight	kknn	kknn	Schliep; Hechenbichler; Lizee (2016)
<i>Artificial Neural Network</i> (ANN)	Size = seq (1, 10, 1); Decay = seq (0.01, 0.9,0.3)	nnet	nnet	Venables; Ripley (2002)
<i>Support Vector Regression</i> (SVR)	Sigma = seq (0.0001, 10, 0.1); C = seq (1, 100, 5)	svmRadial	Kernlab	Karatzoglou; Smola e Hornik (2004)

FONTE: A autora (2020).

Antes do processo de treinamento, os dados foram padronizados para uma mesma escala, onde para cada observação subtraiu-se o valor médio para a respectiva variável e dividiu-se pelo desvio padrão, assim todas as observações apresentaram média igual a 0 e desvio padrão igual a 1. Para o treinamento dos modelos foi necessário o ajuste dos hiperparâmetros, este se deu a partir dos limites estabelecidos na tabela acima. Para os hiperparâmetros de melhor desempenho, realizou-se o processo iterativo novamente, especificando novos valores com intervalo próximo ao melhor hiperparâmetro escolhido no processamento anterior. Esse processo iterativo foi sendo realizado à medida que o ajuste dos modelos apresentava melhora. Em caso negativo, voltava-se aos hiperparâmetros anteriormente encontrados.

O conjunto de 192 parcelas, fora treinado e avaliado a partir da validação cruzada *k-folds*, na qual, a amostra original é dividida aleatoriamente em *k* subamostras, definiu-se *k* = 10, mesmo processo realizado com a regressão linear.

2.3.5 Avaliação dos modelos

A avaliação da acurácia dos modelos ajustados foi realizada a partir do Viés (Equação 7), coeficiente de determinação (Equação 8), coeficiente de determinação ajustado pelo número de parâmetros do modelo (Equação 9), erro padrão da estimativa absoluto (Equação 10) e percentual (Equação 11), além da distribuição gráfica dos resíduos (Equação 12).

$$Viés = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \quad (7)$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (8)$$

$$R^2_{aj.} = 1 - \left((1 - R^2) * \left(\frac{n-1}{n-p} \right) \right) \quad (9)$$

$$SYX = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-p}} \quad (10)$$

$$SYX(\%) = \frac{SYX}{\bar{y}} * 100 \quad (11)$$

$$Resíduo(\%) = \frac{y_i - \hat{y}_i}{y_i} * 100 \quad (12)$$

Em que: y_i = variável dependente obtida a partir do inventário florestal; \hat{y}_i = variável dependente estimada pelos modelos; \bar{y} = média da variável dependente; n = número de observações; p = número de parâmetros do modelo de regressão ou o número de variáveis de entrada para os modelos de aprendizado de máquina.

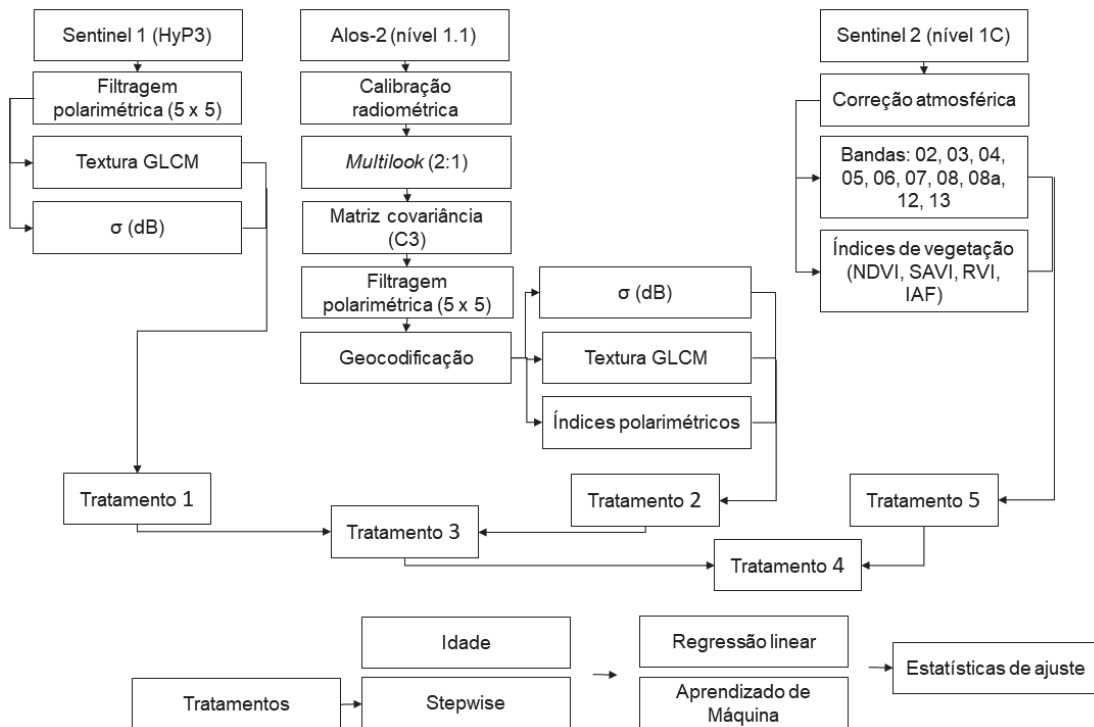
Para os resíduos dos modelos ajustados, testou-se as condicionantes da regressão aplicáveis para este tipo de estudo, sendo a normalidade e homocedasticidade. A normalidade foi avaliada pelo teste de Shapiro-Wilk e a homocedasticidade com o teste de Breusch-Pagan, ambos com nível de significância de 0,05.

Para as métricas obtidas pelo inventário florestal e estimadas pelo melhor modelo PolSar de cada cenário calculou-se o erro de amostragem e o intervalo de confiança da média.

2.3.6 Fluxograma Metodológico

Um resumo das atividades realizadas com os dados obtidos por sensoriamento remoto pode ser observado na FIGURA 10. As imagens de banda C (SENTINEL-1) passaram pelo processo de filtragem polarimétrica, geração de textura e transformação dos valores de retroespalhamento em decibéis. As cenas de banda L (ALOS/PALSAR-2) foram calibradas radiométricamente, *multilook*, transformação das polarizações para matriz covariância para que fosse possível realizar a filtragem polarimétrica e finalmente a geocodificação das imagens, sendo possível a transformação dos valores de retroespalhamento da decibéis, geração da textura e índices polarimétricos. A imagem óptica oriunda do satélite SENTINEL-2 passou pelo processo de correção atmosférica, onde foram gerados índices de vegetação a partir das bandas. Os cinco tratamentos de dados de entrada passaram pelo processo de *stepwise* para a seleção das variáveis de entrada dos modelos de regressão linear e aprendizado de máquina.

FIGURA 10 - FLUXOGRAMA METODOLÓGICO



FONTE: A autora (2020).

2.4 RESULTADOS

2.4.1 Análise dos dados

As 192 parcelas de eucalipto apresentam idade de plantio entre 1 e 10 anos, área basal (G) variando entre 8,45 e 39,80 m².ha⁻¹, o volume total com casca (V) obtido a partir do DAP e altura total variando de 41 a 616 m³.ha⁻¹. A AGB apresentou alta relação com o volume, já que o utiliza para seu cálculo, além do fator de expansão e densidade da madeira, esta última varia principalmente com a espécie e idade (TABELA 8).

TABELA 8 - ESTATÍSTICA DESCRITIVA DAS VARIÁVEIS DENDROMÉTRICAS E ÍNDICES DE VEGETAÇÃO GERADOS A PARTIR DAS IMAGENS DO SATÉLITE SENTINEL-2

Estatísticas	G (m ² .ha ⁻¹)	V (m ³ .ha ⁻¹)	AGB (Mg.ha ⁻¹)	Idade (anos)	RVI	NDVI	SAVI	IAF
Mínimo	8,45	41,00	16,61	1,40	0,02	0,22	3,03	0,50
Média	22,03	256,60	129,44	4,94	0,04	0,34	9,42	0,80
Mediana	22,90	278,50	142,08	5,40	0,04	0,33	8,92	0,80
Máximo	39,80	616,00	318,21	10,10	0,17	0,54	17,31	0,89
Desvio padrão	6,73	130,05	71,38	2,18	0,01	0,05	2,19	0,05
CV (%)	30,53	50,68	55,14	44,15	38,14	14,72	23,23	5,78
Curtose	-0,73	-0,99	-1,05	-1,40	67,65	0,75	0,91	12,94
Assimetria	-0,09	0,01	0,04	-0,24	7,68	0,75	0,67	-2,28

FONTE: A autora (2020).

LEGENDA: G = área basal (m².ha⁻¹); V = volume total (m³.ha⁻¹); AGB = biomassa aérea (Mg.ha⁻¹); RVI: índice de vegetação; NDVI: índice de vegetação normalizada; SAVI: índice de vegetação ajustado ao solo; IAF: índice de área foliar.

Observou-se que o menor coeficiente de variação para as variáveis dendrométricas é o de G (30,53%), seguida de V e AGB. Acredita-se que esse aumento de variação esteja relacionado ao aumento do número de variáveis empregadas para a obtenção dessas variáveis.

Os índices de vegetação foram criados com o intuito de ressaltar o comportamento espectral da vegetação em relação ao solo e a outros alvos da superfície terrestre, utilizando para a sua obtenção o comprimento de onda que possibilita essa diferenciação da cobertura florestal com as demais classes de uso e ocupação da terra, sendo relações entre a refletância obtidas nas bandas do vermelho e infravermelho próximo.

Para as parcelas em estudo, o índice de Razão Simples (RVI) variou entre 0,02 e 0,17, Índice de Vegetação da Diferença Normalizada (NDVI) entre 0,22 e 0,54, Índice de Vegetação Ajustado ao Solo (SAVI) entre 3,03 e 17,31, e Índice de Área Foliar (IAF) entre 0,5 e 0,89. O povoamento quando jovem, apresenta maior atividade fotossintética e crescimento,

consequentemente maior refletância e valores de índice de vegetação, à medida que o povoamento envelhece, os valores decrescem.

A idade, G e IAF possuem valores negativos de assimetria, indicando que a cauda do lado esquerdo da função de densidade de probabilidade é maior que a do lado direito. Já os valores de assimetria do RVI, NDVI e SAVI indicam comportamento contrário, cauda direita é maior, e as variáveis V e AGB apresentaram valores de assimetria praticamente nulos, indicando que os valores são distribuídos de maneira relativamente iguais em ambos os lados da média.

O valor de curtose indica que as variáveis obtidas a partir do inventário florestal assumem distribuição leptocúrtica, e os índices de vegetação obtidos a partir da imagem multiespectral assumem distribuição platicúrtica.

Os valores de retroespalhamento para a banda C (SENTINEL-1, polarizações C_{VV} e C_{VH}) apresentaram baixo desvio padrão, 1,84 e 1,63 dB, sendo que os menores valores se referem a polarização VH. Para a banda L (ALOS/PALSAR-2), o desvio padrão é o dobro do encontrado para a banda C, demonstrando que a banda L consegue capturar maior variância no retroespalhamento dos alvos. A polarização HV (L22) apresentou valores menores aos das polarizações VV (L33) e HH (L11). Dos índices polarimétricos obtidos a partir da banda L, o VSI e σ_{forest} apresentaram uma pequena amplitude de valores e, portanto, baixo desvio padrão. O índice que corresponde à estrutura do dossel (VSI) e à biomassa florestal (BMI) apresentaram coeficiente de variação maior que 20%, e o índice de degradação florestal maior que 41%, indicando uma maior diferenciação das parcelas (TABELA 9).

TABELA 9 - ESTATÍSTICA DESCRITIVA DAS POLARIZAÇÕES DO SATÉLITE SENTINEL-1, ALOS/PALSAR-2 E ÍNDICES POLARIMÉTRICOS DO ALOS/PALSAR-2

Estatísticas	C_{VV} (dB)	C_{VH} (dB)	L11 (dB)	L22 (dB)	L33 (dB)	CSI	BMI	VSI	RFDI	σ_{forest}
Mínimo	-20,86	-32,45	-22,84	-27,93	-25,67	-107,25	-25,67	1,07	0,07	1,07
Média	-15,72	-26,66	-14,25	-19,78	-16,69	-42,98	-16,69	1,21	0,19	1,19
Mediana	-15,31	-26,75	-14,00	-19,29	-16,38	-40,58	-16,38	1,20	0,18	1,18
Máximo	-12,24	-22,93	-2,62	-11,92	-5,11	-25,94	-5,11	1,74	0,65	1,65
Desvio padrão	1,84	1,63	3,70	3,49	3,62	10,93	3,62	0,09	0,08	0,08
CV%	-11,73	-6,10	-25,95	-17,65	-21,67	-25,44	-21,67	7,54	41,79	6,54
Curtose	-0,44	0,48	-0,36	-0,81	-0,03	6,36	-0,03	4,90	5,81	5,81
Assimetria	-0,54	-0,20	-0,12	-0,20	-0,16	-1,84	-0,16	1,36	1,48	1,48

FONTE: A autora (2020).

LEGENDA: C_{VV} e C_{VH} : respectivas polarizações em decibéis de banda C; L11, L22 e L33: polarizações HH, HV e VV em decibéis de banda L; CSI: estrutura do dossel; BMI: biomassa florestal; VSI: espalhamento volumétrico; RFDI: índice de degradação de Radar; σ_{forest} : coeficiente de retroespalhamento da floresta.

A assimetria dos índices polarimétricos VSI, RFDI e σ_{forest} apontaram que a cauda da distribuição do lado direito é maior, já as demais variáveis obtidas por PolSar apresentaram assimetria à esquerda.

O índice polarimétrico RFDI normalmente apresenta valores maiores que 0,3 em florestas densas, maior que 0,4 para florestas degradadas e maior que 0,6 para paisagens desmatadas. Para as plantações em estudo, uma das parcelas apresentou o valor de 0,65, mas não se referiu a desmatamento e sim a uma plantação de 1,4 ano de idade, em que o dossel ainda não fechou.

A idade apresentou alta correlação linear com G, V e AGB, acima de 0,85, com exceção da idade, a maioria das demais informações apresentaram correlação linear negativa com as variáveis dendrométricas (TABELA 10). Seguida da idade, a variável de correlação mais forte com G, V e AGB são os índices de vegetação RVI e IAF, correlação inversa, indicando que à medida que aumenta os valores das variáveis dendrométricas os índices diminuem.

TABELA 10 - CORRELAÇÃO DE SPEARMAN (ρ) ENTRE AS VARIÁVEIS DENDROMÉTRICAS E AS VARIÁVEIS SELECIONADAS PARA CADA TRATAMENTO

Tratamento	Variáveis selecionadas	G (m ² . ha ⁻¹)	V (m ³ .ha ⁻¹)	AGB (Mg.ha ⁻¹)
Tratamento 1 e 3	C _{VH}	-0,404	-0,401	-0,399
Tratamento 4	C _{VH} ASM	-0,070	-0,128	-0,139
Tratamento 1 e 4	C _{VH} contraste	0,034	0,092	0,105
Tratamento 1, 3 e 4	C _{VH} correlacao	-0,017	-0,054	-0,067
Tratamento 1, 3 e 4	C _{VH} media	-0,379	-0,378	-0,371
Tratamento 1 e 3	C _{VH} variância	0,076	0,130	0,136
Tratamento 1, 3 e 4	C _{VV}	-0,126	-0,117	-0,121
Tratamento 1 e 3	C _{VV} ASM	-0,023	-0,040	-0,055
Tratamento 4	C _{VV} contraste	0,057	0,079	0,097
Tratamento 1, 3 e 4	C _{VV} correlacao	-0,030	-0,058	-0,070
Tratamento 1, 3 e 4	C _{VV} homogeneidade	-0,057	-0,079	-0,097
Tratamento 4 e 5	IAF	-0,487	-0,537	-0,530
Tratamento 1, 2, 3, 4 e 5	Idade	0,857	0,874	0,874
Tratamento 2, 3 e 4	L11ASM	-0,147	-0,150	-0,156
Tratamento 2, 3 e 4	L11correlacao	0,079	0,041	0,036
Tratamento 2 e 3	L11dissimilaridade	0,155	0,150	0,150
Tratamento 2, 3 e 4	L12imag	0,067	0,072	0,058
Tratamento 2, 3 e 4	L12real	0,139	0,163	0,148
Tratamento 2, 3 e 4	L13imag	-0,067	-0,066	-0,061
Tratamento 2, 3 e 4	L13real	-0,173	-0,130	-0,142
Tratamento 2 e 3	L22correlacao	0,096	0,087	0,090
Tratamento 4	L22dissimilaridade	0,155	0,150	0,150
Tratamento 2 e 3	L22media	0,280	0,268	0,257
Tratamento 2, 3 e 4	L23imag	-0,015	0,009	0,014
Tratamento 2, 3 e 4	L23real	0,128	0,138	0,119
Tratamento 2, 3 e 4	L33	-0,211	-0,173	-0,194
Tratamento 2, 3 e 4	L33dissimilaridade	0,169	0,153	0,146
Tratamento 2, 3 e 4	RFDI	-0,175	-0,171	-0,155
Tratamento 4 e 5	RVI	-0,523	-0,561	-0,550
Tratamento 4 e 5	S2_B01	0,459	0,480	0,471
Tratamento 4 e 5	S2_B10	0,066	0,054	0,027
Tratamento 2, 3 e 4	σforest	-0,167	-0,122	-0,130

FONTE: A autora (2020).

LEGENDA: G: área basal (m². ha⁻¹); V: volume total (m³.ha⁻¹); AGB: biomassa aérea (Mg.ha⁻¹); C_{VV}: retroespalhamento em decibéis na polarização VV do satélite SENTINEL-1; C_{VH}: retroespalhamento em decibéis na polarização VH do satélite SENTINEL-1; L11: retroespalhamento em decibéis na polarização HH do satélite ALOS/PALSAR-2; L22: retroespalhamento em decibéis na polarização VH e HV do satélite ALOS/PALSAR-2; L33: retroespalhamento em decibéis na polarização VV do satélite ALOS/PALSAR-2; RFDI: Índice de Degradação de Radar do satélite SENTINEL-1; σforest: Coeficiente de retroespalhamento da floresta; RVI: Índice de Vegetação; IAF: Índice de área foliar; S2_B01 e S2_B10: banda 1 e 10 do satélite SENTINEL-2; ASM, dissimilaridade, correlação, homogeneidade, contraste, média, variância: texturas geradas para as polarizações de banda C e banda L.

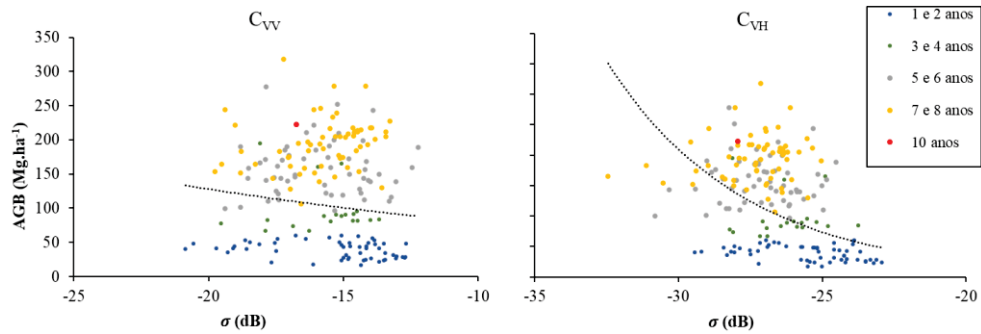
A polarização VH (recepção vertical e transmissão horizontal) da banda C possui correlação linear mais forte com as variáveis dendrométricas ($\rho = -0,399$ a $-0,404$), comparado

aos valores para a polarização VV ($\rho = -0,117$ a $-0,126$). Para a banda L, o L33 da matriz de coerência, que representa a polarização VV, apresentou a correlação mais forte para essa banda ($\rho = -0,173$ a $-0,211$).

Dos índices polarimétricos gerados com as polarizações da banda L, os índices RFDI e σ_{forest} apresentaram valor semelhante de correlação com as variáveis dendrométricas, ambos calculam a relação entre as polarizações HH e HV.

Pela dispersão gráfica do retroespalhamento tanto para as polarizações do satélite SENTINEL-1 (FIGURA 11) quanto para o satélite ALOS/PALSAR-2 (FIGURA 12), nota-se que o para o mesmo intervalo de AGB o retroespalhamento varia até 15 dB.

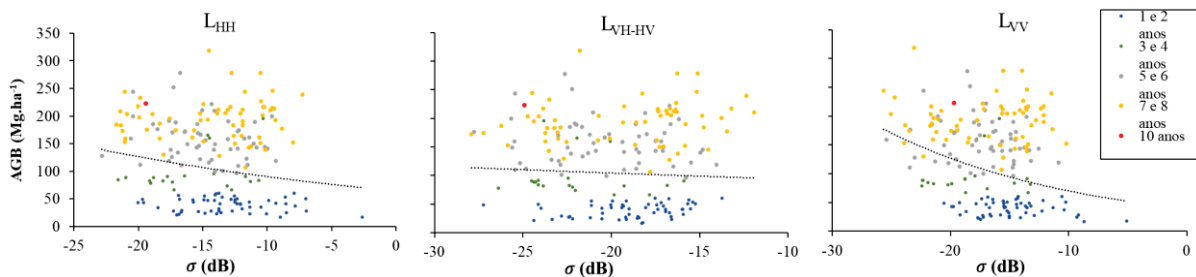
FIGURA 11 - DISPERSÃO DO RETROESPALHAMENTO DA IMAGEM DUAL-POL DO SATÉLITE SENTINEL-1 SOBRE OS VALORES DE BIOMASSA AÉREA



FONTE: A autora (2020).

LEGENDA: C_{VV} e C_{VH} polarizações da banda C.

FIGURA 12 - DISPERSÃO DO RETROESPALHAMENTO DA IMAGEM FULL-POLARIMÉTRICA DO SATÉLITE ALOS/PALSAR-2 SOBRE OS VALORES DE BIOMASSA AÉREA



FONTE: A autora (2020).

LEGENDA: L11, L22 e L33: polarizações da banda L.

2.4.2 Seleção de variáveis e ajuste dos modelos

2.4.2.1 Regressão linear múltipla

Pelas estatísticas de acurácia dos modelos de regressão linear para a estimativa de G (TABELA 11), observa-se que houve um ganho de acurácia ao utilizar PolSar com maior comprimento de onda (banda L, tratamento 2) em relação ao de menor comprimento de onda (banda C, tratamento 1).

TABELA 11 - ESTATÍSTICAS DE AJUSTE E DESEMPENHO DA REGRESSÃO LINEAR PARA A ESTIMATIVA DE ÁREA BASAL ($m^2 \cdot ha^{-1}$)

Estatísticas	Tratamentos				
	1	2	3	4	5
Viés ($m^2 \cdot ha^{-1}$)	0,209	0,183	0,171	0,166	0,202
R ²	0,730	0,775	0,792	0,791	0,746
R ² aj.	0,726	0,768	0,781	0,779	0,740
S _{YX} ($m^2 \cdot ha^{-1}$)	3,510	3,229	3,140	3,155	3,419
S _{YX} (%)	15,934	14,659	14,255	14,322	15,520

FONTE: A autora (2020).

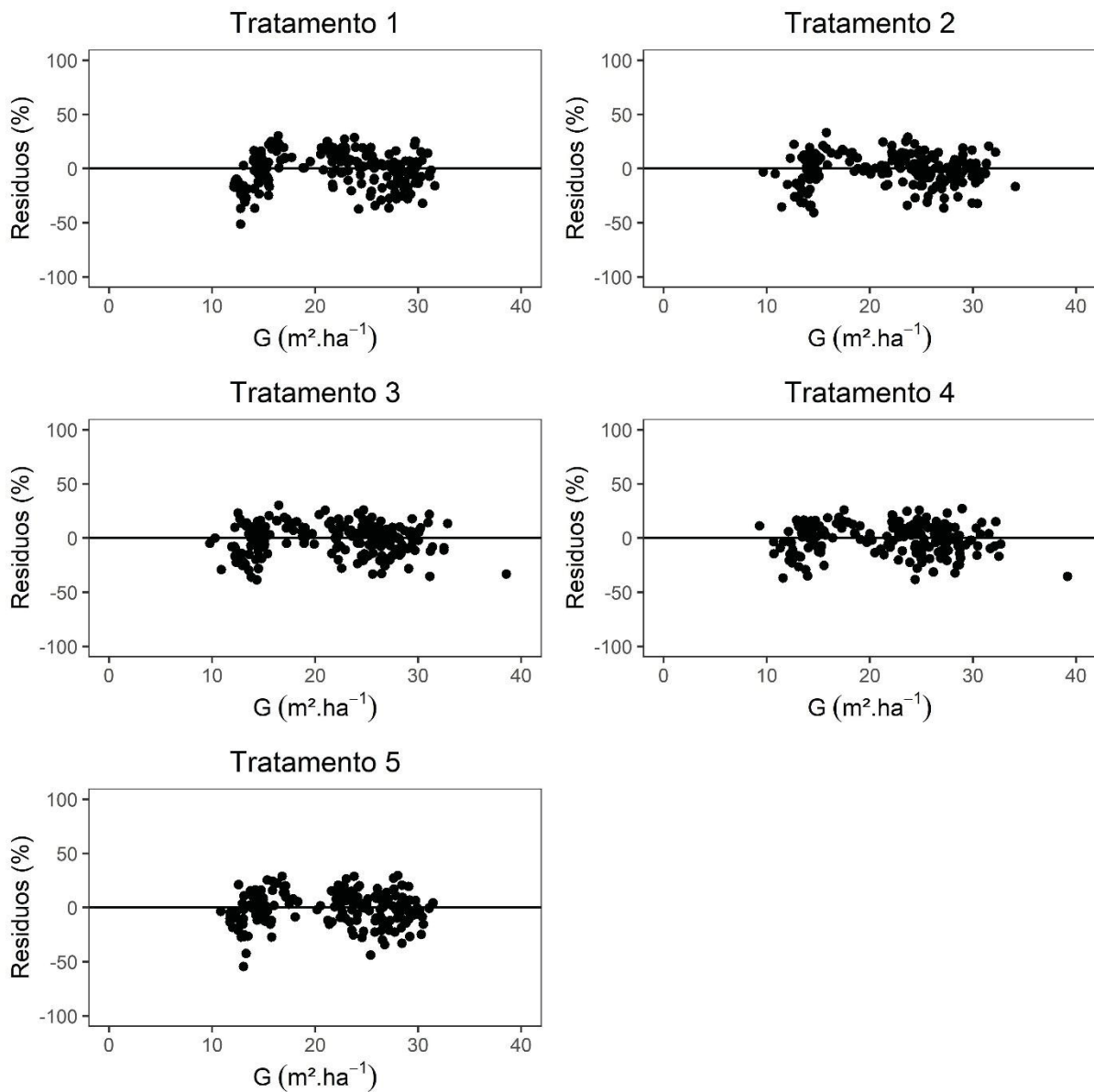
LEGENDA: Tratamento 1: idade e banda C; Tratamento 2: idade e banda L; Tratamento 3: idade, banda C e banda L; Tratamento 4: idade, tratamento PolSar de melhor desempenho, bandas e índices de vegetação oriundos do satélite SENTINEL-2; Tratamento 5: idade, bandas do satélite SENTINEL-2 e índices de vegetação.

A integração das bandas C e L (tratamento 3) proporcionou numericamente melhor desempenho em relação a sua utilização separadamente, demonstrando que a banda C tem capacidade de agregar informações, possivelmente devido a sua capacidade de penetração no dossel que é menor, agregando informações da superfície do dossel.

A adição da reflectância e índices de vegetação obtidos a partir do SENTINEL-2 (tratamento 4), proporcionou menor viés dos resíduos em relação a sua não utilização (tratamento 3), no entanto, pelas demais estatísticas utilizadas, percebeu-se que houve um declínio destas ao estimar G, como se essa informação estivesse confundindo o algoritmo de regressão linear. Porém, com a utilização isolada (tratamento 5), percebeu-se alta acurácia, superior ao uso da banda C (tratamento 1).

A dispersão gráfica dos resíduos corrobora a hipótese nula de que a população é normalmente distribuída e que a homocedasticidade das variâncias não foi rejeitada, não sendo heterocedásticos os resíduos. Percebe-se que alguns tratamentos (tratamento 1 e 5) tem maior tendência de separar as estimativas em dois grupos. O tratamento 3 e 4 apresentam maior homogeneidade na distribuição de resíduos (FIGURA 13).

FIGURA 13 - DISTRIBUIÇÃO DOS RESÍDUOS PARA A ESTIMATIVA DE ÁREA BASAL ($\text{m}^2.\text{ha}^{-1}$) POR REGRESSÃO LINEAR



FONTE: A autora (2020).

LEGENDA: Tratamento 1: idade e banda C; Tratamento 2: idade e banda L; Tratamento 3: idade, banda C e banda L; Tratamento 4: idade, tratamento PolSar de melhor desempenho, bandas e índices de vegetação oriundos do satélite SENTINEL-2; Tratamento 5: idade, bandas do satélite SENTINEL-2 e índices de vegetação.

Após realizar o processo de *stepwise* para a retirada das variáveis que inflam as equações e as com menor valor de AIC, ajustou-se os modelos de regressão linear para a estimativa das variáveis dendrométricas. Para a estimativa de G, V e AGB, foi necessária a transformação da variável dependente do modelo, assumindo valor em logaritmo natural (\ln), para que, assim, as condicionantes da regressão fossem aceitas. Para o cálculo das estatísticas

de ajuste e desempenho e geração dos gráficos de resíduos, estas variáveis foram retransformadas para a escala de valores absolutos.

O tratamento 1 para a estimativa da área basal foi a equação com o menor número de variáveis independentes, no qual pelos processos de *stepwise*, a equação resultante é composta de informações de textura das duas polarizações da banda C e idade da plantação (TABELA 12).

TABELA 12 - EQUAÇÕES AJUSTADAS PARA a ESTIMATIVA DA ÁREA BASAL (G) ($\text{m}^2 \cdot \text{ha}^{-1}$) E O P-VALOR DO TESTE DE NORMALIDADE E HOMOGENEIDADE

Trat.	Equações	Shapiro- Wilk	Breusch- Pagan
1	$\ln(G)=2,5759+(-0,0530 \cdot C_{VV} \text{correlação})+(-0,0953 \cdot C_{VH} \text{média})+(0,1302 \cdot I)$	0,506	0,277
2	$\ln(G)=2,2690+(0,0025 \cdot L13 \text{imag})+(-0,0016 \cdot L23 \text{imag})+(-0,0132 \cdot L33)+(-0,0154 \cdot L11 \text{dissimilaridade})+(0,1388 \cdot I)+(-0,1204 \cdot RFDI)$	0,119	0,167
3	$\ln(G)=2,5543+(0,0024 \cdot L13 \text{imag})+(0,0016 \cdot L23 \text{real})+(-0,0020 \cdot L23 \text{imag})+(-0,0125 \cdot L33)+(-0,1785 \cdot RFDI)+(-0,0160 \cdot L11 \text{dissimilaridade})+(-0,0521 \cdot C_{VV} \text{correlação})+(-0,1012 \cdot C_{VH} \text{média})+(0,0462 \cdot C_{VH} \text{correlação})+(0,1303 \cdot I)$	0,453	0,180
4	$\ln(G)=2,7774+(-0,0022 \cdot L13 \text{real})+(0,0018 \cdot L13 \text{imag})+(0,0020 \cdot L23 \text{real})+(-0,0023 \cdot L23 \text{imag})+(-0,0100 \cdot L33)+(-0,2154 \cdot RFDI)+(-0,0113 \cdot L33 \text{dissimilaridade})+(-0,0464 \cdot C_{VH} \text{média})+(-175,7368 \cdot B10)+(-0,1165 \cdot IAF)+(0,1192 \cdot I)$	0,474	0,874
5	$\ln(G)=3,1189+(-11,3598 \cdot B01)+(-234,8943 \cdot B10)+(-0,0293 \cdot RVI)+(0,1260 \cdot I)$	0,901	0,472

FONTE: A autora (2020).

LEGENDA: I: idade; L13imag, L23imag, L33, C23real: polarizações da matriz C3 oriundas do satélite ALOS/PALSAR-2; C_{VV} correlação, C_{VH} média, L11dissimilaridade, L33dissimilaridade: textura das respectivas polarizações; RFDI: índice polarimétrico correspondente a degradação de radar; B01, B10: bandas multiespectrais oriundas do satélite SENTINEL-2.

O tratamento 2 foi composto por informações da banda L, sendo escolhidas a variável da diagonal principal da matriz de coerência (L33) que representa o retroespalhamento na polarização VV, outras duas representações ditas imaginárias das polarizações (L13imag, L23img), a textura de dissimilaridade para a polarização HH (L11), e o índice polarimétrico referente à degradação florestal (RFDI).

O tratamento 3 é a equação com o segundo maior número de variáveis, sendo seis correspondentes à banda L e três para a banda C, além da idade. Isso indica que a banda L agrega mais informações para a estimativa de G.

Com todas as informações obtidas por sensoriamento remoto e idade, o tratamento 4 foi constituído por sete variáveis referentes à banda L, uma da banda C e duas obtidas pelo satélite SENTINEL-2. É o tratamento mais complexo em número de variáveis, porém foi o tratamento com maior número de variáveis apresentado para o processo de *stepwise*.

Das bandas e índices de vegetação do satélite SENTINEL-2, a banda 1 (aerosol), banda 10 (cirrus), e RVI foram selecionadas para compor o tratamento 5.

As estatísticas de acurácia demonstram um declínio de aproximadamente 10% ao modelar o volume comparado ao desempenho da regressão linear em modelar a área basal (TABELA 13). Para o volume é observado o mesmo comportamento dos tratamentos, melhora ao utilizar banda com comprimento de onda maior (tratamento 2), maior acurácia ao utilizar as duas bandas PolSar (tratamento 3), melhor desempenho ao utilizar informação da reflectância e com índices de vegetação (tratamento 5) ao comparar com a utilização da banda C (tratamento 1), e ganho no desempenho da maioria das estatísticas ao utilizar de todas as informações obtidas por sensoriamento remoto e idade do povoamento (tratamento 4).

TABELA 13 - ESTATÍSTICAS DE AJUSTE E DESEMPENHO DA REGRESSÃO LINEAR PARA A ESTIMATIVA DE VOLUME TOTAL ($m^3 \cdot ha^{-1}$)

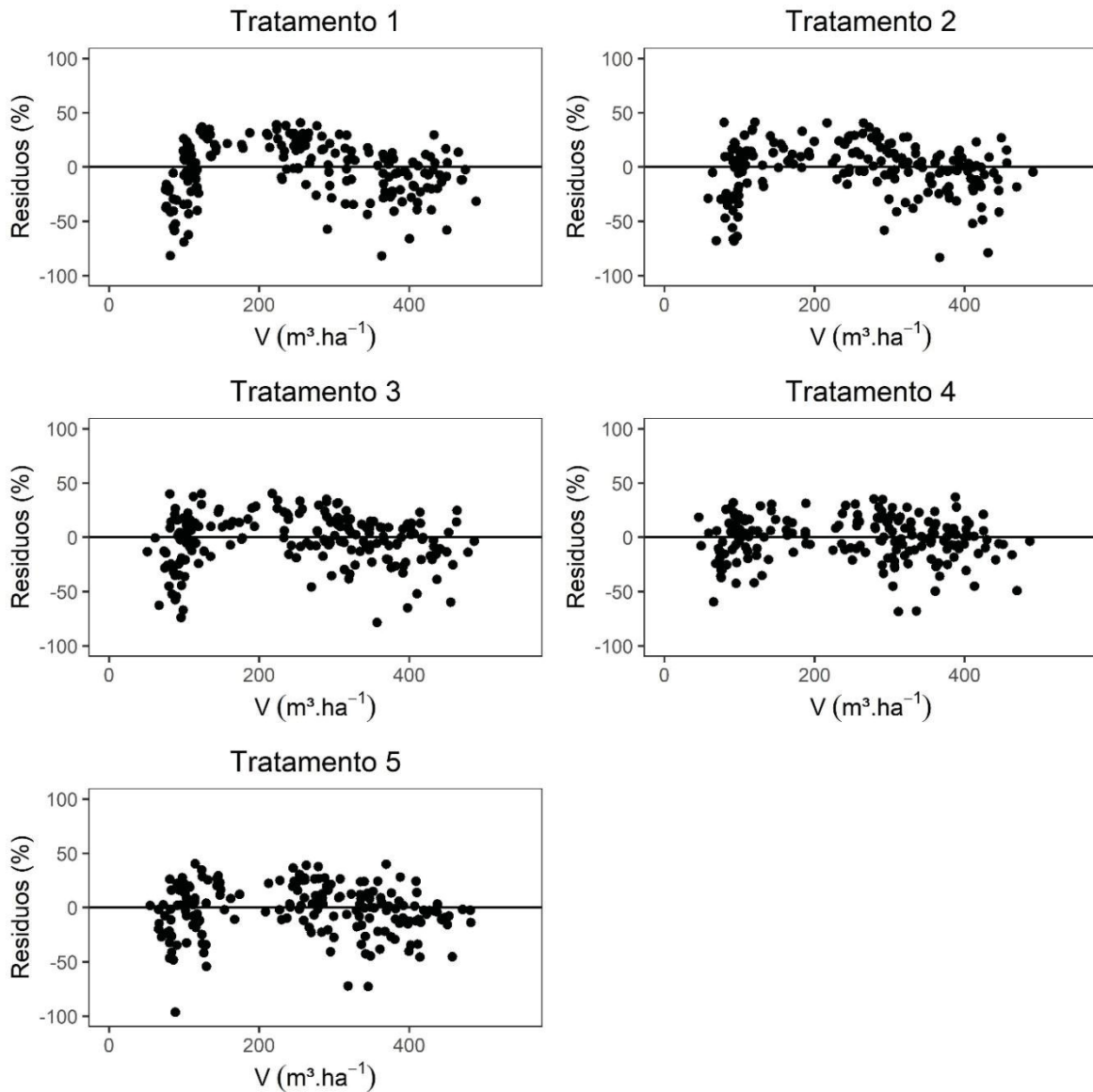
Estatísticas	Tratamentos				
	1	2	3	4	5
Viés ($m^3 \cdot ha^{-1}$)	3,435	3,259	3,159	3,341	3,629
R ²	0,701	0,751	0,771	0,801	0,747
R ² aj.	0,697	0,744	0,761	0,789	0,742
S _{YX} ($m^3 \cdot ha^{-1}$)	71,437	65,650	63,523	59,593	65,899
S _{YX} (%)	27,840	25,585	24,756	23,224	25,682

FONTE: A autora (2020).

LEGENDA: Tratamento 1: idade e banda C; Tratamento 2: idade e banda L; Tratamento 3: idade, banda C e banda L; Tratamento 4: idade, tratamento PolSar de melhor desempenho, bandas e índices de vegetação oriundos do satélite SENTINEL-2; Tratamento 5: idade, bandas do satélite SENTINEL-2 e índices de vegetação.

Corroborando com os valores das estatísticas de acurácia, os resíduos resultantes da regressão linear para a estimativa do volume são maiores e estão mais dispersos em relação a estimativa da área basal (FIGURA 14). Pela dispersão dos resíduos o tratamento 4 apresenta maior homogeneidade na distribuição dos resíduos.

FIGURA 14 - DISTRIBUIÇÃO DOS RESÍDUOS PARA A ESTIMATIVA DE VOLUME TOTAL ($\text{m}^3 \cdot \text{ha}^{-1}$) POR REGRESSÃO LINEAR



FONTE: A autora (2020).

LEGENDA: Tratamento 1: idade e banda C; Tratamento 2: idade e banda L; Tratamento 3: idade, banda C e banda L; Tratamento 4: idade, tratamento PolSar de melhor desempenho, bandas e índices de vegetação oriundos do satélite SENTINEL-2; Tratamento 5: idade, bandas do satélite SENTINEL-2 e índices de vegetação.

As variáveis selecionadas para estimar o volume (TABELA 14) do tratamento 1 e 5 foram as mesmas definidas para estimar G. Em relação aos demais tratamentos, a maioria das variáveis coincidem, demonstrando a alta importância delas para a estimativa de variáveis dendrométricas.

TABELA 14 - EQUAÇÕES AJUSTADAS PARA ESTIMATIVA DO VOLUME TOTAL (V) (m³.ha⁻¹) E O P-VALOR DO TESTE DE NORMALIDADE E HOMOGENEIDADE

Trat.	Equações	Shapiro- Wilk	Breusch- Pagan
1	$\ln(V)=4,4033+(-0,0966*C_{VV}correlação)+(-0,1628*C_{VH}média)+(0,2592*I)$	0,105	0,138
2	$\ln(V)=3,7254+(0,0035*L13imag)+(-0,0203*L33)+(0,2153*L11ASM)+(-0,0180*L33dissimilaridade)+(0,2806*I)$	0,235	0,150
3	$\ln(V)=5,3971+(0,0039*L13imag)+(-0,0217*L33)+(-0,0312*L11dissimilaridade)+(0,1098*L11ASM)+(-0,9319*L11correlação)+(-0,4165*C_{VV}homogeneidade)+(-0,1416*C_{VH}média)+(0,2709*I)$	0,453	0,131
4	$\ln(V)=5,5963+(0,0032*L13imag)+(-0,0161*L33)+(-0,0730*\sigma_{forest})+(0,2507*L11ASM)+(-0,0114*L33dissimilaridade)+(-0,3456*C_{VV}homogeneidade)+(-24,6768*B01)+(-227,4500*B10)+(-0,0561*RVI)+(-0,1122*IAF)+(0,2431*I)$	0,711	0,802
5	$\ln(V)=5,6902+(-19,5130*B01)+(-513,6054*B10)+(-0,0702*RVI)+(0,2402*I)$	0,122	0,323

FONTE: A autora (2020).

LEGENDA: I: idade; L13imag, L23imag, L33, L23real: polarizações da matriz C3 oriundas do satélite ALOS/PALSAR-2; C_{VV}correlação, C_{VH}média, L11dissimilaridade, L33dissimilaridade, L11ASM: textura das respectivas polarizações; RFDI, σ_{forest} : índices polarimétricos correspondente a degradação de radar e coeficiente de retroespalhamento da floresta; B01, B10: bandas multiespectrais oriundas do satélite SENTINEL-2.

Observa-se que as estatísticas de acurácia da regressão linear para a estimativa da biomassa aérea apresentaram erro padrão entre 25 e 32%, em média 3% de erro a mais comparado às estimativas de volume e 14% a mais que as estimativas da área basal. A acurácia no desempenho da regressão linear para a estimativa da AGB seguiu o mesmo padrão encontrado nas demais estimativas das variáveis dendrométricas, no qual o melhor resultado é descrito pelo tratamento 4 e pior para o tratamento 1 (TABELA 15).

TABELA 15 - ESTATÍSTICAS DE AJUSTE E DESEMPENHO DA REGRESSÃO LINEAR PARA A ESTIMATIVA DE BIOMASSA AÉREA (Mg.ha⁻¹)

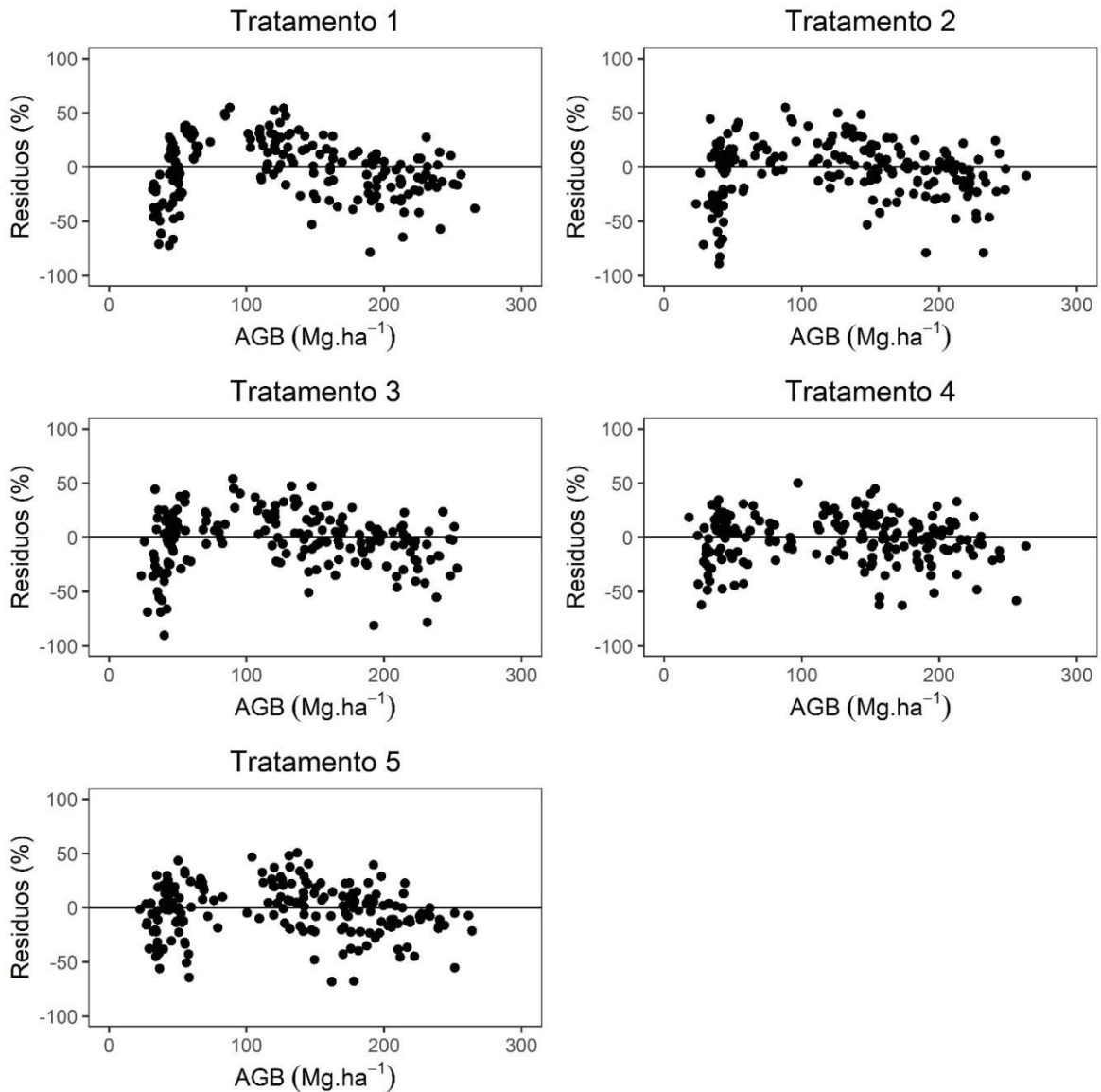
Estatísticas	Tratamentos				
	1	2	3	4	5
Viés (Mg.ha ⁻¹)	1,687	1,644	1,601	1,762	1,866
R ²	0,662	0,729	0,736	0,793	0,719
R ² aj.	0,656	0,722	0,726	0,781	0,713
S _{YX} (Mg.ha ⁻¹)	41,729	37,562	37,266	33,292	38,150
S _{YX} (%)	32,239	29,019	28,791	25,721	29,474

FONTE: A autora (2020).

LEGENDA: Tratamento 1: idade e banda C; Tratamento 2: idade e banda L; Tratamento 3: idade, banda C e banda L; Tratamento 4: idade, tratamento PolSar de melhor desempenho, bandas e índices de vegetação oriundos do satélite SENTINEL-2; Tratamento 5: idade, bandas do satélite SENTINEL-2 e índices de vegetação.

A distribuição gráfica dos resíduos corrobora com os testes de normalidade e homogeneidade em todos os tratamentos. O tratamento 4 demonstra menor erro e maior homogeneidade das estimativas. Os tratamentos 1 e 5 que são os de menor acurácia, tendendo a separar as estimativas em dois grupos (FIGURA 15).

FIGURA 15 - DISTRIBUIÇÃO DOS RESÍDUOS PARA A ESTIMATIVA DE BIOMASSA AÉREA ($\text{Mg}\cdot\text{ha}^{-1}$) POR REGRESSÃO LINEAR



FONTE: A autora (2020).

LEGENDA: Tratamento 1: idade e banda C; Tratamento 2: idade e banda L; Tratamento 3: idade, banda C e banda L; Tratamento 4: idade, tratamento PolSar de melhor desempenho, bandas e índices de vegetação oriundos do satélite SENTINEL-2; Tratamento 5: idade, bandas do satélite SENTINEL-2 e índices de vegetação.

Assim como foi para a estimativa da área basal e volume, as variáveis selecionadas para compor os modelos de estimativa da biomassa aérea por regressão linear foram as mesmas nos tratamentos 1 e 5, e para os demais tratamentos, poucas as variáveis que diferiram (TABELA 16).

TABELA 16 - EQUAÇÕES AJUSTADAS PARA ESTIMATIVA DA BIOMASSA AÉREA (AGB) (Mg.ha⁻¹) E O P-VALOR DO TESTE DE NORMALIDADE E HOMOGENEIDADE

Trat	Equações	Shapiro- Wilk	Breusch- Pagan
1	$\ln(\text{AGB})=3,5105+(-0,1161*C_{VV}\text{correlação})+(-0,1633*C_{VH}\text{média})+(0,2954*I)$	0,580	0,074
2	$\ln(\text{AGB})=+(2,7829+(0,0040*L13\text{imag})+(-0,0253*L33)+(0,2117*L11\text{ASM})+(-0,0240*L33\text{dissimilaridade})+(0,3171*I)$	0,342	0,073
3	$\ln(\text{AGB})=2,1391+(0,0041*L13\text{imag})+(-0,0254*L33)+(0,1842*L11\text{ASM})+(-0,0261*L33\text{dissimilaridade})+(-0,0294*C_{VH})+(-0,0984*C_{VV}\text{correlação})+(0,3063*I)$	0,564	0,125
4	$\ln(\text{AGB})=4,8189+(0,0038*L13\text{imag})+(-0,0260*L33)+(0,2637*L11\text{ASM})+(-0,0170*L33\text{dissimilaridade})+(-0,4976*C_{VV}\text{homogeneidade})+(-24,9729*B01)+(-317,727*B10)+(-0,0593*RVI)+(-0,1293*IAF)+(0,2773*I)$	0,797	0,831
5	$\ln(\text{AGB})=4,9961+(-17,5371*B01)+(-684,0858*B10)+(-0,0764*RVI)+(0,2731*I)$	0,468	0,327

FONTE: A autora (2020).

LEGENDA: Trat.: tratamentos; I: idade; L13imag, L23imag, L33, L23real: polarizações da matriz C3 oriundas do satélite ALOS/PALSAR-2; C_{VV}correlação, C_{VH}média, L11dissimilaridade, L33dissimilaridade, L11ASM: textura das respectivas polarizações; RFDI: índice polarimétrico correspondente a degradação de radar; B01, B10: bandas multiespectrais oriundas do satélite SENTINEL-2.

2.4.2.2 Aprendizado de máquina

Aplicou-se modelos de aprendizado de máquina ao tratamento 4, pois foi o tratamento de melhor desempenho para a estimativa por regressão linear do volume total e biomassa aérea, além de ser o tratamento que integra todas as informações oriundas dos diferentes comprimentos de onda e a idade do povoamento de eucalipto.

O desempenho pela validação cruzada dos modelos de aprendizado de máquina ao estimar a área basal apresentou coeficiente de determinação (R²) entre 0,806 e 0,963 e erro padrão da estimativa variando entre 6,269 e 14,29%, sendo o algoritmo *Model Tree* (M5) o responsável pelo pior desempenho e o algoritmo *Random Forest* (RF) pelo melhor (TABELA 17).

TABELA 17 - ESTATÍSTICAS DE AJUSTE E DESEMPENHO DOS MODELOS DE APRENDIZADO DE MÁQUINA PARA A ESTIMATIVA DE ÁREA BASAL ($\text{m}^2.\text{ha}^{-1}$)

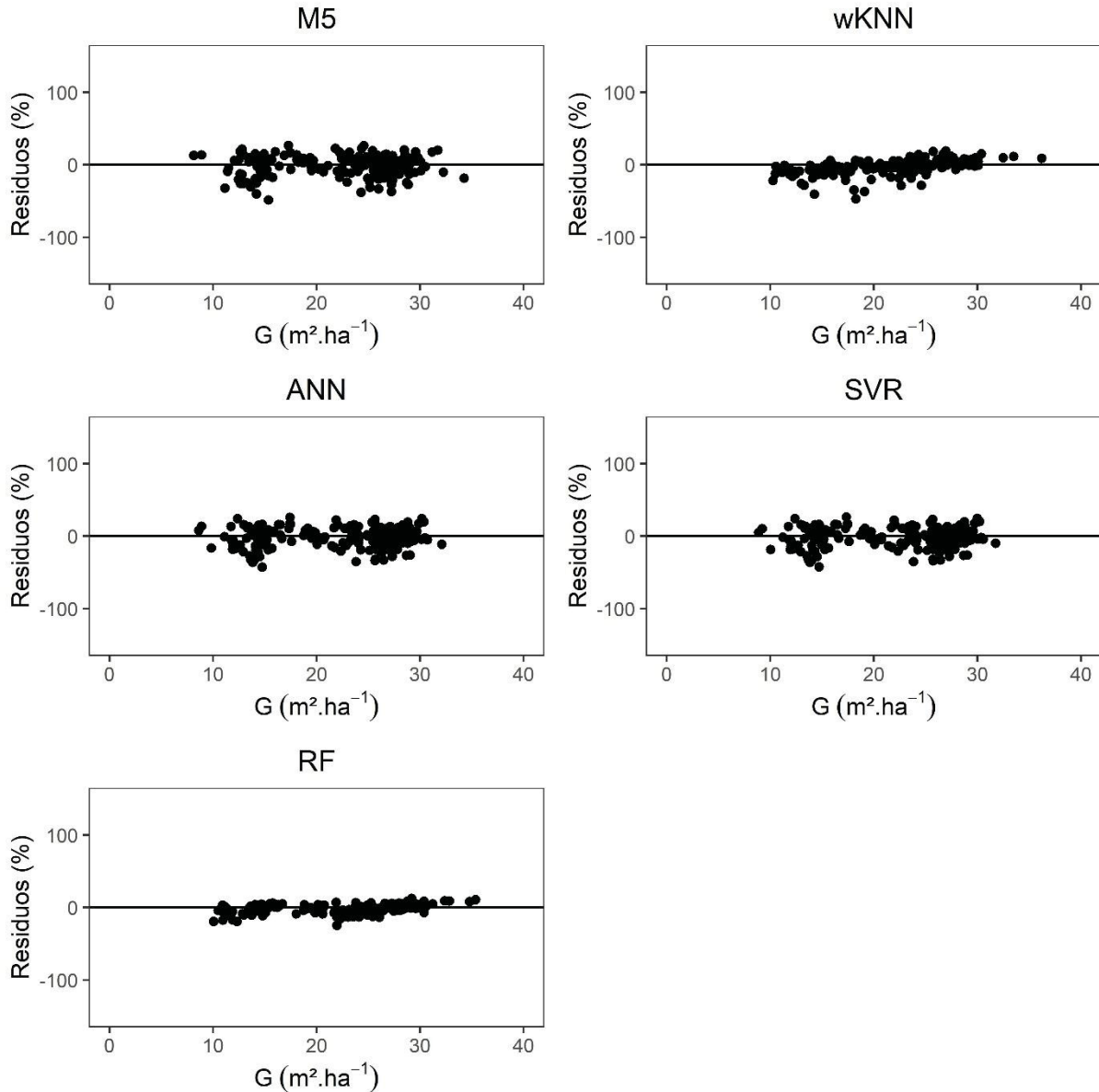
Hiperparâmetros		Viés ($\text{m}^2.\text{ha}^{-1}$)	R ²	R ² aj.	S _{YX} ($\text{m}^2.\text{ha}^{-1}$)	S _{YX} (%)
M5	Pruned= Yes; Smoothed= No; Rules= No	0,000	0,806	0,780	3,148	14,290
ANN	Size= 1; Decay= 0,02	0,000	0,815	0,789	3,077	13,970
SVR	Sigma= 0,001; C= 31	0,000	0,814	0,788	3,084	14,000
wKNN	Kmax= 17; Distance= 2; Kernel= rank	-0,120	0,919	0,908	2,035	9,238
RF	Mtry= 20; Ntree= 200	-0,082	0,963	0,958	1,381	6,269

FONTE: A autora (2020).

LEGENDA: M5: *Model Tree*; RF: *Random Forest*; wKNN: *Weighted K-Nearest Neighbor*; ANN: *Artificial Neural Network*; SVR: *Support Vector Regression*.

Corroborando com as estatísticas de desempenho, pela distribuição dos resíduos percebe-se que o algoritmo RF apresentou os menores resíduos e com distribuição uniforme ao longo do eixo x, o qual representa os valores de área basal estimada. Os algoritmos M5, ANN e SVR estimaram valores de área basal próximos, constado pela distribuição gráfica dos resíduos e pelas estatísticas de desempenho (FIGURA 16).

FIGURA 16 - DISTRIBUIÇÃO DOS RESÍDUOS PARA A ESTIMATIVA DE ÁREA BASAL ($\text{m}^2 \cdot \text{ha}^{-1}$) POR APRENDIZADO DE MÁQUINA



FONTE: A autora (2020).

LEGENDA: M5: *Model Tree*; RF: *Random Forest*; wKNN: *Weighted K-Nearest Neighbor*; ANN: *Artificial Neural Network*; SVR: *Support Vector Regression*.

Assim como o encontrado pela regressão linear, os erros ao estimar o volume total e biomassa aérea foram maiores (TABELA 18, TABELA 19). Ao estimar o volume total o desempenho dos modelos de aprendizado de máquina apresentou R^2 de 0,715 a 0,968 e S_{YX} entre 9,676 e 28,762%, com o pior desempenho atribuído ao modelo wKNN e o melhor ao RF (TABELA 18).

TABELA 18 - ESTATÍSTICAS DE AJUSTE E DESEMPENHO DOS MODELOS DE APRENDIZADO DE MÁQUINA PARA A ESTIMATIVA DO VOLUME TOTAL ($\text{m}^3 \cdot \text{ha}^{-1}$)

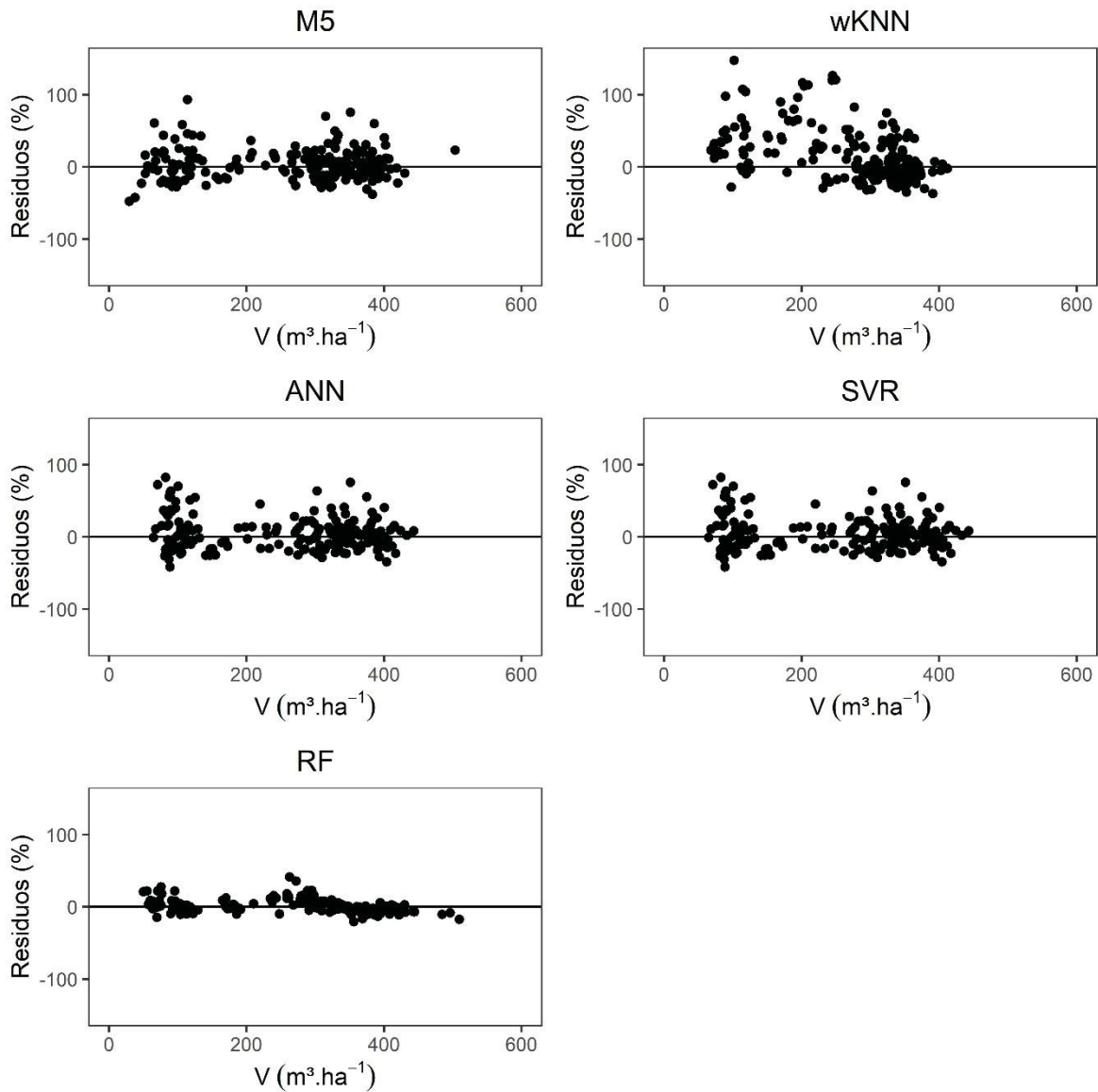
Hiperparâmetros		Viés ($\text{m}^3 \cdot \text{ha}^{-1}$)	R ²	R ² aj.	S _{YX} ($\text{m}^3 \cdot \text{ha}^{-1}$)	S _{YX} (%)
M5	Pruned= Yes; Smoothed= No; Rules= No	0,000	0,844	0,822	54,671	21,306
ANN	Size= 1; Decay= 0,1	0,011	0,854	0,834	52,780	20,569
SVR	Sigma= 0,001; C= 30	0,011	0,854	0,834	52,780	20,569
wKNN	Kmax= 17; Distance= 2; Kernel= rank	9,943	0,715	0,676	73,803	28,762
RF	Mtry= 20; Ntree= 43	-1,015	0,968	0,963	24,828	9,676

FONTE: A autora (2020).

LEGENDA: M5: *Model Tree*; RF: *Random Forest*; wKNN: *Weighted K-Nearest Neighbor*; ANN: *Artificial Neural Network*; SVR: *Support Vector Regression*.

Corroborando com as estatísticas de ajuste e desempenho dos modelos ao estimar o volume total, observa-se que o algoritmo wKNN apresentou a pior distribuição gráfica dos resíduos, com tendência a superestimar as estimativas. Em média, todos os algoritmos apresentaram tendência a superestimar os resíduos (FIGURA 17).

FIGURA 17 - DISTRIBUIÇÃO DOS RESÍDUOS PARA A ESTIMATIVA DE VOLUME TOTAL ($\text{m}^3 \cdot \text{ha}^{-1}$) POR APRENDIZADO DE MÁQUINA



FONTE: A autora (2020).

LEGENDA: M5: *Model Tree*; RF: *Random Forest*; wKNN: *Weighted K-Nearest Neighbor*; ANN: *Artificial Neural Network*; SVR: *Support Vector Regression*.

Os modelos de aprendizado de máquina seguiram o mesmo comportamento ao estimar a biomassa aérea, tendo o melhor desempenho de ajuste para o modelo RF ($R^2 = 0,968$ e S_{YX} de 10,445%) e o pior para o modelo wKNN ($R^2 = 0,774$ e S_{YX} de 27,843%) (TABELA 19).

TABELA 19 - ESTATÍSTICAS DE AJUSTE E DESEMPENHO DOS MODELOS DE APRENDIZADO DE MÁQUINA PARA A ESTIMATIVA DA BIOMASSA AÉREA (Mg.ha⁻¹)

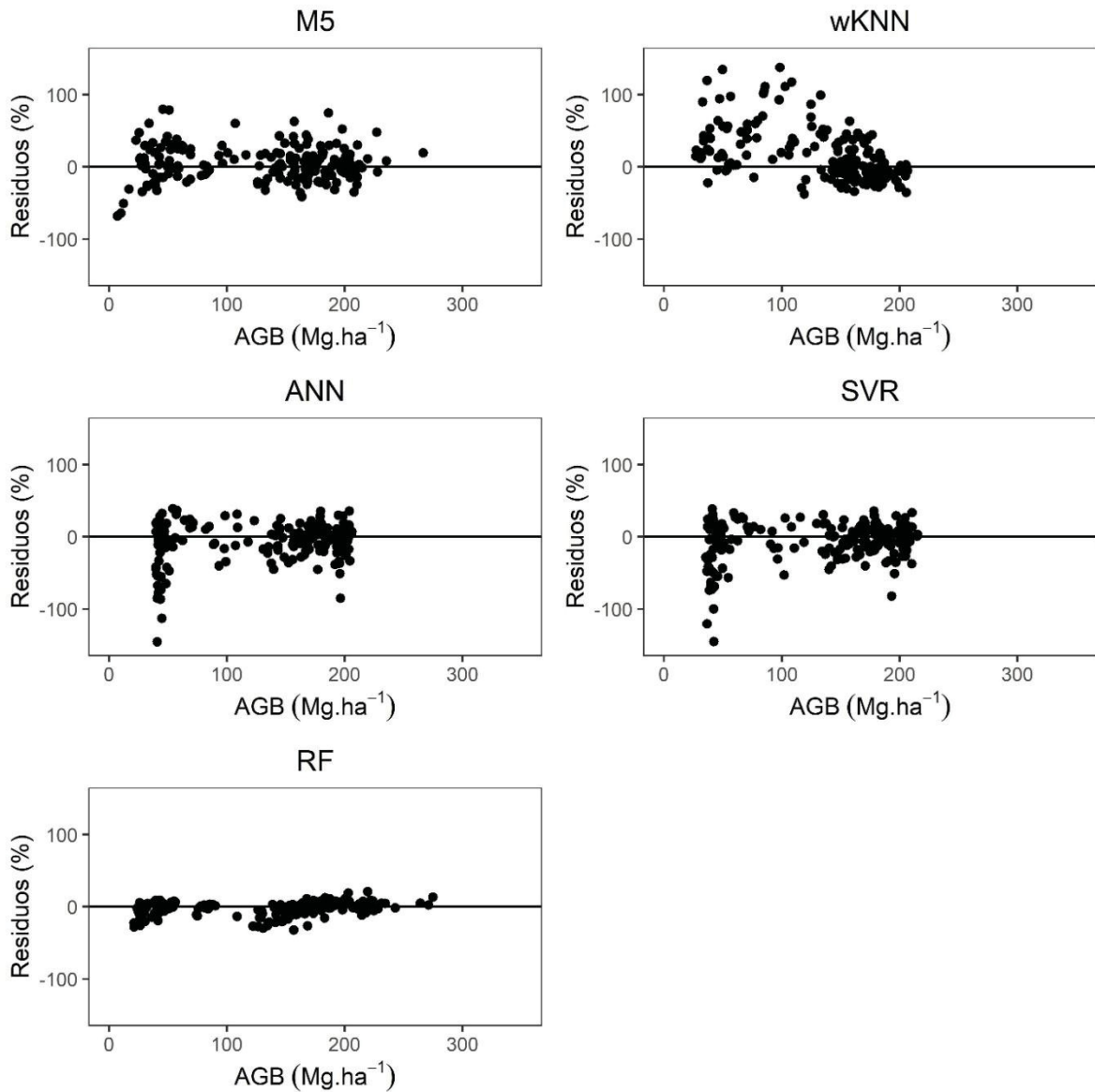
Hiperparâmetros		Viés (Mg.ha ⁻¹)	R ²	R ² aj.	S _{YX} (Mg.ha ⁻¹)	S _{YX} (%)
M5	Pruned= Yes; Smoothed= No; Rules= No	0,000	0,840	0,818	30,341	23,441
ANN	Size= 1; Decay= 0,1	0,179	0,849	0,828	29,478	22,774
SVR	Sigma= 0,001; C= 30	0,061	0,853	0,833	29,076	22,463
wKNN	Kmax= 17; Distance= 2; Kernel= rank	3,121	0,774	0,744	36,040	27,843
RF	Mtry= 20; Ntree= 43	-0,541	0,968	0,964	13,520	10,445

FONTE: A autora (2020).

LEGENDA: M5: *Model Tree*; RF: *Random Forest*; wKNN: *Weighted K-Nearest Neighbor*; ANN: *Artificial Neural Network*; SVR: *Support Vector Regression*.

Observa-se pela distribuição gráfica que o modelo wKNN apresenta tendência de superestimar a biomassa aérea e o algoritmo RF apresenta os menores resíduos, corroborando com as estatísticas de desempenho. Os modelos ANN e SVR apresentaram comportamento semelhante, superestimando principalmente os menores valores de biomassa aérea. As estimativas dos modelos foram em média próximo aos valores de referência (FIGURA 18).

FIGURA 18 - DISTRIBUIÇÃO DOS RESÍDUOS PARA A ESTIMATIVA DE BIOMASSA AÉREA ($\text{Mg}\cdot\text{ha}^{-1}$) POR APRENDIZADO DE MÁQUINA



FONTE: A autora (2020).

LEGENDA: M5: *Model Tree*; RF: *Random Forest*; wKNN: *Weighted K-Nearest Neighbor*; ANN: *Artificial Neural Network*; SVR: *Support Vector Regression*.

Analisando o tempo de processamento entre os modelos de aprendizado de máquina, constatou-se que o modelo wKNN fora o de desempenho mais lento, seguido do SVR, ANN, RF e M5. Estes podem ser atribuídos ao número de hiperparâmetros testados, por exemplo, o modelo M5 foi testado seu desempenho apenas avaliando a poda ou não.

2.4.3 Regressão linear x aprendizado de máquina

Para modelar as variáveis dendrométricas a partir da regressão linear, viu-se a necessidade da transformação das variáveis dependentes para logaritmo neperiano, para que assim os modelos atendessem aos pressupostos da regressão. Por não ser necessário atender esse critério, foi possível modelar a variável dendrométricas sem transformação a partir dos modelos de aprendizado de máquina. O problema da não linearidade desaparece com o uso de aprendizado de máquina, como esperado.

Ao comparar o desempenho dos modelos de aprendizado de máquina com a regressão linear múltipla, observou-se que: i) todos os modelos de aprendizado de máquina com exceção do M5 apresentaram desempenho superior ao da regressão linear ao estimar a área basal (G); ii) todos os modelos de aprendizado de máquina, com exceção do wKNN, apresentaram desempenho superior a regressão linear ao modelar o volume total (V) e a biomassa aérea (AGB).

O RF se sobressaiu sobre os demais modelos de aprendizado de máquina na estimativa da área basal, volume e biomassa aérea. Ao comparar o desempenho da regressão linear frente ao RF, ambos utilizando as mesmas variáveis do tratamento 4 que abrange o conjunto total dos dados selecionados, é notável o superior desempenho do modelo de aprendizado de máquina (TABELA 20).

Assim como observado para a regressão linear, o modelo RF obteve menores erros ao estimar a área basal (G), seguido do volume total (V) e biomassa aérea (AGB). Porém, observa-se pelo erro padrão da estimativa que o maior ganho de desempenho do modelo RF em relação a regressão linear se dá no sentido contrário, com a maior diferença encontrada ao estimar a AGB, seguido de V e G.

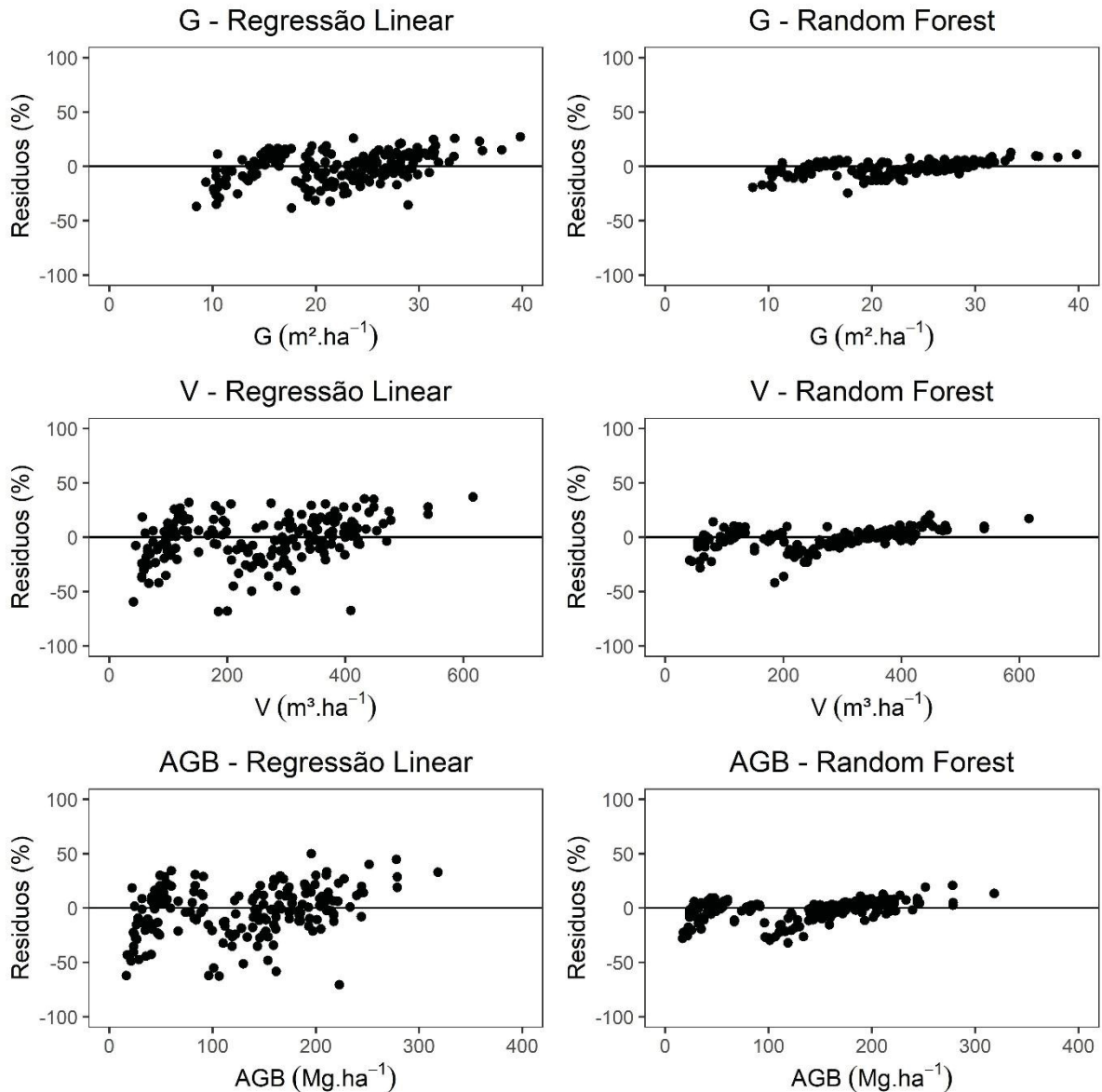
TABELA 20 - ESTATÍSTICAS DE AJUSTE E DESEMPENHO DA REGRESSÃO LINEAR E DO ALGORITMO RANDOM FOREST (RF) PARA A ESTIMATIVA DA ÁREA BASAL (G), VOLUME TOTAL (V) E BIOMASSA AÉREA (AGB)

	G (m ² .ha ⁻¹)		V (m ³ .ha ⁻¹)		AGB (Mg.ha ⁻¹)	
	Regressão Linear	RF	Regressão Linear	RF	Regressão Linear	RF
Viés	0,166	-0,082	3,341	-1,015	1,762	-0,541
R ²	0,791	0,963	0,801	0,968	0,793	0,968
R ² aj.	0,779	0,958	0,789	0,963	0,781	0,964
S _{YX}	3,155	1,381	59,593	24,828	33,292	13,520
S _{YX} (%)	14,322	6,269	23,224	9,676	25,721	10,445

FONTE: A autora (2020).

Corroborando com as estatísticas de ajuste e desempenho, a distribuição gráfica dos resíduos demonstra menores erros ao utilizar do modelo RF. Em média, as estimativas oriundas tanto da regressão linear quanto do modelo RF tiveram tendência a superestimar (FIGURA 19).

FIGURA 19 - DISTRIBUIÇÃO DOS RESÍDUOS PARA A ESTIMATIVA REGRESSÃO LINEAR E DO ALGORITMO DE MELHOR DESEMPENHO (RF) PARA A ESTIMATIVA DA ÁREA BASAL (G), VOLUME TOTAL (V) E BIOMASSA AÉREA (AGB)



FONTE: A autora (2020).

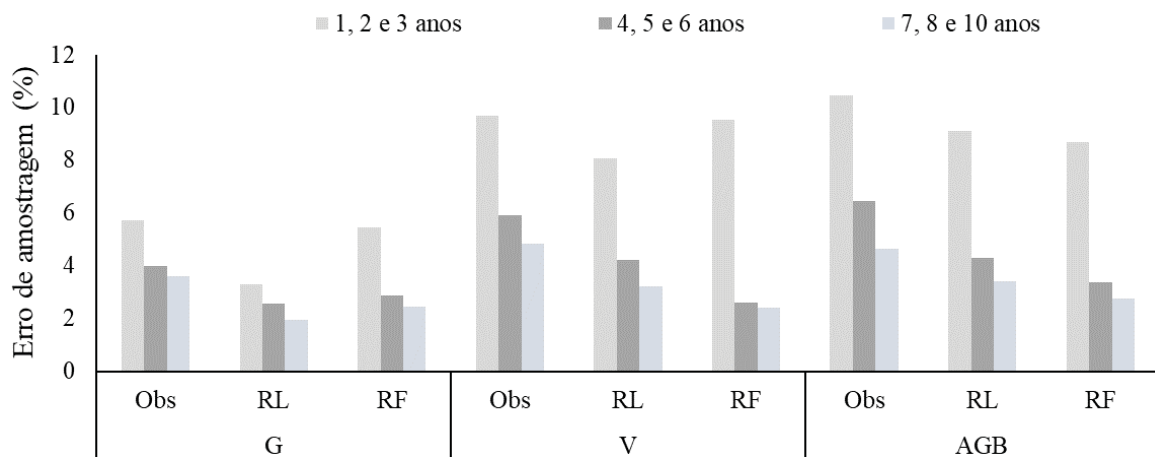
LEGENDA: G: área basal (m².ha⁻¹); V: volume total (m³.ha⁻¹); AGB: biomassa aérea (Mg.ha⁻¹).

2.4.4 Erros oriundos do inventário e da modelagem

Os erros amostrais resultam por não se medir toda a população florestal, ou seja, se empregam de técnicas de amostragem para a estimativa de variáveis de interesse (SANQUETTA et al., 2014), estas devem ser consideradas ao realizar o inventário florestal. Já esta é uma das principais vantagens da utilização de dados obtidos por sensoriamento remoto, pois cobre 100% da área, em essência eliminando o erro de amostragem, porém, em decorrência disso essa técnica insere outras fontes de erro as quais serão tratadas adiante.

Comparando-se o erro de amostragem obtido por inventário e o resultante do melhor modelo PolSar, o erro resultante da modelagem foi menor em todas as classes de idade. Para a variável AGB obtida a partir de métricas do inventário o erro foi maior que 10%, para as demais métricas o erro ficou dentro do limite de erro de amostragem aceitável (FIGURA 20).

FIGURA 20 - ERRO DE AMOSTRAGEM ENTRE OS VALORES ORIUNDOS DO INVENTÁRIO E DAS ESTIMATIVAS COM MODELAGEM POR REGRESSÃO LINEAR E RANDOM FOREST

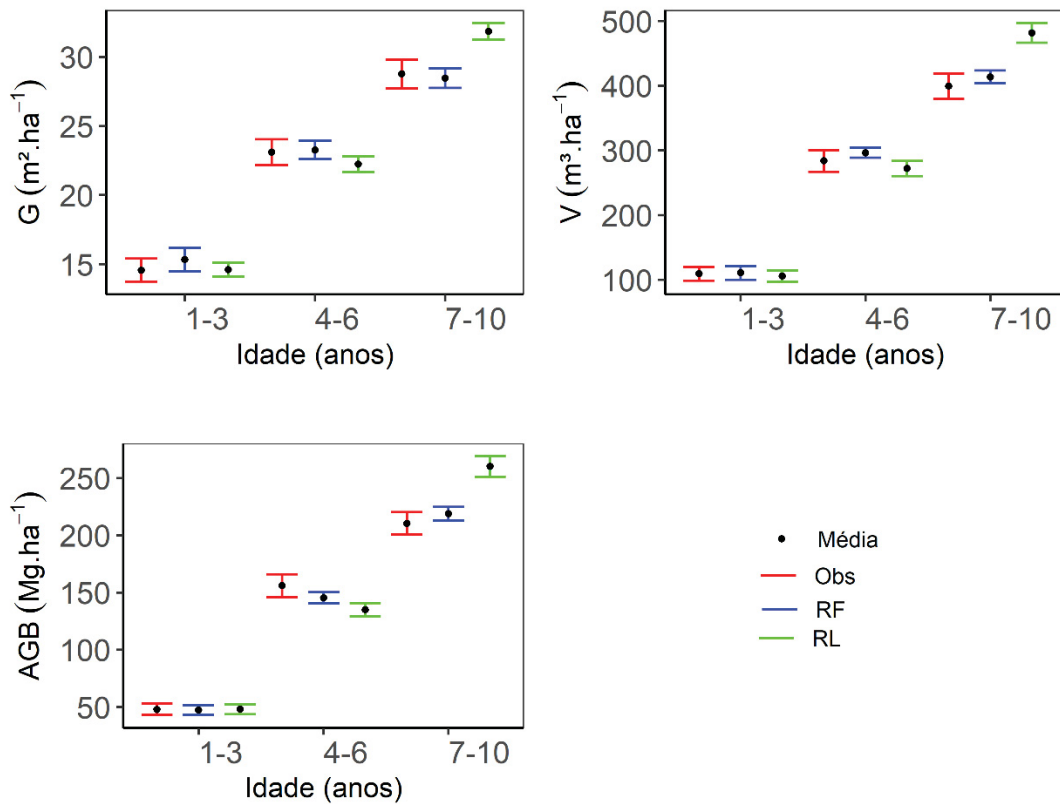


FONTE: A autora (2020).

LEGENDA: Obs: observações oriundas do inventário; RL: valores estimados pela regressão linear; RF: valores estimados pelo modelo *Random Forest*; G: área basal ($\text{m}^2 \cdot \text{ha}^{-1}$); V: volume total ($\text{m}^3 \cdot \text{ha}^{-1}$); AGB: biomassa aérea ($\text{Mg} \cdot \text{ha}^{-1}$).

Quando separado os dados por classe etária as médias das variáveis dendrométricas obtidas pelo inventário e com a modelagem por regressão linear e *Random Forest* ficaram dentro do limite de intervalo de confiança. As variáveis dendrométricas tanto obtidas a partir das métricas do inventário florestal quanto as estimadas por modelos de regressão linear e *Random Forest* utilizando PolSar apresentaram valores semelhantes e estão dentro do intervalo de confiança do inventário florestal (FIGURA 21).

FIGURA 21 - COMPARAÇÃO ENTRE AS MÉDIAS E INTERVALO DE CONFIANÇA POR CLASSE DE IDADE PARA AS MÉTRICAS ORIUNDAS DO INVENTÁRIO (Obs) E ORIUNDAS DAS ESTIMATIVAS POLSAR



FONTE: A autora (2020).

LEGENDA: Obs: observações oriundas do inventário; RL: valores estimados pela regressão linear; RF: valores estimados pelo modelo *Random Forest*; G: área basal ($m^2 \cdot ha^{-1}$); V: volume total ($m^3 \cdot ha^{-1}$); AGB: biomassa aérea ($Mg \cdot ha^{-1}$).

2.5 DISCUSSÃO

Nesse estudo, uma abordagem de estimativa de variáveis dendrométricas foi desenvolvida para povoamentos de eucalipto plantados no Sul do Brasil, combinando informações do espectro eletromagnético na região do microondas (satélites ALOS/PALSAR-2 e SENTINEL-1), visível e infravermelho (satélite SENTINEL-2). As variáveis dendrométricas das 192 parcelas possibilitaram ajustar modelos baseados em regressão linear múltipla, utilizando de diferentes combinações de dados de entrada (tratamentos), com o retroespalhamento, textura das polarizações, índices polarimétricos, índices de vegetação e bandas do visível e infravermelho. Para o tratamento de dados de entrada de melhor desempenho aplicou-se de modelos de aprendizado de máquina a fim de compará-los. Através dos modelos ajustados, foi possível estimar variáveis dendrométricas de povoamentos de eucalipto, com resultados promissores.

As variáveis dendrométricas e os índices de vegetação (RVI e IAF) apresentam relação forte e negativa, demonstrando comportamento inversamente proporcional. A alta correlação deve-se ao fato desses índices de vegetação utilizarem das bandas espectrais na região do espectro eletromagnético abrangendo o visível e infravermelho, considerada como sendo uma região de forte absorção e máxima reflectância da vegetação.

Observa-se que a correlação da biomassa florestal com os índices de vegetação mencionados anteriormente, tem relação direta com a idade e conseqüentemente estado das plantações florestais. A correlação para esses povoamentos é maior até dois anos de idade, porém com relação inversa, e enfraquece de acordo com o seu crescimento. Acredita-se que a correlação negativa esteja relacionada ao tamanho da copa das árvores e fechamento do dossel florestal, visto que de acordo com Spanner et al. (1990) as correlações positivas são encontradas para áreas florestais com dosséis fechados e as correlações negativas são encontradas para áreas com dosséis abertos.

O comportamento do gráfico de dispersão entre a AGB e os valores de retroespalhamento corrobora com os estudos anteriores, os quais evidenciaram que o retroespalhamento aumenta com o aumento da biomassa da floresta. No entanto, gradualmente perde sua sensibilidade para altos valores de AGB, na qual ocorre a saturação, resultando em uma relação logarítmica ou sigmoideal entre biomassa e retroespalhamento (DOBSON et al., 1992; LE TOAN et al., 1992; SAATCHI et al., 2011). Para este estudo, não ficou evidente a ocorrência de saturação, apenas a L22 que se refere a polarização HV da banda L que não apresentou tendência logarítmica ou sigmoideal como esperado.

Para a banda L, a polarização VV (C33) apresenta maior correlação positiva com as variáveis dendrométricas comparado às demais polarizações, HH, HV/VH. Já Kumar et al. (2019) ao modelarem a AGB do Himalaia no Nordeste da Índia com a banda L, obtiveram maior correlação com a polarização HV ($R^2 = 0,63$) em comparação a HH ($R^2 = 0,58$). Os autores atribuíram este resultado à maior sensibilidade do retroespalhamento HV da banda L as diversas estruturas florestais. Atribui-se esse fato a estrutura da floresta, onde a plantação de eucalipto apresenta maior representatividade na altura, ou seja, na estrutura vertical tendo maior correlação com a polarização nessa orientação. Já a floresta estudada por Kumar et al. (2019), por se tratar de floresta nativa, apresenta a copa maior, por isso a polarização que tem o sinal emitido na polarização vertical e o sinal recebido na horizontal apresenta maior correlação com a biomassa.

Para a banda C, a polarização cruzada HV apresenta maior correlação quando comparado à polarização VV. Ao realizar a modelagem utilizando-se das variáveis obtidas da

banda C, obteve-se R^2 variando entre 0,66 e 0,73. Resultados menos significativos foram encontrados por Liu et al. (2019) ao modelar a altura média da floresta e AGB no Nordeste da China com a mesma banda, em que com a polarização VH encontraram R^2 de 0,53 e para a polarização VV o R^2 de 0,42.

O processo de *stepwise* não selecionou as variáveis obtidas por sensoriamento remoto de maior correlação linear com a área basal, volume e biomassa aérea, tais como o SAVI e NDVI. Isso pode ser justificado pela forma com que o *stepwise* seleciona as variáveis, buscando por modelos parcimoniosos, evitando a multicolinearidade e informações redundantes.

Neste estudo optou-se por incluir a variável idade em todos os tratamentos devido a sua correlação com o estoque de madeira existente na floresta e pela sua fácil obtenção, por exemplo, se o interessado em conhecer as métricas da floresta não possuir os dados de inventário, o mesmo consegue saber facilmente a idade do talhão identificando a data do plantio, ao buscar por séries históricas de imagens de satélite.

Outros estudos incluem a idade na modelagem, como por exemplo Ismail et al. (2015) que ao modelarem o volume de madeira em povoamentos de eucaliptos localizados em Zululand na África do Sul. Além da idade, os autores utilizaram-se de informações obtidas a partir do ALOS/PALSAR e SPOT-4, verificaram que ao incluírem a idade como variável independente, obtiveram melhoria de 15 a 27%. Joshi et al. (2017) e Souza et al. (2019) sugerem que a utilização de informações como a idade e tratamentos silviculturais podem superar a perda de sinal por saturação de dados PolSar e multiespectrais.

Do total de 15 modelos de regressão linear ajustados, 12 deles eram compostos por métricas obtidas por PolSar, o processo de *stepwise* selecionou ao todo 27 métricas PolSar, sendo 16 representações de textura das polarizações, dois índices polarimétricos, sete representações da matriz de coerência da banda L, e duas polarizações da banda C. Esse resultado evidencia a relevância que as informações de textura agregam aos modelos, corroborando com o encontrado também em outros estudos, tais como, Chen et al. (2018), Chen et al. (2019), Cutler et al. (2012) e Ozdemir e Karnieli (2011). Dentre as variações da textura GLCM disponíveis, observa-se que a homogeneidade e dissimilaridade são variáveis importantes e selecionadas como variáveis explicativas (GIANNETTI et al., 2018; HENTZ, 2018), comportamento observado em muitos dos modelos ajustados no presente estudo.

Ao modelar as variáveis dendrométricas, a acurácia da regressão linear foi maior ao utilizar a banda L, comparada à utilização da banda C. Esse fato já era esperado, já que a banda C por apresentar menor comprimento de onda, tem menor penetração no dossel e satura com estoque menor de biomassa. Pesquisas anteriores apontaram que as bandas X e C são mais

sensíveis à umidade da superfície que recebe a energia do radar do que a biomassa (PULLIAINEN; KURVONEN; HALLIKAINEN, 1999; WAY et al., 1990), porém para as cenas adquiridas exclui-se a umidade, já que não havia precipitação antecedente que pudesse estar influenciando neste desempenho.

A banda L, por ter maior comprimento de onda neste estudo, apresenta retroespalhamento mais forte nos galhos e troncos, coletando a maior parte da energia do dossel das árvores (KUMAR et al., 2019). No entanto, observa-se que a utilização da banda C, quando aliada à banda L, apresenta modelos mais acurados do que a utilização da banda L de forma isolada, demonstrando que a banda C, apesar de suas limitações, agrega informações importantes.

De forma geral, as informações multiespectrais e índices de vegetação melhoraram a acurácia das estimativas por regressão linear quando adicionadas às demais informações PolSar (tratamento 4), porém a métrica de viés indica que os valores estimados ficaram piores.

Apesar de outros estudos já terem documentado a deficiência de dados ópticos para o mapeamento da AGB, devido a insensibilidade à biomassa a partir do fechamento do dossel e interferência de gramíneas em savanas e bosques (NAIDOO et al., 2016; ZEIDLER; WEGMANN; DECH, 2012), as informações multiespectrais e seus índices de vegetação apresentaram melhores estimativas do que quando comparado a utilização do PolSar de forma isolada. Corroborando com este resultado, Forkuor et al. (2020), ao mapearem a AGB na savana da África Ocidental, encontraram melhor desempenho do estimador ao utilizarem as informações do SENTINEL-2 (RMSE = 60,6 Mg.ha⁻¹), comparado a utilização da banda C (RMSE = 78,6 Mg.ha⁻¹), e a utilização conjunta das informações (RMSE = 45,4 Mg.ha⁻¹), produziu melhores resultados do que seu uso isolado. Outros autores encontraram resultados contrários, tais como os estudos de Chen et al. (2018) em uma área de 2500 km² no nordeste da China, e Navarro et al. (2019) em áreas ocupadas por mangue, no Senegal.

O tratamento 4 que é composto da combinação de todas as informações oriundas do PolSar e dados ópticos apresentou o melhor resultado nas estimativas do volume e biomassa aérea. Já para modelar a área basal, o tratamento 3 foi mais eficiente, este constituído das bandas C e L. Esse resultado pode ser atribuído às informações que cada imagem agrega na modelagem, como por exemplo, as cenas obtidas do PolSar e suas derivações, fornecendo informações referentes a estrutura das plantações florestais e os dados ópticos com os índices de vegetação fornecendo informações sobre a densidade do dossel e caracterização da folhagem.

Resultados superiores ao do presente estudo foram encontrados por Gama, Santos e Mura (2010), ao ajustarem modelos de regressão linear entre as representações das bandas X e P, para a estimativa de volume e biomassa de *Eucalyptus saligna*, localizados no estado de São Paulo - Brasil. Os modelos ajustados foram constituídos pela altura interferométrica e o índice polarimétrico CSI, resultando em R^2 superior a 0,81 e erros menores que 10%. Esses resultados superiores podem ser justificados pela qualidade das informações que os autores detinham, tais como, banda P que apresenta o maior comprimento de onda na região do microondas, o qual possibilita total penetração do sinal de radar no povoamento florestal e a saturação do sinal que ocorre em maiores taxas de biomassa. Também por possuírem dados de altura interferométrica, a qual representa a altura total dos povoamentos e essa variável apresentar altíssima correlação com volume e biomassa.

No presente estudo optou-se inicialmente pela utilização da regressão linear múltipla por ser um método amplamente conhecido, que gera uma equação a partir do modelo ajustado, a qual pode ser aplicada para outras parcelas de *Eucalyptus* de mesma espécie. Além do motivo dos resíduos terem atendido aos pressupostos da regressão e os resultados serem satisfatórios para um modelo de tal simplicidade.

Após a avaliação dos tratamentos de dados de entrada a partir da regressão linear foi possível definir que o tratamento 4, composto de todas as informações oriundas dos diferentes comprimentos de onda fora o de melhor desempenho para a estimativa da maioria das variáveis dendrométricas. Então, utilizando-se das mesmas variáveis selecionadas para a regressão linear pode-se aplicar a modelagem por aprendizado de máquina, a qual apresentou desempenho superior. Dentre os modelos de aprendizado de máquina o modelo *Random Forest* se destacou dos demais na estimativa de G, V e AGB. O desempenho superior pelos modelos de aprendizado de máquina pode ser justificado por se tratar de técnicas robustas de identificação e previsão de padrões, as quais tem sido utilizada com sucesso nas mais diversas áreas do conhecimento, assim como para a estimativa de parâmetros biofísicos da floresta a partir de dados de sensoriamento remoto (VERRELST et al., 2012; DOMINGUES et al., 2020; SOUZA et al., 2019).

Dentre as variáveis dendrométricas, observou-se que os modelos de regressão linear e o *Random Forest* apresentaram maior precisão ao estimar área basal ($S_{YX} = 14,322\%$ e $S_{YX} = 6,269\%$), seguido pelos modelos de volume total ($S_{YX} = 23,224\%$ e $S_{YX} = 9,676\%$) e biomassa aérea ($S_{YX} = 25,721\%$ e $S_{YX} = 10,445\%$). Acredita-se que este erro esteja associado, primeiro a altura para a obtenção do volume e depois a densidade da madeira. Para a área basal é considerado apenas a área transversal dos povoamentos, e para a obtenção do volume é utilizada

além das métricas de área transversal, a altura. Já para a obtenção da biomassa aérea, é levado em consideração um fator de expansão fixo para todo o povoamento, o volume que já resultou em maior erro e a densidade específica da madeira, sendo a última variável altamente influenciada pela idade do povoamento e da espécie. Essa hipótese se apoia também na alta correlação que as variáveis dendrométricas possuem com a idade dos povoamentos, variável está inclusa em todos os modelos ajustados.

O ganho ao utilizar o modelo RF foi de 8,053% no erro padrão da estimativa para área basal, 13,548% ao estimar o volume e 15,276% nas estimativas de biomassa aérea. O maior ganho na precisão das estimativas com o modelo RF deu-se justamente na estimativa da AGB e volume as quais apresentaram os maiores erros na regressão linear. Atribui-se esse melhor desempenho não só à capacidade da modelagem do RF, mas também à notável piora da regressão linear em ajustar os coeficientes adequados para a estimativa dessas variáveis dendrométricas.

O desempenho da modelagem com regressão linear e técnicas de aprendizado de máquina para a estimativa de variáveis biofísicas da floresta já foi avaliada por outros autores, tais como, Domingues et al. (2020) que modelaram a AGB de *Eucalyptus*. Os autores também destacaram que a combinação de dados PolSar (ALOS/PALSAR) e ópticos (AVNIR-2), melhoraram as estimativas de AGB, e quando utilizados isoladamente, os dados ópticos apresentaram melhor resultado ($R^2 = 0,86$ para dados ópticos e $R^2 = 0,52$ para a banda L). Para a banda L, as polarizações de maior relevância foram na seguinte ordem: VH, VV e HH. Ao utilizarem de regressão linear múltipla, os autores encontraram resultados menos significativos que o presente estudo ($R^2 = 0,71$), e com o uso de uma técnica de inteligência artificial, ANN, também alcançaram desempenho superior que dá regressão linear, R^2 de 0,95 e raiz do erro médio quadrático de 2,87%.

Outra pesquisa que corrobora com os resultados deste estudo é o de Souza et al. (2019), os quais estimaram o volume de povoamentos de *Eucalyptus*. Como variáveis de entrada os autores utilizaram a banda L (ALOS/PALSAR) e imagem óptica (AVNIR-2), avaliaram o modelo SVR. De forma similar, concluíram que a integração dos dados obtidos por sensoriamento remoto proporcionou melhores resultados que seu uso isolado (Banda L (ALOS/PALSAR) e imagem óptica (AVNIR-2)), seus resultados com o uso de SVR alcançaram R^2 de 0,926 e erro médio quadrático de $1,007 \text{ m}^3 \cdot \text{ha}^{-1}$.

O erro de amostragem por classe etária das variáveis dendrométricas obtidas a partir do inventário e do estimado pela regressão linear e por aprendizado de máquina, ficou menor que o erro estabelecido por Péllico Neto e Brena (1997) como de máximo aceitável ($\pm 10\%$).

Os excelentes resultados da modelagem para a estimativa das variáveis dendrométricas podem ser atribuídos ao conjunto de informações e metodologia utilizada, sendo a homogeneidade das plantações de eucalipto, qualidade das imagens de satélite e a abordagem de aprendizado de máquina.

Neste estudo não foram mensuradas e retropropagadas todas as fontes de erros, foi considerado apenas o erro na modelagem, mas deve-se ponderar que existem outros erros, tais como: mensuração das variáveis dendrométricas, amostragem do inventário, geolocalização das parcelas em campo, erro das equações alométricas para a estimativa das variáveis altura, volume, densidade básica da madeira e biomassa aérea.

2.6 CONCLUSÃO

A estimativa de variáveis dendrométricas de área basal, volume e biomassa aérea a partir de dados obtidos por sensoriamento remoto ativo e passivo mostrou-se possível, porém com resultados influenciados por variáveis que representam as características dos povoamentos florestais, variáveis estas selecionadas para compor os tratamentos de dados de entrada.

No total, 15 experimentos envolvendo diferentes combinações de dados foram realizados para alcançar os objetivos do estudo. Para a estimativa da área basal, a utilização de variáveis da banda C, L e idade (tratamento 3), apresentaram o melhor desempenho. Para a estimativa do volume e biomassa aérea, a integração de todas as informações disponíveis juntamente com a idade, possibilitou as estimativas mais acuradas.

As estimativas a partir da banda C foram satisfatórias. A banda de maior comprimento de onda se mostrou mais apropriada para a estimativa das variáveis dendrométricas (banda L). Os índices polarimétricos, textura GLCM e índices de vegetação se mostraram variáveis relevantes para a modelagem. A modelagem utilizando apenas as informações da imagem multiespectral apresentaram desempenho superior ao da utilização da banda C.

Dentre os modelos de aprendizado de máquina avaliados, o modelo *Random Forest* apresentou desempenho superior, este também superior ao desempenho da regressão linear, com melhoria de aproximadamente 8% no erro padrão da estimativa para a área basal, 13% para o volume e 15% para a estimativa da biomassa aérea. Os erros foram menores que aproximadamente 10%.

Os modelos de aprendizado de máquina foram aplicados apenas no tratamento que engloba todas as variáveis para a entrada da modelagem, acredita-se que se aplicado as demais

combinações de dados de entrada (tratamento) as estimativas tendem a ser mais precisas também, então recomenda-se sua aplicação em todas as situações.

As estatísticas de desempenho dos modelos ajustados indicam um grande potencial de informações obtidas por PolSar e sensor multiespectral, como suporte ao inventário florestal de grandes áreas, sendo dependente apenas de informações obtidas por sensoriamento remoto. Com isso, pode-se concluir que com as informações e metodologias utilizadas é possível produzir estimativas precisas, auxiliando dessa forma, na redução dos custos de inventário florestal.

CAPÍTULO 3: ESTIMATIVA DE BIOMASSA EM FLORESTA TROPICAL POR MEIO DE APRENDIZAGEM DE MÁQUINA E DADOS MULTISENSOR OBTIDO POR SENSORIAMENTO REMOTO DE LIVRE ACESSO

Resumo: A quantificação da capacidade que a floresta amazônica possui em estocar carbono se faz necessária para os programas de mudanças climáticas. O objetivo do presente estudo é avaliar o potencial de Radar Polarimétrico de Abertura Sintética (PolSar) de banda C e banda L, imagens multiespectrais e de algoritmos de aprendizado de máquina para a estimativa da biomassa florestal acima do solo (AGB). Este estudo foi realizado na Floresta Nacional do Jamari, localizada na Amazônia brasileira. Utilizou-se de dados oriundos do satélite SENTINEL-1, SENTINEL-2, ALOS/PALSAR-2 e dados LiDAR. Realizou-se a seleção das variáveis mais relevantes para a composição de cada tratamento de variáveis de entrada para a modelagem, sendo: 1) polarizações oriundas de banda C e banda L; 2) tratamento 1 com adicional de textura PolSar e índices polarimétricos; 3) tratamento 2 com as bandas multiespectrais e índices de vegetação derivados. Como variável resposta da modelagem utilizou-se a AGB estimada a partir de levantamento laser aerotransportado. A modelagem foi realizada com cinco algoritmos de aprendizado de máquina: *Model Tree* (M5), *Random Forest* (RF), *Weighted K-Nearest Neighbor* (wKNN), *Artificial Neural Network* (ANN) e *Support Vector Regression* (SVR). A adição de informações de textura proporcionou pequena melhora em relação ao tratamento 1. As informações multiespectrais (tratamento 3) contribuíram com uma melhora de 7% no RMSE e os algoritmos conseguiram representar melhor a magnitude da AGB. Para os três tratamentos o algoritmo que apresentou a melhor representação da AGB florestal foi o modelo M5. As estimativas geradas a partir de informações disponibilizadas gratuitamente demonstraram grande potencial.

Palavras-chaves: Floresta Amazônica; modelagem; aprendizado de máquina, SENTINEL, ALOS/PALSAR-2.

3.1 INTRODUÇÃO

As florestas são ecossistemas de vegetação amplamente distribuídas no planeta, cobrindo aproximadamente 4000 milhões de hectares (FAO, 2015). Informações sobre o estado atual e monitoramento da floresta tem importância mundialmente reconhecida devido ao papel desempenhado pelas florestas no fornecimento de serviços ambientais, principalmente na atuação como o principal sumidouro terrestre de carbono (WARING; RUNNING, 2007), e portanto no clima global.

No entanto, esse papel das florestas permanece pouco caracterizado quantitativamente, em função de dificuldades práticas de medir os estoques de biomassa florestal em larga escala (MINH et al., 2016), principalmente se tratando das florestas tropicais que são ambientes complexos.

Comumente a biomassa florestal acima do solo (AGB) pode ser estimada por meio de três métodos: i) medições de inventário: envolve medições de parcelas amostrais pelo método

destrutivo, as quais produzem informações confiáveis sobre a AGB em escalas locais ou regionais (CHAVE et al., 2004); ii) simulações baseadas em modelos: geralmente fornecem estimativas em escala local à global, com base em modelos alométricos pré-ajustados com dados de inventário (LU, 2006); iii) estimativa a partir de dados oriundos de satélites: são geralmente combinações de medições remotas e baseadas nas medições obtidas pelo inventário de campo (STICKLER et al., 2009).

Em comparação com o modelo e as abordagens de inventário florestal, as técnicas de sensoriamento remoto melhoram significativamente a eficiência do mapeamento da AGB da floresta em áreas de difícil acesso (LU, 2006), com melhor cobertura espacial e temporal.

As medições em campo permanecem essenciais para a calibração e validação dos indicadores obtidos pelo sensoriamento remoto, no entanto, existem várias fontes potenciais de erro no inventário, como a variação no tamanho das parcelas e as equações alométricas utilizadas (CHAVE et al., 2004). Além disso, de acordo com Houghton; Hall e Goetz (2009), as atuais estimativas de AGB e estoque de carbono em florestas tropicais são oriundos de extrapolações de um número limitado de locais de campo, deixando muitas regiões e tipos de floresta sub-representadas. Como aliada às medições em campo, algumas tecnologias a exemplo do LiDAR, são promissoras e capazes de estimar variáveis da floresta para uma grande escala e com alta precisão (RUIZ et al., 2014; DEO; RUSSELL; DOMKE, 2016; SILVA et al.; 2017).

Porém, a tecnologia *Light Detection And Ranging* (LiDAR) não tem capacidade de penetrar sobre as nuvens, quando aerotransportada apresenta a limitação espacial e conseqüentemente do alto custo, tornando inviável a aquisição de dados sistemáticos para o monitoramento, sendo necessária a busca por outra tecnologia que possa realizar o monitoramento e quantificação dos produtos da floresta. Dessa forma, surge a alternativa de dados Radar Polarimétrico de Abertura Sintética (PolSar) para as investigações florestais, sendo mais evidenciado após o lançamento de novos satélites que têm sensores radar a bordo (por exemplo, TerraSAR-X, ALOS/PALSAR-2 e SENTINEL-1) e obtêm imagens com boa resolução espacial e revisita frequente (SANTI et al., 2017). Trata-se também de tecnologia ativa assim como a LiDAR.

Os radares operam na frequência do microondas, sendo sensores ativos, o que possibilita o imageamento mesmo a noite e com cobertura de nuvens, característica favorável para o monitoramento da Amazônia. Uma vantagem adicional do radar é que ele pode operar em diferentes combinações de polarização, tornando-o adequado para monitorar o dossel e parâmetros estruturais da vegetação (NINGTHOUJAM; JOSHI; ROY, 2018).

A abordagem do retroespalhamento baseado em dados PolSar é amplamente conhecida para o mapeamento da AGB (BERNINGER et al., 2018). Consiste na mensuração da energia recebida pelo sensor após a transmissão, conhecido como retroespalhamento ou *backscatter*, em que posteriormente essa medida é relacionada com as medições de AGB em campo. O retroespalhamento é sensível à quantidade de AGB, aumentando o retroespalhamento de acordo com o aumento de AGB, até um certo ponto no qual ocorre a saturação, ou seja, a sensibilidade do PolSar à AGB fica estagnada. Este ponto está relacionado ao comprimento de onda do sensor (JOSHI et al., 2017), como por exemplo, a banda C apresenta comprimento de onda capaz de penetrar através das folhas, porém, espalhada com a existência de pequenos ramos. Já a banda L, apresenta comprimento de onda maior e é espalhado por troncos e galhos grossos, enquanto a banda P é a mais adequada para a quantificação da AGB, pois nesse comprimento de onda a energia do sensor é capaz de penetrar o dossel da floresta. Porém, dados da banda P ainda não estão disponíveis de forma gratuita (GHASEMI et al., 2011).

Sensores de banda L têm sido amplamente empregados para estimar a AGB da floresta (CARREIRAS; MELO; VASCONCELOS, 2013; CARTUS; SANTORO; KELLNDORFER, 2012; SAATCHI et al., 2011b; SANTOS et al., 2002), oriundos principalmente do satélite ALOS/PALSAR. Existe uma lacuna de disponibilidade gratuita para esse comprimento de onda, a não ser pelo mosaico global que é confeccionado a partir de várias cenas adquiridas ao longo do ano, processado e disponibilizado pela Agência de exploração Aeroespacial do Japão (JAXA).

Considerando a limitação da série de dados PolSar gratuitos e a limitação do comprimento de onda indicada para a alta AGB, característica da Floresta Amazônica, há necessidade de explorar a combinação de outros dados disponíveis gratuitamente para garantir a continuidade das estimativas de cobertura florestal. Este estudo tem como contribuição explorar o potencial da integração das bandas C e L, a integração de dados de textura e de informações multiespectrais juntamente com índices de vegetação, para a estimativa de biomassa aérea de floresta tropical. A textura de imagens PolSar pode fornecer informações relacionadas às propriedades estruturais e geométricas dos dosséis florestais (DEGRANDI; LUCAS; KROPACEK, 2009), que podem apresentar correlação com a AGB e serem especialmente úteis na discriminação entre tipos de florestas (PODEST; SAATCHI, 2002).

3.2 OBJETIVOS

O objetivo do presente capítulo é analisar a potencialidade do uso de imagens de livre acesso em estimar a AGB de floresta tropical. De maneira específica, os objetivos são:

1) Avaliar a precisão das estimativas de AGB a partir de modelos construídos com diferentes fontes de dados obtidos por sensoriamento remoto.

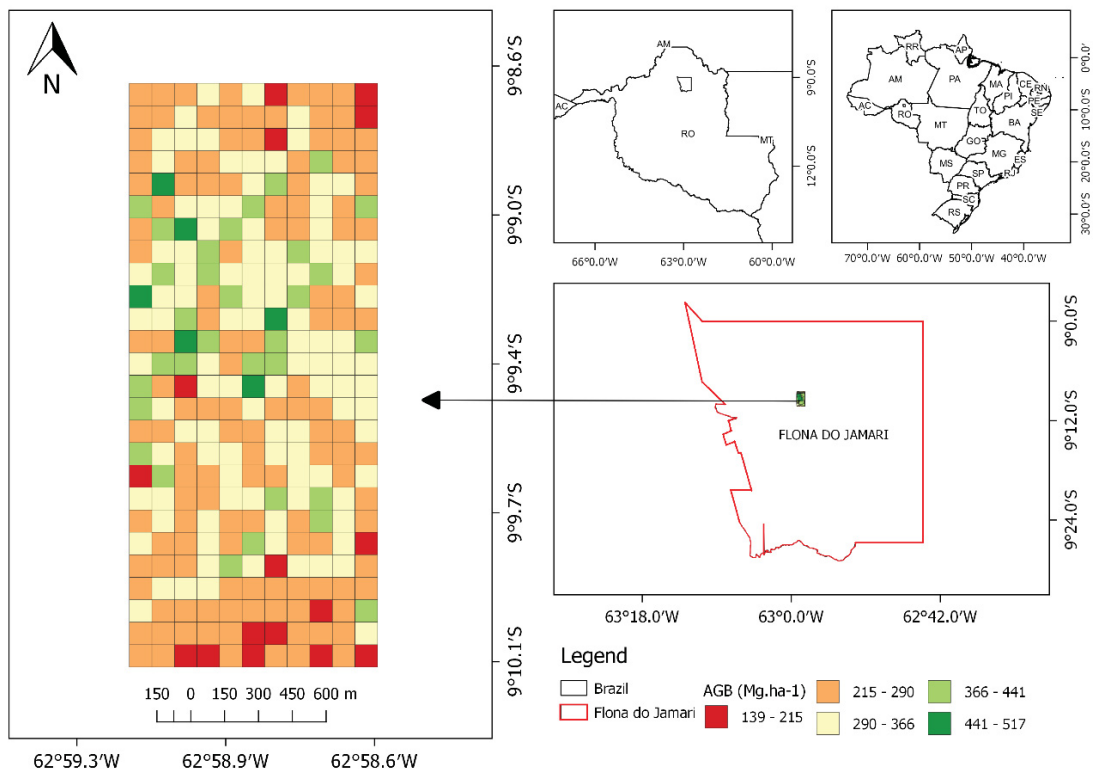
2) Comparar o desempenho de modelos de aprendizado de máquina na estimativa de AGB.

3.3 MATERIAL E MÉTODOS

3.3.1 Área de estudo

A área de estudo localiza-se na Floresta Nacional (Flona) do Jamari (FIGURA 22), Estado de Rondônia, na região norte do Brasil. A Flona do Jamari possui uma área de cerca de 220.000 ha de Floresta Ombrófila Densa, sendo uma Unidade de Conservação de Uso Sustentável da Floresta Amazônica.

FIGURA 22 - AGB ($\text{Mg}\cdot\text{ha}^{-1}$) OBTIDA POR MEIO DE DADOS LASER AEROTRANSPORTADO E LOCALIZAÇÃO DA ÁREA DE ESTUDO SOBRE O BRASIL



FONTE: A autora (2020).

Esta Unidade resguarda amostras da flora características da Amazônia sul-ocidental, região submetida a altas taxas de desmatamento. De acordo com o Manual Técnico da Vegetação Brasileira (IBGE, 2012), o relevo é suavemente ondulado, a classificação fisionômica-ecológica predominantemente é de Floresta Ombrófila Densa, com algumas porções de Floresta Ombrófila Aberta, com predominância de palmeiras ou com cipós.

3.3.2 Aquisição e processamento de dados SENTINEL

Foram utilizadas imagens do satélite SENTINEL-1 (PolSar) e SENTINEL-2 (multiespectral) da Agência Espacial Europeia. A imagem multiespectral foi adquirida com o centro de dados científicos da Agência *Copernicus Sentinels* (ESA, 2016), e um resumo das especificações pode ser observado na TABELA 21.

TABELA 21 - RESUMO DAS ESPECIFICAÇÕES DOS SATÉLITES SENTINEL-1 E SENTINEL-2

SENTINEL-1			SENTINEL-2	
Data de aquisição	01-07 e 09-08-2016	13-07 e 21-08-2016	Data de aquisição	2015-08-25
Produto	GRDM	GRDM	Número do satélite	A
Modo do instrumento	EW (Faixa extralarga)	EW (Faixa extralarga)	Instrumento	MSI
Órbita	156/32	156/32	Órbita	908
Ponto	621/1144	621/1144	Ponto	53
Polarização	HH	VV, VH	NSSDC	2015-000A
			identificador	
Direção	Descendente	Descendente	Direção	Descendente
			Nível do produto	1C

FONTE: A autora (2020).

Sabendo que a umidade tem influência sobre os valores de retroespalhamento (KIRIMI et al., 2016), buscou-se por estações meteorológicas na área de estudo para verificar o volume de precipitação acumulado, porém não foram encontrados estes dados disponíveis, apenas em outra estação próximo da região. A estação meteorológica Santo Antônio (código 963001) registrou precipitação acumulada igual a 0 mm nos últimos sete dias que antecederam a aquisição das cenas escolhidas (ANA, 2016).

Fez-se o download das imagens do satélite SENTINEL-1A na plataforma *Hybrid Pluggable Processing Pipeline* (HyP3), as quais, são fornecidas com correção geométrica e

radiométrica do terreno (RTC - GAMMA) (HOGENSON et al., 2016). Aplicou-se o filtro polarimétrico *Lee Filter* (LEE; GRUNES; GRANDI, 1999) com janela de 5 x 5 com o objetivo de redução dos ruídos.

Posteriormente, realizou-se a média para cada pixel entre as duas imagens de mesma polarização, buscando dessa forma reduzir os ruídos do uso da cena de uma data apenas e obtendo valores para as polarizações HH, VH e VV. Para cada polarização, gerou-se os índices polarimétricos descritos a seguir: 1) Índice de Degradação de Radar (RFDI); 2) índices de Pope, Rey-Benayas e Paris (1994): CSI, BMI, VSI; 3) coeficiente de retroespalhamento da floresta (σ_{forest}^0), ambos os índices estão descritos com maiores detalhes no Capítulo 2.

A partir da imagem contendo a média do retroespalhamento das cenas adquiridas em datas distintas e com o auxílio do software Envi 5.3, foram geradas as imagens de textura, computadas em uma janela de 5 x 5 e geradas em ambas as direções (MAHDIANPARI; MOTAGH, 2017; OON; AZHAR, 2019), estas são baseadas na Matriz de Co-ocorrência de Nível de Cinza (GLCM) para as bandas HH, VV e VH. As imagens de textura geradas foram: contraste, dissimilaridade, homogeneidade, segundo momento angular (ASM), entropia, média, variância e correlação. Descrições detalhadas de cada medida de textura podem ser obtidas em Haralick (1979). A textura das imagens pode fornecer informações importantes relacionadas às propriedades estruturais e geométricas das florestas que podem ser úteis na discriminação da floresta (DEGRANDI; LUCAS; KROPACEK, 2009).

Realizou-se a calibração radiométrica para a conversão do valor do pixel em retroespalhamento a partir da Equação 2, capítulo 2.

As imagens produto do satélite SENTINEL-2A foram adquiridas com nível de processamento 1C, as quais já foram processadas para as correções radiométricas e geométricas (SENTINEL-2_TEAM, 2015), totalizando 13 bandas, entre elas 4 bandas com resolução espacial de 10 m, 6 bandas com resolução espacial de 20 m e 3 bandas de 60 m. Foi realizada a correção atmosférica para a imagem multiespectral no *semi-automatic plugin* do software Qgis versão 2.18.16, método DOS1 (*Dark Object Subtraction*).

Gerou-se índices de vegetação a partir da imagem multiespectral para servir como informação complementar para a modelagem, sendo eles, Índice de Vegetação (RVI), Índice de Vegetação Normalizada (NDVI), Índice de Vegetação Ajustado ao Solo (SAVI) e Índice de área Foliar (IAF), sendo os cálculos e autores dos índices demonstrados na TABELA 4.

3.3.3 Aquisição e processamento do mosaico ALOS/PALSAR-2

Utilizou-se de informações obtidas do comprimento de onda correspondente a banda L, oriundas do mosaico ortorretificado do satélite ALOS-2, sensor PALSAR-2. Os mosaicos são produtos globais gerados a partir de muitas imagens de radar as quais passam pela aplicação de sofisticadas técnicas de processamento pela Agência de Exploração Aeroespacial do Japão (JAXA). Os mosaicos gerados são disponibilizados gratuitamente (https://www.eorc.jaxa.jp/ALOS/en/palsar_fnf/data/index.htm). Estes produtos são processados pelo algoritmo Sigma-SAR, com correção geométrica e radiométrica, em que, o coeficiente de retroespalhamento é corrigido para amenizar o efeito da inclinação do terreno, sendo disponibilizadas as polarizações HH e HV. Esse conjunto de dados é agregado da observação original com resposta mínima à umidade da superfície (SHIMADA et al., 2014).

Abrangendo a área de estudo, obteve-se o mosaico com resolução espacial de 25 m, imageado em 2015 e identificado como: S09W064. Obteve-se a partir da amplitude do sinal em cada polarização, os índices polarimétricos, Índice de Degradação de Radar (RFDI) e coeficiente de retroespalhamento da floresta (σ_{forest}^0). Após isso, realizou-se a conversão dos valores da amplitude da polarização para seção transversal normalizada em decibéis (dB), a partir da Equação 2, em que para o satélite ALOS/PALSAR-2 o fator de calibração (CF) é de -83.

Como mencionado, o mosaico é constituído por várias cenas adquiridas ao longo do ano, e só é aplicável por se tratar de floresta tropical. Mas, por ser uma floresta estabelecida com baixo incremento anual em biomassa, esse mosaico tende a ser representativo e uma importante variável para a modelagem, seguindo com o objetivo de utilizar de tecnologias disponibilizadas gratuitamente.

3.3.4 Dados LiDAR e Processamento

Foram utilizados dados obtidos com LiDAR de domínio do Projeto Paisagens Sustentáveis, apoiado pela Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), pelo Serviço Florestal dos Estados Unidos (USFS), pela Agência dos Estados Unidos para o Desenvolvimento Internacional (USAID) e pelo Departamento de Estado dos EUA. Os dados LiDAR utilizados neste trabalho referem-se a sobrevoos na área 7, as configurações do voo constam na TABELA 22.

TABELA 22 - ESPECIFICAÇÕES DO AEROLEVANTAMENTO COM LIDAR

Especificações	Características
Data de aquisição	21/09/2015
Datum	Sirgas 2000
Sistema de coordenadas da UTM:	20S
Área total	249,9 ha
Média de densidade de pulsos/m ²	34,03 ppm ²
Altitude média do voo	750 m
Campo de visão	15°
Frequência de varredura	40 Hz
Porcentagem de sobreposição	70%

FONTE: A autora (2020).

Os dados LiDAR foram processados com auxílio do *software* FUSION versão 3.60 (MCGAUGHEY, 2016), que foi desenvolvido pelo Serviço Florestal Norte Americano. O processamento da nuvem de pontos obtidos por LiDAR seguiu uma sequência lógica que pode ser observada na FIGURA 23. Primeiramente os dados foram selecionados conforme a área de interesse, em seguida deu-se início à classificação dos pontos e retirada de possíveis pontos com ruídos, em que foram classificados os pontos pertencentes ao solo para gerar os Modelos Digitais de Terreno (MDT).

Em um segundo momento, as nuvens de pontos obtidos por LiDAR foram submetidas à uma normalização de valores, nesse passo, é obtido os valores reais de cada objeto, como a altura das árvores, gerando o Modelo Digital de Superfície (MDS). Posteriormente, foram extraídas as métricas das nuvens em grids de 100 m, totalizando 286 grids para a área.

Tendo a extração das métricas foram obtidos os valores de AGB por meio da Equação 13. Esta equação foi desenvolvida com base em parcelas de campo alocadas na Floresta Estadual Antimary, localizadas no estado do Acre, resultando em erro RMSE de 40 Mg.ha⁻¹ (D'OLIVEIRA et al., 2012). Esta equação foi utilizada por ser confeccionada com dados do mesmo biomas e ambas as áreas de estudo apresentarem manejo florestal.

$$AGB = (3,119 + 0,564 P25 + 0,062 Var)^2 + 1,74 \quad (13)$$

Em que: AGB = representa a biomassa seca acima do solo (Mg.ha⁻¹, árvores DAP ≥ 10 cm); P25 = equivale aos retornos do primeiro quartil ou 25% percentil; Var = variância da elevação.

Optou-se em utilizar como variável dependente a estimativa da AGB obtida a partir de levantamento com a tecnologia LiDAR principalmente pela indisponibilidade de dados de

campo para o ano de interesse. Para evitar erros de mensuração de variáveis dendrométricas em campo, imprecisão do GPS no interior de florestas, escala espacial das parcelas de medição de campo, emprego de equações alométricas inadequadas, entre outros (THAPA et al., 2015). A utilização da biomassa resultante de dados obtidos com levantamento LiDAR como variável dependente já foi utilizada por Enghart, Keuck e Siegert (2011) que encontraram resultados satisfatórios. Além disso, essa técnica de obtenção de dados biofísicos de floresta tropical já foi comprovada como tendo alta acurácia nas estimativas quando comparada a dados de campo (D'OLIVEIRA et al., 2012; MAUYA et al., 2015; LONGO et al., 2016).

A partir de uma análise prévia do tamanho dos pixels e grid de dados LiDAR, optou-se por utilizar grid com resolução espacial de 100 m. Então, ao invés de simplesmente reamostrar as imagens do satélite SENTINEL-1, SENTINEL-2 e ALOS/PALSAR-2 para essa mesma resolução, foram calculadas algumas estatísticas (mínimo, máximo, média, soma, desvio padrão e variância) englobando os conjuntos de pixels admitidos dentro de cada grid. Corroborando com esta decisão do tamanho de grid, Morel et al. (2011) reamostraram as imagens para a resolução de 100 m e Urbazaev et al. (2018) utilizaram a reamostragem dos vizinhos mais próximos e a média de blocos para agregar os pixels em uma resolução espacial de 100 m.

Morel et al. (2011) destacaram que o tamanho do grid é um fator importante para relacionar dados de campo com retroespalhamento e que grid de 100 m é um tamanho confiável, estando menos propenso a erros relacionado à expansão de variáveis mensuradas em áreas menores e impacto de erros de geolocalização. Garcia et al. (2018) destacaram que o tamanho do pixel também influencia na precisão do modelo, em que encontraram melhores resultados com resolução espacial de 100 m.

3.3.5 Modelagem da biomassa a partir das imagens SENTINEL

A modelagem foi constituída de três tratamentos de variáveis predictoras da AGB, sendo:

- 1) valores de retroespalhamento nas polarizações HH e HV (ALOS/PALSAR-2), VV, HH e VH (SENTINEL-1) (total de 30 variáveis);
- 2) polarizações e subprodutos (texturas GLCM e índices polarimétricos) oriundos do satélite SENTINEL-1 e ALOS/PALSAR-2 (total de 277 variáveis); e

3) as variáveis pré-selecionadas no tratamento 2 com a adição das bandas (01, 02, 03, 04, 05, 06, 07, 08, 08a, 09, 10, 11, 12) e subprodutos (LAI, RVI, SAVI, NDVI) oriundos do satélite SENTINEL-2 (total de 100 variáveis).

Visando construir um modelo parcimonioso, foi utilizado o algoritmo CFS (CfsSubsetEval) implementado no pacote Fselector (ROMANSKI; KOTTHOFF, 2018) do *software* R, para encontrar as melhores variáveis e suas representações (mínimo, máximo, média, desvio padrão, variância) para os três tratamentos.

O CFS é uma das famosas abordagens de mineração de dados que avalia o valor de um subconjunto de atributos considerando a capacidade preditiva individual de cada característica, juntamente com o grau de redundância entre elas. A principal ideia do CFS é que boas variáveis incluem valores altamente correlacionados com a classe, mas não entre si. Sendo assim, os subconjuntos de dados são primeiro selecionados usando um algoritmo de busca heurística juntamente com uma função de correlação de Pearson (HALL, 1999).

Iniciou-se a modelagem pelo método comumente empregado, a regressão linear múltipla, porém, os pressupostos da regressão não foram atendidos, além de que a regressão não foi capaz de ajustar a relação entre as informações obtidas por sensoriamento remoto e a AGB da floresta tropical. Por isso, constatou-se a necessidade da aplicação de métodos mais complexos, como os de inteligência artificial ou também conhecidos como aprendizado de máquina.

Foram selecionados cinco modelos de aprendizado de máquina para este estudo, sendo eles, *model tree* (M5), *Random Forest* (RF), *Weighted K-Nearest Neighbor* (wKNN), *Artificial Neural Network* (ANN) e *Support Vector Regression* (SVM). Todos os algoritmos desses modelos foram implementados no pacote Caret (KUNH, 2008) do *software* R, com o auxílio do script disponibilizado por Souza (2020).

Os dados utilizados para a modelagem foram padronizados para a mesma escala, tendo média igual a 0 e desvio padrão igual a 1. Posteriormente foi realizado o treinamento a partir da busca dos hiperparâmetros ótimo contendo a amplitude constante na Tabela 23.

TABELA 23 - MODELOS DE APRENDIZADO DE MÁQUINA UTILIZADOS PARA A MODELAGEM DA AGB DA FLORESTA TROPICAL

Modelo	Amplitude de Hiperparâmetros	Método	Pacote do R	Autor
<i>Model Tree</i> (M5)	Poda = sim; não	M5	RWeka	Hornik; Buchta; Zeileis (2009)
<i>Random Forest</i> (RF)	Mtry = (1:13); ntree = (50, 100, 150, 200, 500)	rf	randomForest	Liaw; Wiener (2002)
<i>Weighted K-Nearest Neighbor</i> (wKNN)	Kmax = seq (2, 150, 2); distance = 1:3; Kernel = rectangular, biweight, cos, epanechnikov, gaussian, inv, optimal, rank, triangular, triweight	kknn	kknn	Schliep; Hechenbichler; Lizee (2016)
<i>Artificial Neural Network</i> (ANN)	Size = seq (1, 10, 1); Decay = seq (0.1, 0.7, 0.1)	nnet	nnet	Venables; Ripley (2002)
<i>Support Vector Regression</i> (SVR)	Sigma = seq (0.001, 10, 0.1); C = seq (1, 100, 10)	svmRadial	Kernlab	Karatzoglou; Smola e Hornik (2004)

FONTE: A autora (2020).

O grid composto de 286 amostras, fora treinado e avaliado a partir da validação cruzada *k-folds*, na qual, a amostra original é dividida aleatoriamente em *k* sub-amostras, definiu-se $k = 10$. Esse processo consiste que uma única sub-amostras *k* é mantida para validar o modelo e as restantes *k-1* são usadas para treinamento do modelo. Esse processo é então repetido *k* vezes, com cada uma das sub-amostras para validação.

A avaliação do desempenho dos modelos de aprendizado de máquina fora por meio de algumas das estatísticas descritas no item 2.3.5 do Capítulo 2, sendo a Raiz do Erro Médio Quadrático (RMSE) na forma absoluta e relativa, correlação de Spearman (ρ) e Viés na forma relativa.

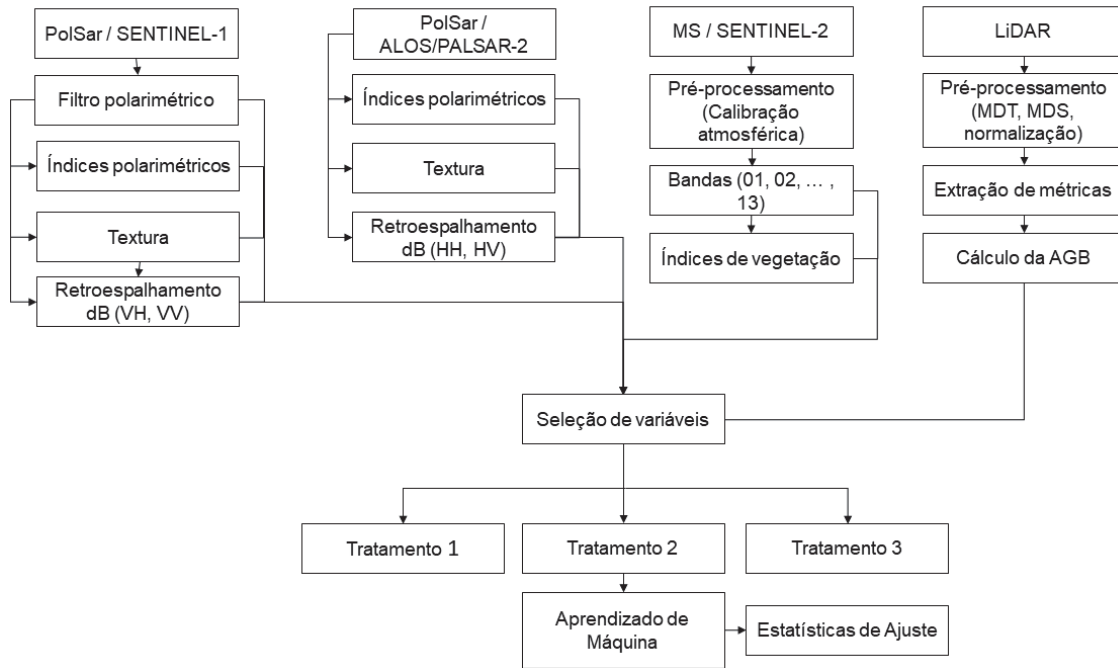
O erro total foi calculado assumindo fontes de erros independentes proposta pela metodologia de Saatchi et al. (2011), como mostrado na Equação 1.

$$\varepsilon_{AGB} = \left(\varepsilon_{measurement}^2 + \varepsilon_{allometry}^2 + \varepsilon_{sampling}^2 + \varepsilon_{prediction}^2 \right)^{1/2}, \quad (1)$$

Em que, $\varepsilon_{measurement}$: é o erro de medição de inventário; $\varepsilon_{allometry}$: é o erro associado a equações alométricas para obter AGB; $\varepsilon_{sampling}$: é o erro espacial no aumento de escala de estimativas derivadas de lidar; $\varepsilon_{prediction}$: é o erro de previsão.

Um resumo dos dados disponível e da metodologia aplicada pode ser observado na FIGURA 23.

FIGURA 23 - FLUXO DE TRABALHO REALIZADO COM OS DADOS DISPONÍVEIS NESTE ESTUDO



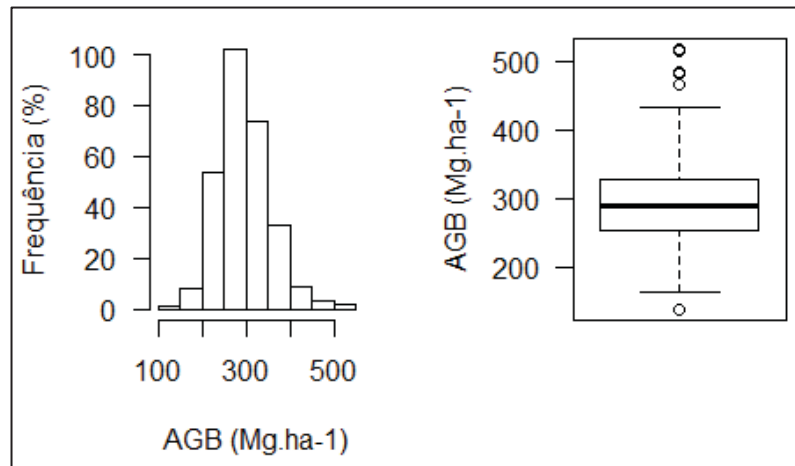
FONTE: A autora (2020).

3.4 RESULTADOS

3.4.1 Banco de dados e seleção de atributos

A AGB dos 286 grids variou de 139,3 a 516,7 Mg.ha⁻¹, com valores médios de 296 Mg.ha⁻¹ e coeficiente de variação em 20%. A distribuição da AGB é platicúrtica (curtose 1,075) com assimetria à direita (0,672) (FIGURA 24). O histograma da distribuição dos valores da AGB demonstra que os valores extremos ocorrem com menor frequência e a AGB próximo de 300 Mg.ha⁻¹ é comumente mensurada nessa área de estudo.

FIGURA 24 - DISTRIBUIÇÃO DOS VALORES DE AGB REPRESENTADOS SOBRE UM HISTOGRAMA DE FREQUÊNCIA E BOXPLOT



FONTE: A autora (2020).

LEGENDA: AGB: biomassa aérea.

Para o tratamento 1, das 30 variáveis apresentadas, o algoritmo CFS selecionou apenas três variáveis para compor o experimento. No tratamento 2, das 277 variáveis, 18 foram selecionadas, e para o tratamento 3 das 100 apenas nove foram consideradas relevantes de acordo com o algoritmo CFS. As variáveis selecionadas para compor os tratamentos estão listadas na TABELA 24, juntamente com a correlação linear de Spearman entre as variáveis independentes com a dependente (AGB).

TABELA 24 - CORRELAÇÃO DE SPEARMAN (ρ) ENTRE AGB E AS VARIÁVEIS PREDITORAS SELECIONADAS PELO ALGORITMO CFS

Tratamento 1	ρ	Tratamento 2	ρ	Tratamento 3	ρ
SD (C_{VV})	0,067	Ave (Méd C_{VV})	0,115	SD (banda 8-S2)	0,401
Máx (C_{VV})	0,052	Min (Méd C_{VV})	0,103	Var (banda 8-S2)	0,401
Ave (L_{HH})	0,049	SD (C_{VV})	0,067	SD (banda 3-S2)	0,373
		Máx (Homog C_{VH})	0,040	Var (banda 4-S2)	0,271
		Max (Méd C_{HH})	0,011	Ave (Méd C_{VV})	0,115
		Min (Homog C_{VH})	-0,001	Max (Homog C_{VH})	0,040
		SD (Homog L_{HH})	0,086	Min (banda 1-S2)	0,150
		Máx (Dis L_{HH})	0,115	Ave (Cont C_{HH})	-0,111
		Ave (Cont C_{HH})	-0,111	Min (Méd L_{HH})	0,078
		Ave (Homog C_{VH})	0,027		
		Min (Méd L_{HH})	0,078		
		Min (Cont C_{VV})	-0,141		
		Min (Dis C_{VV})	-0,141		
		Var (C_{VH})	0,119		
		Min (Homog L_{VH})	-0,001		
		Min (Cor C_{VV})	0,132		
		Sum (Cont C_{VH})	-0,021		
		Máx (ASM C_{VV})	0,146		

FONTE: A autora (2020).

LEGENDA: Var: representa a variância da variável abrangendo o grid de 100 m; Ave: representa a média da variável dentro do grid; SD: desvio padrão da variável; Max: máximo valor; Min: mínimo valor; S2: satélite SENTINEL-2; HH, VV, VH: polarizações que podem ser para a banda C ou L; ASM: textura segundo momento angular; Cont: textura de contraste; Cor: textura de correlação; Dis: textura de dissimilaridade; Homog: textura de homogeneidade; Méd: textura de média; Tratamento 1: valores de retroespalhamento nas polarizações HH e HV (ALOS/PALSAR-2), VV, HH e VH (SENTINEL-1); Tratamento 2: polarizações e subprodutos oriundos do satélite SENTINEL-1 e ALOS/PALSAR-2; Tratamento 3: as variáveis pré-selecionadas no tratamento 2 com a adição das bandas e subprodutos oriundos do satélite SENTINEL-2.

Para compor o tratamento 1, a polarização VV da banda C foi selecionada em duas representações, o máximo valor (Max) e o desvio padrão (SD), e para a banda L o algoritmo CFS selecionou a média (Ave) da polarização HH. Em média o valor das polarizações selecionadas apresenta maior correlação com a AGB do que as demais polarizações, sendo que para a banda C a VV apresenta $\rho = 0,052$ e para HH da banda L a correlação de 0,049.

No tratamento 2 nenhum dos índices polarimétricos foi selecionado para a modelagem. Sendo constituído por representações do retroespalhamento nas polarizações VH e VV (Var (C_{VH}) e SD (C_{VV}), mais 16 variáveis que representam a textura GLCM. Em que, somente quatro

destes é referente as polarizações da banda L, demonstrando a relevância que a textura e as polarizações da banda C possuem para a modelagem.

O tratamento 3 é constituído por quatro representações de textura, cinco representações oriundas das bandas multiespectrais e nenhum índice de vegetação foi selecionado para a modelagem. Acredita-se que o fato de os índices de vegetação não terem sido selecionados estejam indicando uma saturação dos valores dos índices, característico de situações como da área de estudo em que há altos valores de AGB.

A correlação linear de Spearman demonstrou que a relação entre AGB e as variáveis selecionadas pelo algoritmo CFS é fraca, sendo as maiores correlações associadas as representações das bandas multiespectrais, banda 8 ($\rho = 0,401$) e banda 3 ($\rho = 0,373$).

3.4.2 Modelagem da biomassa

A média da AGB para as 286 parcelas foi de 297 Mg.ha⁻¹, a modelagem com o tratamento 1 resultou em RMSE variando de 57 Mg.ha⁻¹ (wKNN) a 60 Mg.ha⁻¹ (M5) e viés de 0,14% (RF) e 1,218% (SVM). O melhor desempenho para o tratamento 1 é atribuído ao modelo constituído pelo wKNN e o pior para o M5 (TABELA 25).

TABELA 25 - DESEMPENHO DOS ALGORITMOS FRENTE ÀS VARIÁVEIS PREDITORAS DO TRATAMENTO 1, UTILIZAÇÃO DOS VALORES DE RETROESPALHAMENTO DAS BANDAS C E L

Modelos	Melhor ajuste	RMSE (Mg.ha ⁻¹)	RMSE (%)	Viés (%)	Rank
M5	Poda = sim	60,305	20,351	0,212	3°
RF	Mtry = 1; ntree = 150	59,398	20,040	-0,140	2°
wKNN	Kmax = 54; Kernel = inv, dist = 2	57,066	19,253	0,267	1°
ANN	Size = 1; decay = 0,4	57,947	19,553	-0,271	3°
SVR	Sigma = 0,001; C = 21	57,780	19,496	-1,218	3°

FONTE: A autora (2020).

LEGENDA: M5: *Model Tree*; RF: *Random Forest*; wKNN: *Weighted K-Nearest Neighbor*; ANN: *Artificial Neural Network*; SVR: *Support Vector Regression*.

O tratamento 2 que foi composto por informações de retroespalhamento, textura e índices polarimétricos, resultou em RMSE variando de 55 Mg.ha⁻¹ (wKNN) a 61 Mg.ha⁻¹ (M5) e viés entre 0,217 % (RF) e 1,714% (SVM). O RF foi o modelo que apresentou o melhor desempenho para a modelagem, e o pior modelo da mesma forma que para o tratamento 1, foi o M5 (TABELA 26).

TABELA 26 - DESEMPENHO DOS ALGORITMOS FRENTE ÀS VARIÁVEIS PREDITORAS DO TRATAMENTO 2, ADIÇÃO DE VARIÁVEIS DE TEXTURA E ÍNDICES POLARIMÉTRICOS

Modelos	Melhor ajuste	RMSE (Mg.ha ⁻¹)	RMSE (%)	Viés (%)	Rank
M5	Poda = sim	61,383	20,708	0,791	5°
RF	Mtry = 2; ntree = 500	55,900	18,859	0,217	1°
wKNN	Kmax = 151; Kernel = triangular, dist = 3	55,736	18,802	-0,592	2°
ANN	Size = 1; decay = 0,4	58,617	19,774	0,344	3°
SVR	Sigma = 0,001; C = 21	55,956	18,873	-1,714	4°

FONTE: A autora (2020).

LEGENDA: M5: *Model Tree*; RF: *Random Forest*; wKNN: *Weighted K-Nearest Neighbor*; ANN: *Artificial Neural Network*; SVR: *Support Vector Regression*.

Comparando os resultados do tratamento 1 para o 2, observa-se em média uma sutil melhora no desempenho dos modelos (0,981 Mg.ha⁻¹). A melhoria com maior impacto é sentida pelo modelo RF (3,498 Mg.ha⁻¹), seguido pelo SVM (1,82 Mg.ha⁻¹), porém o modelo M5 apresentou desempenho contrário, com piora ao utilizar as informações de textura (1,079 Mg.ha⁻¹).

Ao unir todas as variáveis selecionadas do tratamento 2 para a seleção do tratamento 3, mais as informações oriundas das bandas multiespectrais e seus índices de vegetação derivados, o desempenho dos modelos de aprendizado de máquina apresentou melhoria. O RMSE variou entre 48 (M5) e 52 (ANN) e viés de 0 (M5) a 1,797% (SVM). Para o tratamento 3 o modelo M5 que vinha apresentando o pior desempenho foi o melhor, e o pior desempenho é atribuído a ANN (TABELA 27).

TABELA 27 - DESEMPENHO DOS ALGORITMOS FRENTE ÀS VARIÁVEIS PREDITORAS DO TRATAMENTO 3. USO DO RETROESPALHAMENTO, TEXTURA COM A ADIÇÃO DAS BANDAS E SUBPRODUTOS DO SATÉLITE SENTINEL-2

Modelos	Melhor ajuste	RMSE (Mg.ha ⁻¹)	RMSE (%)	Viés (%)	Rank
M5	Poda = sim	48,284	16,297	0,000	1°
RF	Mtry = 1; ntree = 200	50,006	16,873	0,210	2°
wKNN	Kmax = 14; Kernel = gaussiano, dist = 3	49,724	16,781	-0,577	3°
ANN	Size = 1; decay = 0,1	52,503	17,717	-0,330	5°
SVR	Sigma = 0,001; C = 11	48,900	16,502	-1,797	3°

FONTE: A autora (2020).

LEGENDA: M5: *Model Tree*; RF: *Random Forest*; wKNN: *Weighted K-Nearest Neighbor*; ANN: *Artificial Neural Network*; SVR: *Support Vector Regression*.

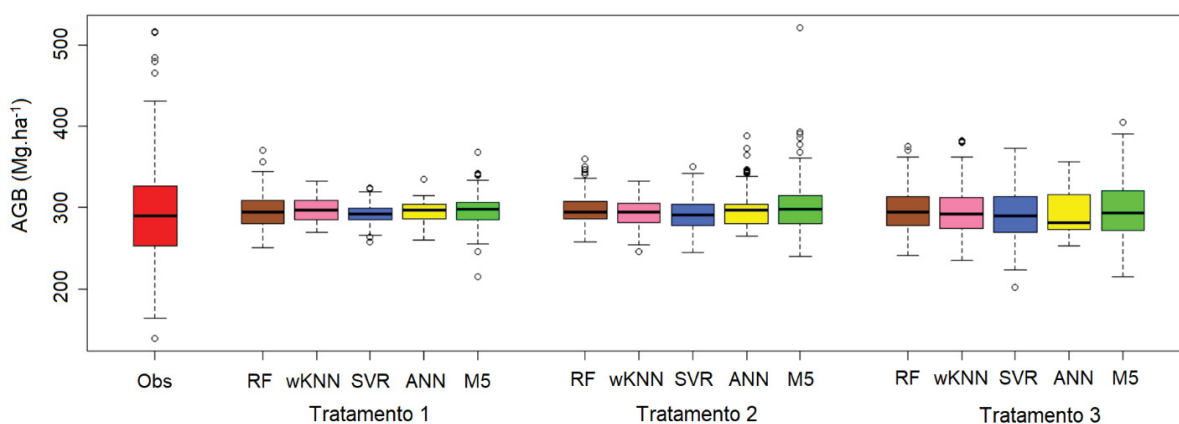
Observa-se que a adição das informações multiespectrais trouxeram melhorias ao desempenho dos modelos de aprendizado de máquina, em média 7%, sendo mais sentida pelo

modelo M5 (13,1 Mg.ha⁻¹), seguida pelo SVM (7,05 Mg.ha⁻¹), o modelo wKNN que já vinha apresentando o melhor desempenho não apresentou tanto ganho adicional, porém ainda com melhoria de 6,02 Mg.ha⁻¹ no tratamento 3.

Corroborando com as Tabelas 25, 26 e 27, ao analisar o boxplot dos valores estimados (FIGURA 26), percebe-se uma melhoria na amplitude dos valores simulados pelos modelos de aprendizado de máquina ao adicionar as informações de textura (tratamento 2) e melhoria ainda mais significativa ao adicionar as informações multiespectrais (tratamento 3).

Pelas estatísticas de desempenho, o modelo de melhor desempenho para o tratamento 1 é o wKNN, para o tratamento 2 os modelos RF e wKNN e para o tratamento 3 é o modelo M5. Porém, ao analisar a amplitude de valores mensurados e os simulados pelos modelos de aprendizado de máquina (FIGURA 25), percebe-se que a maioria deles não consegue representar a amplitude de valores, simulando valores intermediários, quase que repetitivos, independente dos valores das variáveis predictoras.

FIGURA 25 - BOXPLOT DOS VALORES OBSERVADOS E ESTIMADOS EM CADA UM DOS TRATAMENTOS



FONTE: A autora (2020).

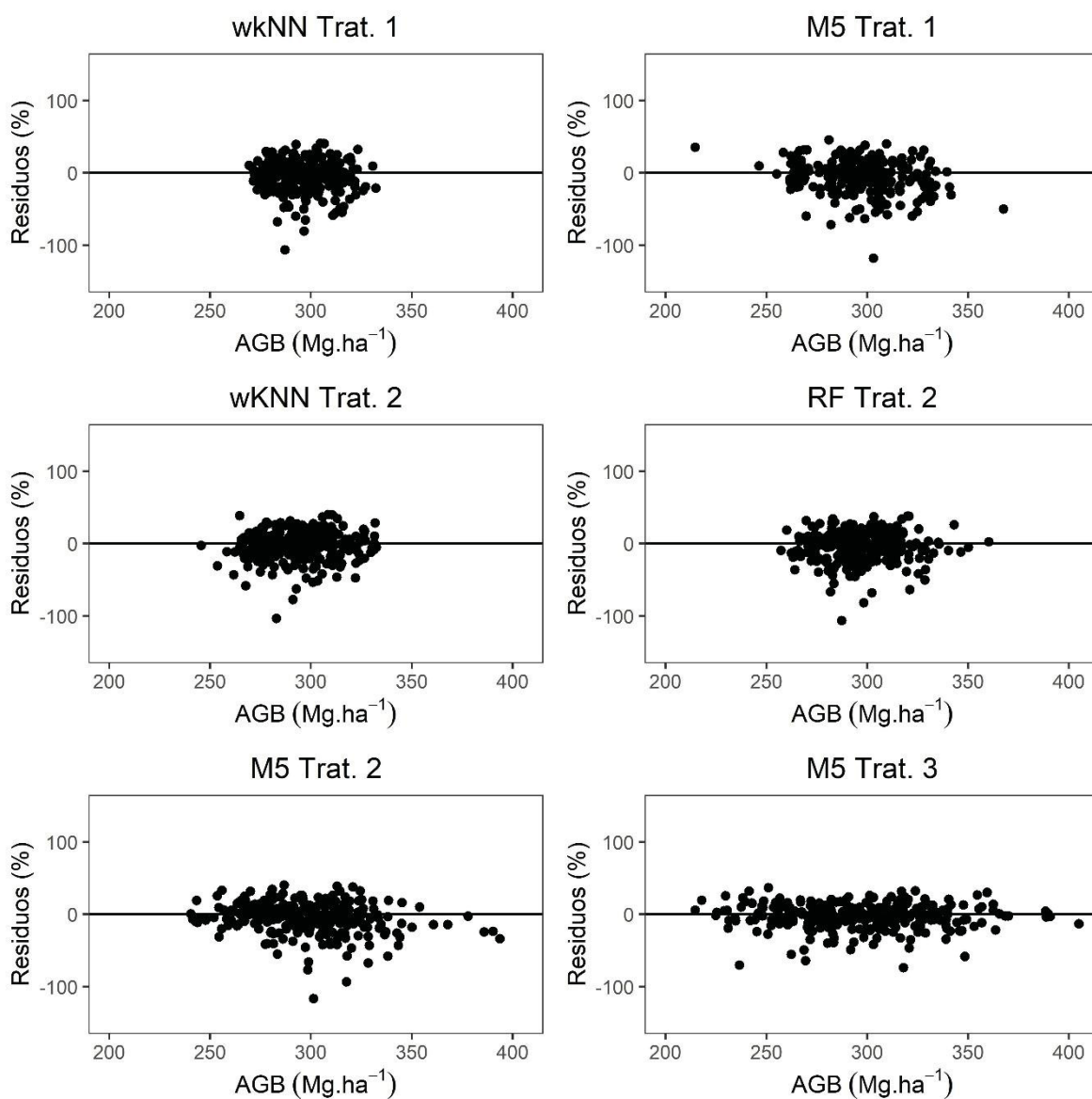
LEGENDA: Tratamento 1: valores de retroespalhamento nas polarizações HH e HV (ALOS/PALSAR-2), VV, HH e VH (SENTINEL-1); Tratamento 2: polarizações e subprodutos oriundos do satélite SENTINEL-1 e ALOS/PALSAR-2; Tratamento 3: as variáveis pré-selecionadas no tratamento 2 com a adição das bandas e subprodutos oriundos do satélite SENTINEL-2.

A AGB extraída a partir das métricas LiDAR apresenta amplitude de 377,4 Mg.ha⁻¹, no tratamento 1, 2 e 3 o modelo M5 foi o que estimou valores de AGB que mais se aproximou da amplitude de valor desejada (153, 280,8 e 190,2 Mg.ha⁻¹).

Observa-se pela dispersão dos resíduos que para o tratamento 1 o modelo wKNN estimou os valores de AGB próximo aos valores médios. No tratamento 2, os modelos wKNN

e RF apresentaram comportamento semelhante entre si, com valores estimados próximo ao valor médio da AGB. Porém, entre estes dois algoritmos, percebe-se que o RF possui valores levemente dispersos do centro das estimativas. A dispersão dos resíduos gerados pelo M5 no tratamento 3 demonstra que é formado pelo melhor conjunto de variáveis de entrada e o que possui as melhores estimativas. Para os três tratamentos os valores estimados pelo modelo M5 são mais próximos a amplitude de valores observados (FIGURA 26).

FIGURA 26 - DISPERSÃO GRÁFICA DOS RESÍDUOS GERADOS PELOS MODELOS DE APRENDIZADO DE MÁQUINA COM MELHOR DESEMPENHO, SOBRE O EIXO X QUE REPRESENTA OS VALORES ESTIMADOS



FONTE: A autora (2020).

LEGENDA: Trat. 1: valores de retroespalhamento nas polarizações HH e HV (ALOS/PALSAR-2), VV, HH e VH (SENTINEL-1); Trat. 2: polarizações e subprodutos oriundos do satélite SENTINEL-1 e ALOS/PALSAR-2; Trat. 3: as variáveis pré-selecionadas no tratamento 2 com a adição das bandas e subprodutos oriundos do satélite SENTINEL-2.

Para o cálculo do erro total, não foram considerados erros de medição devido à falta de dados de inventário. Foram considerados os seguintes erros: erro de alometria de 40,2 Mg.ha⁻¹ ou 17,47% considerando o AGB total de 230 Mg.ha⁻¹ (D'Oliveira et al. 2012); erro amostral de 22,8% (SAATCHI et al. 2011) e erro de previsão de 16% para o algoritmo *Model Tree*. O total de erros seguindo a metodologia do presente estudo é de aproximadamente 40%.

3.5 DISCUSSÃO

3.5.1 Análise do banco de dados

O banco de dados utilizado para esse estudo resultante de mensurações da Floresta do Jamari apresenta AGB inferior ao encontrado por Kuplich, Curran e Atkinson (2005), média de $387 \text{ Mg}\cdot\text{ha}^{-1}$ de AGB, para uma floresta estabelecida no Tapajós, situada também na Floresta Amazônica. Esse valor inferior de AGB pode estar relacionado a intensidade e data de manejo aplicada a essa área da Flona do Jamari.

Os valores de retroespalhamento das bandas C e L correlacionou-se com a AGB de forma fraca, acredita-se que tenha ocorrido saturação dessas bandas devido aos altos valores de AGB encontrados na área de estudo. Da mesma forma, para uma área próxima da cidade de Manaus, também na floresta tropical brasileira, Cutler et al. (2012) encontraram correlação de 0,16 entre a AGB e dados PolSar de banda L (satélite JERS-1). A banda L apresenta comprimento de onda mais apropriado para as estimativas de AGB comparado a banda C devido a sua maior capacidade de penetração no dossel da floresta, mesmo assim, Lucas et al. (2016) relata também sobre saturação da banda L ao realizar estimativas em florestas tropicais.

O resultado de baixa correlação não é o mesmo verificado em outras coberturas do solo com baixa AGB. Tal como apresentado por Castillo et al. (2017), os quais avaliaram áreas de usos não florestais (agricultura, aquicultura e áreas urbanas) e manguezais, os quais encontraram valores de correlação entre AGB e retroespalhamento mais fortemente correlacionados, variando de 0,28 - 0,84, para dados do satélite SENTINEL-1.

Das variáveis selecionadas para compor os tratamentos, a medida de variância e desvio da banda 8 (infravermelho próximo) apresentou a maior correlação com AGB ($r = 0,423$). A relevância da banda 8 é justificada por esse comprimento de onda realçar a vegetação, a qual é utilizada no computo da maioria dos índices de vegetação. Resultados mais fortemente correlacionados ($r = 0,69-0,72$) foram encontrados por Castillo et al. (2017) ao relacionar as bandas multiespectrais do satélite SENTINEL-2 com a AGB de manguezais. Acredita-se que esse fato esteja relacionado a complexidade da estrutura florestal em áreas de floresta tropical, comparada a áreas de manguezais.

A seleção de variáveis para compor o conjunto de entrada na modelagem da AGB tem papel fundamental, no qual Almeida et al. (2019) demonstraram que é possível reduzir consideravelmente o número de informações de entrada para modelos de aprendizado de máquina sem perder muito a precisão das estimativas da AGB em floresta tropical. Ao tentar modelar um problema baseado em dados oriundos de várias fontes é essencial o equilíbrio entre

a dimensionalidade dos dados, perda de informações e parcimônia dos modelos a serem construídos, este foi o intuito ao realizar a seleção de variáveis pelo algoritmo CFS.

3.5.2 Modelagem da biomassa

Os algoritmos utilizados apresentaram desempenho numericamente semelhante para o mesmo tratamento de dados de entrada. Acredita-se que este fato esteja relacionado ao alto potencial que estes apresentam, conseguindo desempenho semelhante na modelagem. O tratamento 1 que é composto por dados de retroespalhamento das bandas C e L proporcionou bom desempenho, sendo que o menor RMSE foi de 19% para o modelo wKNN. Verificou-se que a correlação das polarizações de banda L apresentaram menor correlação com a AGB, comparado aos da banda C. Porém, pode-se comprovar que a banda L agrega informações relevantes devido ao desempenho quando comparado com o relatado por Debastiani et al. (2019), que utilizaram dessa mesma base de dados, com exceção dos dados do mosaico ALOS/PALSAR-2, e encontraram desempenho menos significativo e obtiveram o melhor desempenho ao utilizarem de regressão robusta.

A utilização dos índices polarimétricos e da textura GLCM associada as polarizações da banda C e L (tratamento 2) proporcionaram uma melhora nas estimativas de AGB pela maioria dos algoritmos, acredita-se que a textura quebre a limitação da saturação da banda e agregue informações relevantes. Corroborando com estes resultados, Cutler et al. (2012) ajustaram uma série de RNAs para a estimativa de AGB da floresta tropical, utilizando para tal, dados do Brasil, Malásia e Tailândia. Utilizaram dados multiespectrais do satélite Landsat TM, imagens SAR do satélite JERS-1 e imagens de textura deste (análise wavelet e GLCM). Concluíram que a inclusão das informações de textura das imagens SAR aos dados multiespectrais, melhora a relação com a AGB. Kuplich, Curran e Atkinson (2005) também ressaltaram que a adição da textura as imagens de retroespalhamento aumentam o potencial da utilização de imagens SAR para estimar a biomassa de floresta tropical. Estes autores nos motivaram a adicionar ao nosso experimento tal informação, fato que pudemos comprovar.

Erros menores aos gerados pelos tratamentos 1 e 2 foram descritos por Thapa et al. (2015), estes utilizaram de dados SAR derivadas do satélite ALOS/PALSAR e imagens de textura para a estimativa do estoque de carbono acima do solo na Indonésia, por meio de regressão linear múltipla. Somente com as polarizações das imagens SAR os autores conseguiram ajustar a regressão com RMSE de 45,64 Mg.C.ha⁻¹, e com a adição das métricas de textura os modelos melhoraram, apresentando RMSE entre 28,01 a 37,70 Mg.C.ha⁻¹.

Para o tratamento 3 em que todas as variáveis de maior relevância participaram da modelagem, foi perceptível a melhora nos resultados com média de 7%, porém é muito mais representativo visualmente, em que os algoritmos conseguiram estimar a AGB com uma maior amplitude dos valores. Pode-se atribuir essa melhora às informações de bandas como a verde e infravermelho próximo, explicadas pela correlação da refletância nesse comprimento de onda (560 e 842 nm) e a quantidade de material fotossintético, as quais tem grande relevância para a caracterização da vegetação.

Para cada tratamento, as estatísticas de ajuste apresentaram bons resultados e com valores próximos entre os modelos. Sendo para o tratamento 1 as melhores estimativas foram provenientes do modelo wKNN, para o tratamento 2 os modelos wKNN e RF e para o tratamento 3 o modelo M5. Porém essas estatísticas de desempenho não se mostraram apropriadas para avaliar o desempenho já que mensuram a precisão e não a acurácia, sendo que em média os valores estimados são semelhantes com a média da AGB observada na floresta. No entanto, os modelos não conseguem representar os extremos da amplitude de valores que foram utilizados para o ajuste. Visualmente pode-se concluir que o modelo M5 foi o melhor, em média também apresentou valores estimados parecidos entre cada observação, mas sendo capaz de representar uma maior amplitude de valores. Esse comportamento dos modelos acaba afetando as estatísticas que não demonstram grandes erros médios, podendo levar a conclusões precipitadas sobre a real acurácia.

É notável o baixo RMSE para os três tratamentos e o bom desempenho dos algoritmos de aprendizado de máquina, resultado também constatado por Garcia et al. (2018) que ressaltam que estas técnicas são atraentes porque não fazem suposições explícitas sobre a distribuição dos dados e permitem modelar relações complexas entre as variáveis independentes sobre a dependente. Além disso, esses algoritmos têm a vantagem de serem versáteis e flexíveis, além de não terem que levar em conta alguns pressupostos de regressão, como normalidade dos dados e homocedasticidade da variância (SANQUETTA et al., 2015).

Estatisticamente o RMSE variou entre 48 a 61 Mg.ha⁻¹ no presente estudo. Resultados menos expressivos foram encontrados por Morel et al. (2011) que ajustaram um modelo de regressão para mapear a AGB de áreas ocupadas por dendezeiros, a partir da polarização HV do satélite ALOS/PALSAR, encontraram RMSE de 125 Mg.ha⁻¹. Da mesma forma, Ningthoujam, Joshi e Roy (2018) estimaram por meio de regressão a AGB de uma floresta tropical decídua mista, utilizando o mosaico de dados do satélite ALOS/PALSAR. Os autores encontraram valores de RMSE variando de 92 a 94 Mg.ha⁻¹ ao utilizar como entrada as

polarizações HH e HH+HV, e RMSE superior a 107 Mg.ha⁻¹ ao empregar apenas a polarização HV.

Ao aplicar metodologia semelhante a este estudo, Castillo et al. (2017) encontraram valores menores de RMSE ao utilizarem de dados do satélite SENTINEL, regressão linear e aprendizado de máquina para estimar e mapear a AGB de manguezais e áreas não florestais nas Filipinas. Ao utilizarem apenas as polarizações do SENTINEL-1, encontraram RMSE variando entre 27 e 55 Mg.ha⁻¹, e ao utilizarem apenas das bandas multiespectrais do satélite SENTINEL-2 encontraram valores de RMSE variando de 27 a 48 Mg.ha⁻¹, pequena melhoria ao comparar apenas com a utilização dos dados SAR, corroborando com os resultados obtidos no presente estudo.

Estes resultados encontrados por Castillo et al. (2017) podem ser justificados pela conclusão de Urbazaev et al. (2018) que utilizaram de uma abordagem de aprendizado de máquina, a árvore híbrida Cubist, para a estimativa de AGB no México. Para tal, utilizaram das imagens polarizadas da banda L do satélite ALOS/PALSAR, DEM oriundo do projeto SRTM, mosaico de imagens do satélite Landsat e métricas obtidas por LiDAR como variáveis predictoras. Observaram uma subestimação dos maiores valores de AGB, o qual justificaram pela sensibilidade insuficiente dos dados de satélite para estimar a alta AGB.

A quantificação do erro total considerando a metodologia de propagação de erros de Saatchi et al. (2011) foi de aproximadamente 40%. Além do erro mensurado pelas estatísticas de desempenho, acredita-se que uma parte não quantificável de erros ao estimar a AGB estejam relacionados às condições de umidade do solo e da vegetação a qual afeta os valores de retroespalhamento, a umidade não foi totalmente contabilizada pela falta de estações meteorológicas com maior proximidade a área de estudo. Outra questão que pode influenciar e deve ser levada em consideração é que apenas parte da vegetação de porte arbóreo (acima de determinado diâmetro mínimo de inclusão) é captada pelo LiDAR, deixando, portanto, parte do estrato da árvore não sendo considerada.

Mesmo com as limitações do SAR de banda C e L, com a sua integração juntamente com informações multiespectrais foi possível obter estimativas precisas. Para romper a limitação de saturação é indicada a utilização da banda P, pois esta apresenta maior penetração e sensibilidade com a AGB (SAATCHI et al., 2017). Mas infelizmente nenhum sensor dos satélites atuais disponibilizam de forma gratuita imagens com tal característica. Está previsto para 2020 o lançamento do satélite BIOMASS que abrange esta faixa espectral, e irá fornecer informações quantitativas sobre a estrutura da floresta a partir do espaço (LE TOAN et al.,

2011), até que estejam disponíveis tais dados ainda são necessários estudo de metodologias alternativas que considerem os dados que são disponibilizados gratuitamente.

3.6 CONCLUSÃO

Dados PolSar de banda C e L aliados a modelos de aprendizado de máquina produzem estimativas precisas da AGB. Informações de textura SAR e as oriundas das imagens multiespectrais agregam melhoria ao desempenho dos modelos.

A integração de dados PolSar de banda C e L com imagens multiespectrais apresentam potencial para estimativa de AGB em floresta tropical. A utilização dessa tecnologia demonstra que é possível quantificar biomassa florestal com precisão (RMSE: 48 Mg.ha⁻¹ ou 16%) e baixo custo.

O algoritmo *Model Tree* (M5) destaca-se sobre os demais em todos os tratamentos de variáveis de entrada, sendo capaz de se aproximar mais dos valores extremos de AGB encontrados na floresta tropical utilizada como área de estudo.

Este estudo indica resultados satisfatórios e demonstra o potencial da utilização de dados obtidos gratuitamente, aliados a técnicas de aprendizado de máquina para a estimativa de AGB. A metodologia permite o monitoramento em larga escala, regular e de baixo custo para os ecossistemas florestais tropicais.

DISCUSSÃO GERAL

Com a revisão bibliométrica do capítulo 1 foi possível obter um panorama geral dos artigos científicos indexados nas bases de dados *Web of Science* e Scopus. No total foram analisados 159 artigos relacionados a estimativa de biomassa florestal utilizando SAR, publicado desde 1991 até o final do ano 2018.

O país líder em produtividade científica sobre esse tema é os EUA, seguido pela China. O país que mais tem suas florestas estudadas é a China. Os autores que mais estudam sobre o tema são Solberg, Saatchi e Sun. A revista mais procurada para publicação é a *Remote Sensing of Environment* e a banda mais empregada é a banda L.

A revisão bibliométrica deu suporte de uma visão como o todo, a qual possibilitou a escolha de alguns aspectos dos demais capítulos, além de esclarecer o que se tem de mais novo com o uso da tecnologia SAR.

Com a revisão bibliométrica, verificou-se que as florestas plantadas recebem pouca atenção nos estudos de estimativa de biomassa florestal utilizando SAR. Dada a importância do setor de florestas plantadas no Brasil e os poucos estudos nessa área, buscou-se no Capítulo 2 por abordar sobre estimativas de variáveis dendrométricas em plantações de eucalipto. Para isso, buscou-se preencher a lacuna da integração de diferentes fontes de dados obtidos por sensoriamento remoto e explorar mais o potencial da banda L que passará a ter disponibilidade gratuita em breve.

A estimativa de variáveis dendrométricas de plantações florestais a partir de dados obtidos por sensoriamento remoto ativo e passivo se mostrou possível, a precisão das estimativas se mostraram dependentes do conjunto de entrada (tratamento) para a modelagem.

As estimativas a partir da banda C foram satisfatórias, porém a banda L se mostrou mais apropriada para a estimativa das variáveis dendrométricas. Os índices polarimétricos, textura GLCM e índices de vegetação se mostraram variáveis relevantes para a modelagem. Utilizou-se a variável idade em todos os tratamentos de variáveis de entrada por acreditar que seja uma informação fácil de se obter, a partir de séries históricas de imagens de satélite.

As estimativas de área basal, volume e biomassa aérea apresentaram coeficiente de determinação ajustado ($R^2_{aj.}$) superior a 0,65 e erro padrão da estimativa (S_{YX}) inferior a 32% com o uso de modelos de regressão linear múltipla. Os modelos de regressão linear são modelos fáceis de interpretar, replicar e apresentaram bons resultados, porém, buscando por melhor desempenho nas estimativas, realizou-se a modelagem utilizando aprendizado de máquina.

A estimativa da área basal, volume e biomassa aérea para as parcelas de eucalipto melhoraram consideravelmente utilizando dos modelos de aprendizado de máquina. O *Random Forest* foi o modelo que apresentou o melhor desempenho, e quando comparado a regressão linear, apresentou melhora nas estimativas em aproximadamente 8% para área basal, 13% para o volume e 15% para a biomassa. O erro padrão das estimativas das variáveis dendrométricas utilizando o modelo *Random Forest* não ultrapassou a aproximadamente 10%.

Já no segundo experimento (Capítulo 3), com o objetivo de estimar a biomassa aérea de parcelas da floresta amazônica a partir da integração de dados gratuitos, viu-se a necessidade de aplicar de técnicas de aprendizado de máquina. Isso foi motivado pelos pressupostos da regressão linear não serem atendidos e a alta complexidade da floresta, demonstrada pelos dados utilizados para a modelagem.

Os dados SAR de banda C e L aliados a modelos de aprendizado de máquina produziram estimativas precisas da biomassa aérea. Tendo um ganho no desempenho dos modelos com a utilização das imagens multiespectrais e textura SAR.

A integração de todas as informações oriundas de dados de livre acesso (banda C, banda L, multiespectrais e derivações destes), resultaram em RMSE de 48,284 Mg.ha⁻¹ ou 16,297%. Todos os modelos apresentaram desempenho numericamente semelhantes, porém o *Model Tree* (M5) se destacou sobre os demais em todos as variações de conjuntos de dados de entrada, sendo capaz de se aproximar dos valores extremos de biomassa encontrados na área de estudo.

Ao comparar os dois estudos de caso, o desempenho ao modelar as variáveis dendrométricas das plantações florestais apresentou erros menores do que os encontrados para a floresta nativa. Esperava-se exatamente este comportamento, já que os povoamentos florestais apresentam maior homogeneidade dentro de uma mesma unidade amostral, as imagens utilizadas como base para as estimativas apresentam melhor resolução, além da robustez das técnicas utilizadas para a modelagem.

As estimativas de carbono para o mercado de créditos de carbono por exemplo, dependem da qualidade das estimativas de biomassa, necessitando que estas sejam acuradas e precisas para garantir a qualidade das informações. O erro de amostragem para as duas áreas de estudo tanto a partir das métricas obtidas do inventário florestal quando da modelagem, foi menor que 10%, considerado como limite aceitável. Portanto, no que tange os usuários dessas metodologias utilizadas, tanto no setor de florestas plantadas ou florestas nativas, os erros obtidos estão dentro do esperado.

Neste estudo não foram quantificados o total de erros propagados. Porém, deve-se levar em consideração que existem outros erros além dos resultantes dos modelos de regressão linear e das técnicas de aprendizado de máquina apresentados neste estudo. Deve-se levar em consideração que existem ainda as seguintes fontes de erro: erro de amostragem do inventário, erro das equações alométricas (utilizadas para estimar altura, volume e biomassa aérea), erros de geolocalização e erros inerentes do instrumento sensor (ruído *speckle*, erros de calibração, efeitos topográficos e umidade).

Embora o método baseado em campo seja tedioso e demorado, é o método mais preciso, confiável e indispensável inicialmente para estimar as variáveis dendrométricas em ambas as florestas de estudo. Mas o uso de dados obtidos por sensoriamento remoto e sua integração surge como uma forma de mapeamento preliminar de suporte para grandes áreas.

CONCLUSÃO GERAL

Dado todo o processamento de dados e estimativa de variáveis dendrométricas de forma indireta para a plantação florestal e floresta nativa, pode-se concluir que:

- Foi possível obter estimativas dendrométricas de plantações de eucalipto e da floresta tropical de forma indireta com precisão.

- Quanto mais informações foram adicionadas como variáveis independentes dos modelos, melhor foi a sua precisão.

- Foi possível ajustar modelos de regressão linear múltipla para estimar variáveis dendrométricas de plantações de eucalipto, as quais têm o intuito de facilitar essas estimativas por usuários que desconhecem ou não têm acesso a nenhuma métrica obtida na floresta por inventário.

- A utilização de técnicas de aprendizado de máquina apresentou desempenho superior à regressão linear.

- A integração de dados de sensoriamento remoto disponibilizados com livre acesso e modelos de aprendizado de máquina produziram estimativas precisas da biomassa da floresta amazônica.

- A utilização de dados de radar polarimétrico de abertura sintética se mostrou com grande potencial para a estimativa de área basal, volume e biomassa aérea. Porém o desempenho dos modelos de predição apresentou melhor desempenho quando integradas as informações obtidas de imagens multiespectrais.

A principal limitação deste estudo é estimar variáveis que não se tem o seu valor real, pois foram obtidas a partir de estimativa por modelos alométricos. Justifica-se isso pela forma de se obter a biomassa aérea real por exemplo, seria necessários métodos destrutivos da floresta.

Neste estudo, encontrou-se dificuldades em estabelecer a ordem e os processamentos necessários para cada cena nos diferentes níveis de processamento que estão sendo disponibilizadas. Assim como no tempo de processamento e tamanho das imagens PolSar as quais foram baixadas e processadas em um computador físico.

REFERÊNCIAS

- ALBINET, C. et al. **ESA-NASA multi-Mission Analysis Platform for improving global aboveground terrestrial carbon dynamics**. European Geophysical Union 2018. Disponível em: <https://meetingorganizer.copernicus.org/EGU2018/EGU2018-6626.pdf>. Acesso em: 01 fev. 2020.
- ALLEN, R. G.; TASUMI, M.; TREZZA, R. SEBAL (Surface Energy balance Algorithms for land). **Advanced Training and users' manual**. Idaho: Implementation, v. 1.0, 2002.
- ALMEIDA, C. T. et al. Combining LiDAR and hyperspectral data for aboveground biomass modeling in the Brazilian Amazon using different regression algorithms. **Remote Sensing of Environment**, v. 232, 2019.
- ANA – AGÊNCIA NACIONAL DE ÁGUAS. **Séries históricas de estações**. 2016. Disponível em: < http://www.snirh.gov.br/hidroweb/publico/medicoes_historicas_abas.jsf >. Acesso em: 30 de out. de 2019.
- BALDAUF, T.; GARCIA, M. Image Processing of Radar and Lidar in Tropical Forestry. In: PANCEL, L.; KÖHL, M. (Eds.). **Tropical Forestry Handbook**. 1. ed. Berlin: Springer-Verlag Berlin Heidelberg, 2016. p. 635–661.
- BECKER, B. Public R & D policies and private R & D investment: a survey of the empirical evidence. **Journal of Economic Surveys**, v. 29, p. 917–942, 2015.
- BERNINGER, A. et al. SAR-Based estimation of above-ground biomass and its changes in Tropical Forests of Kalimantan using L- and C-band. **Remote Sensing**, v.10, p. 1-22, 2018.
- BINOTI, M. L. M. D. S. et al. Prognose em nível de povoamento de clones de eucalipto empregando redes neurais artificiais. **Cerne**, v. 21, p. 97-105, 2015.
- BINOTI, M. L. M. D. S. et al. Redes neurais artificiais para estimação do volume de Árvores. **Revista Árvore**, v. 38, p. 283-288, 2014.
- BINOTI, M. L. M. D. S. **Redes neurais artificiais para prognose da produção de povoamentos não desbastados de eucalipto**. Dissertação (Mestrado em Ciências Florestais), Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brasil. 2010.
- BOISVENUE, C. et al. Integration of Landsat time series and field plots for forest productivity estimates in decision support models. **Forest Ecology and Management**, v. 376, p. 284-297, 2016.
- BREIMAN, L. E. O. Random forest. **Machine Learning**, v. 45, p. 5-32, 2001.
- CALLON, M. et al. From translations to problematic networks: An introduction to co-word analysis. **Social Science Information**, v. 22, p. 191–235, 1983.
- CAMPBELL, J. B.; WYNNE, R. H. **Introduction to Remote Sensing**. Guilford Press. 2011.

- CAPES. **Portal de periódicos da CAPES**. 2018. Disponível em: http://www-periodicos-capes-gov-br.ez22.periodicos.capes.gov.br/index.php?option%3Dcom_pcollection%26mn%3D70%26s mn%3D79%26cid%3D81%26Itemid%3D%26
- CARREIRAS, J.; MELO, J. B.; VASCONCELOS, M. J. Estimating the above-ground biomass in miombo savanna woodlands (Mozambique, East Africa) using l-band synthetic aperture radar data. **Remote Sensing**, v. 5, p. 1524-1548, 2013.
- CARTUS, O.; SANTORO, M.; KELLNDORFER, J. Mapping forest aboveground biomass in the northeastern united states with Alos Palsar dual polarization l-band. **Remote Sensing of Environment**, p. 466-478, 2012.
- CASSOL, H. L. G. **Aplicação dos dados polarimétrico Alos/Palsar-2 para modelagem de biomassa em florestas secundárias da Amazônia considerando o histórico de uso**. 267 f. Tese (Doutorado em sensoriamento remoto), Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2017.
- CASTILLO, J. A. A. et al. Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 134, p. 70-85, 2017.
- CHAVE J. et al. Error propagation and scaling for tropical forest biomass estimates. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v.359, p. 409-420. 2004.
- CHEN, L. et al. Assessment of multi-wavelength SAR and multispectral instrument data for forest aboveground biomass mapping using random forest kriging. **Forest Ecology and Management**, v. 447, p. 12-25, 2019.
- CHEN, L. et al. Estimation of forest above-ground biomass by geographically weighted regression and machine learning with sentinel imagery. **Forests**, v. 9, p. 582, 2018.
- CLARK, D. B.; KELLNER, J. R. Tropical forest biomass estimation and the fallacy of misplaced concreteness. **Journal of Vegetation Science**, v. 23, p. 1191-1196, 2012.
- CORTE, A. P. D. et al. Fator de expansão de biomassa, razão de raízes-parte aérea e modelos para carbono para *Eucalyptus grandis* plantados no sul do Brasil, **Enciclopédia Biosfera**, v. 11, n. 21, p.1078-1091, 2015.
- CROOM, S. Introduction to research methodology in operations management. In: KARLSSON, C. (Ed.). **Researching Operations Management**. New York: Routledge, 2009.
- CUTLER, M. E. J. et al. Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: An assessment of predictions between regions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 70, p. 66-77, 2012.
- DAI, E. F. et al. Quantifying ecosystem service trade-offs for plantation forest management to benefit provisioning and regulating services. **Ecology and Evolution**, v. 7, p. 7807-7821, 2017.

DEBASTIANI, A. B. et al. Evaluating SAR-optical sensor fusion for aboveground biomass estimation in a Brazilian tropical forest. **Annals of Forest Research**, v. 62, n. 1, p. 109-122, 2019.

DEGRANDI, G.; LUCAS, R. M.; KROPACEK, J. Analysis by wavelet frames of spatial statistics in SAR data for characterizing structural properties of forests. **IEEE Transactions on Geoscience and Remote Sensing**, v. 47, n. 2, p. 494-507, 2009.

DEO, R.; RUSSELL, M.; DOMKE, G. Using Landsat Time-series and LiDAR to Inform Aboveground Forest Biomass Baselines in Northern Minnesota, USA. **Canadian Journal of Remote Sensing**, v. 43, p. 28–47, 2016.

DOBSON, M. C. et al. Dependence of radar backscatter on coniferous forest biomass. **IEEE Transactions on Geoscience and Remote Sensing**, v. 30, n. 2, p. 412-415, 1992.

D'OLIVEIRA, M. V. N. et al. Estimating forest biomass and identifying low-intensity logging areas using airborne scanning lidar in Antimary State Forest, Acre State, Western Brazilian Amazon. **Remote Sensing of Environment**, v. 124, p. 479-491, 2012.

DOMINGUES, G. F. et al. Artificial neural networks on integrated multispectral and SAR data for high-performance prediction of eucalyptus biomass. **Computers and Electronics in Agriculture**, v. 168, p.1-7, 2020.

DUNCANSON, L. et al. Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California. **Remote Sensing of Environment**, v. 242, p. 111779, 2020.

DUDLEY, R. G. A little REDD model to quickly compare possible baseline and policy scenarios for reducing emissions from deforestation and forest degradation. **Mitigation and Adaptation Strategies for Global Change**, v. 15, p. 53-69, 2009.

ENGLHART, V.; KEUCK, V.; SIEGERT, F. Aboveground biomass retrieval in tropical forest – The potential of combined X and L band Sar data use. **Remote Sensing Environment**, v. 115, p. 1260-1271, 2011.

ESA. **Sentinels Scientific Data Hub, European Space Agency**. 2016. Disponível em: <<https://scihub.copernicus.eu/>>. Acesso em: 15 de mai. 2019.

FAO. **Global Forest Resources Assessment 2015**. 2015. Rome. Disponível em: <www.fao.org>. Acesso em: 15 de nov. 2018.

FASSNACHT, F. E. et al. Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. **Remote Sensing of Environment**, v. 154, p. 102-114, 2014.

FORKUOR, G. et al. Above-ground biomass mapping in West African dryland forest using Sentinel-1 and 2 datasets - A case study. **Remote Sensing of Environment**, v. 236, 2020.

GAMA, F. F.; DOS SANTOS, J. R.; MURA, J. C. Continuous monitoring of biophysical Eucalyptus sp parameters using interferometric synthetic aperture radar data in P and X bands. **International Journal of Remote Sensing**, v. 10, p. 1191-1196, 2016.

GAMA, F. F.; SANTOS, J. R.; MURA, J. C. Eucalyptus biomass and volume estimation using interferometric and polarimetric SAR data. **Remote Sensing**, v. 2, n. 4, p. 939-956, 2010.

GARCIA, M. et al. Modelling forest canopy height by integrating airborne LiDAR samples with satellite Radar and multispectral imagery. **International Journal of Applied Earth Observation and Geoinformation**, v. 66, p. 159-173, 2018.

GAVIRIA-MARIN, M.; MERIGÓ, J. M.; BAIER-FLUENTES, H. Knowledge management: A global examination based on bibliometric analysis. **Technological Forecasting and Social Change**, v. 140, p. 194-220, 2019.

GHASEMI, N.; SAHEBI, M. R.; MOHAMMADZADEH, A. A Review on Biomass Estimation Methods Using Synthetic Aperture Radar Data. **International Journal of Geomatics and Geosciences**, v. 1, p. 776-788, 2011.

GIANNETTI, F. et al. A new set of DTM-independent metrics for forest growing stock prediction using UAV photogrammetric data. **Remote Sensing of Environment**, v. 213, p. 195-205, 2018.

GODIN, B. On the origins of bibliometrics. **Scientometrics**, v. 68, p. 109-133, 2006.

GORGENS, E. B. et al. Influência da arquitetura na estimativa de volume de Árvores individuais por meio de redes neurais artificiais. **Revista Árvore**, v. 38, p. 289-295, 2014.

GORGENS, E. B., et al. Estimaco do volume de Árvores utilizando redes neurais artificiais. **Revista Árvore**, v. 33, p. 1141-1147, 2009.

HALL, M. A. **Correlation-based Feature Selection for Machine Learning**. Department of Computer Science University of Waikato, Hamilton, New Zealand, 1999.

HARALICK, R. M. Statistical and structural approaches to texture. **Proceedings of the IEEE**, v. 67, n. 5, p. 786-804, 1979.

HECHENBICHLER, K.; SCHLIEP, K. **Weighted K-Nearest-Neighbor techniques and ordinal classification**. Sonderforschungsbereich 386, 16p. 2004. Disponível em: <https://epub.ub.uni-muenchen.de/1769/1/paper_399.pdf>. Acesso em: 30 de jan. 2020.

HENGL, T. et al. Random Forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. **PeerJ Preprints**, v. 6, p. e5518, 2018.

HENTZ, A. M. K. **Obteno de informaoes dendromtricas para inventrio florestal automatizado por meio de Veculo Areo No Tripulado (VANT)**. 238 f. Tese (Doutorado em Engenharia Florestal) – Setor de Cincias Agrrias, Universidade Federal do Paran, Curitiba, 2018.

HO, Y. S. Top-cited articles in chemical engineering in science citation index expanded: a bibliometric analysis. **Chinese Journal of Chemical Engineering**, v. 20, p. 478-488, 2012.

HOGENSON, K. et al. **Hybrid Pluggable Processing Pipeline (HyP3): A cloud-based infrastructure for generic processing of SAR data**. San Francisco, CA, December. 2016. p. 12-16.

HORNIK, K.; BUCHTA C.; ZEILEIS, A. Open-source machine learning: R meets Weka. **Computational Statistics**, v. 24, n. 2, p. 225-232, 2009.

HOUGHTON, R. A.; HALL, F.; GOETZ, S. J. Importance of biomass in the global carbon cycle. **Journal of Geophysical Research**, v.114, p. G00E03, 2009.

HUETE, A. R. A soil-adjusted vegetation index (SAVI). **Remote Sensing of Environment**, v. 25, p. 295-309, 1988.

HUSSIN, Y. A.; REICH, R. M.; HOFFER, R. M. Estimating slash pine biomass using radar backscatter. **IEEE Transactions on Geoscience and Remote Sensing**, v. 29, n. 3, p. 427-431, 1991.

HYYPPÄ, J. et al. Remote Sensing of Forests from Lidar and Radar. In: THENKABAIL, P. S. (Ed.). **Land Resources Monitoring, Modeling, and Mapping with Remote Sensing**. Boca Raton, USA: CRC Press, 2015. p. 397–428.

IBA. **Indústria Brasileira de Árvores**. Relatório 2019. 79p. 2019. Disponível em: <<https://iba.org/datafiles/publicacoes/relatorios/iba-relatorioanual2019.pdf>>

IBGE – Instituto Brasileiro de Geografia e Estatística. Manual Técnico da Vegetação Brasileira. 2ª ed. revisada e ampliada. 2012. 271p.

INSTITUTO DE TERRAS, CARTOGRAFIA E GEOLOGIA DO PARANÁ (ITCG). **Clima: estado do Paraná**. 2008. Disponível em: http://itcg.pr.gov.br/arquivos/File/Produtos_DGEO/mapas_ITCG/PDF/Mapa_Climas_A3.pdf. Acesso em: 06 de nov, 2019.

INSTITUTO PARANAENSE DE DESENVOLVIMENTO ECONOMICO E SOCIAL (IPARDES). **Leituras regionais: mesorregiões geográficas**, Curitiba: IPARDES, 2004.

ISMAIL, R. et al. Assessing the utility of ALOS PALSAR and SPOT 4 to predict timber volumes in even-aged Eucalyptus plantations located in Zululand, South Africa. **Southern Forests: a Journal of Forest Science**, v. 77, n. 3, 2015.

JAMES, G. et al. **An introduction to statistical learning: with applications in R**. Springer; Edição: 1st ed. 426p. 2013

JAPAN AEROSPACE EXPLORATION AGENCY (JAXA). **ALOS/PALSAR-2**. Tokyo, Japan. 2015. Disponível em: http://global.jaxa.jp/projects/sat/ALOS2/pdf/daichi2_e.pdf. Acesso em: 15 abr. 2017.

- JOSHI, N. et al. Understanding “saturation” of radar signals over forests. **Scientific Reports**, v. 7, p. 3505, 2017.
- KARATZOGLU, A.; SMOLA, A.; HORNIK, K. **Kernlab - An S4 Package for Kernel Methods in R**. 22p. 2004.
- KIRIMI F. et al. Influence of vegetation cover on the oh soil moisture retrieval model: a case study of the Malinda Wetland, Tanzania. **Advances in Remote Sensing**, v. 5, p. 28-42, 2016.
- KÖHL, M.; MAGNUSSEN, S.; MARCHETTI, M. **Sampling Methods, Remote Sensing and GIS Multisource Forest Inventory**. Berlin Heidelberg: Springer, 2006. v. 3.
- KUHN, M. Building predictive models in R using the caret package. **Journal of Statistical Software**, v. 28, n. 5, p. 1-26, 2008.
- KUMAR, A. et al. Tree diversity assessment and above ground forests biomass estimation using SAR remote sensing: A case study of higher altitude vegetation of North-East Himalayas, India. **Physics and Chemistry of the Earth**, v. 111, p. 53-64, 2019.
- KUPLICH, T. M.; CURRAN, P. J.; ATKINSON, P. M. Relating SAR image texture to the biomass of regenerating tropical forests. **International Journal of Remote Sensing**, v. 6, n. 21, 2005.
- LE TOAN, T. et al. The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. **Remote Sensing of Environment**, p. 2850-2860, 2011.
- LE TOAN, T. et al. Relating forest biomass to SAR data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 30, n. 2, p. 403-411, 1992.
- LEE, J. S.; GRUNES, M.R.; DE GRANDI, G. Polarimetric SAR speckle filtering and its implication for Classification. **IEEE Transactions on Geoscience and Remote Sensing**, v. 37, n. 5, p. 2363-2373, 1999.
- LEE, J. S.; POTTIER, E. **Polarimetric radar imaging: from basics to applications**. New York, NY: CRC Press: Taylor & Francis Group, ISBN 978-1-4200- 5497-2, 2009. 440 p.
- LI, J.; HALE, A. Identification of, and knowledge communication among core safety science journals. **Safety Science**, v. 74, p. 70-78, 2015.
- LI, W.; ZHAO, Y. Bibliometric analysis of global environmental assessment research in a 20-year period. **Environmental Impact Assessment Review**, v. 50, p. 158-166, 2015.
- LIAW, A.; WIENER, M. Classification and Regression by randomForest. In: R news. v.2/3, p.18-22, 2002. Disponível em: https://www.r-project.org/doc/Rnews/Rnews_2002-3.pdf. Acesso em: 27 jan. 2020.
- LIU, X. et al. A bibliometric study of earthquake research: 1900–2010. **Scientometrics**, v. 92, p. 747-765, 2012.

LIU, Y. et al. Estimation of the forest stand mean height and aboveground biomass in Northeast China using SAR Sentinel-1B, multispectral Sentinel-2A, and DEM imagery. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 151, p. 277-289, 2019.

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

LU, D. S. The potential and challenge of remote sensing-based biomass estimation. **International Journal of Remote Sensing**, v. 27, n. 7, p. 1297–1328, 2006.

LU, D. et al. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. **International Journal of Digital Earth**, v. 9, p. 63-105, 2014.

LUCAS, R. M. et al. Integration of radar and Landsat-derived foliage projected cover for woody regrowth mapping, Queensland, Australia. **Remote Sensing of Environment**, v. 100, n. 3, p. 388-406, 2006.

LUCAS, R. M.; MOGHADDAM, M.; CRONIN, N. Microwave scattering from mixed-species forests, Queensland, Australia. **IEEE Transactions on Geoscience and Remote Sensing**, v. 42, p. 2142–2159, 2004.

MAHDIANPARI, M.; MOTAGH, M. Random Forest Wetland classification using ALOS-2 L-band, RADARSAT-2 C-band, and TerraSar-X imagery. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 130, p.13-31, 2017.

MAO, G. et al. Research on biomass energy and environment from the past to the future: A bibliometric analysis. **Science of the Total Environment**, v. 635, p. 1081-1090, 2018.

MAUYA, E.W. et al. Effects of field plot size on prediction accuracy of aboveground biomass in airborne laser scanning-assisted inventories in tropical rain forests of Tanzania. **Carbon Balance Management**, v. 10, p. 1-14, 2015.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, v. 5, p. 115-133, 1943.

MCGAUGHEY, R. J. **FUSION / LDV: software para análise e visualização de dados LIDAR**. Seattle, WA, USFS, p.11, 2016.

MERIGÓ, J. M. et al. A bibliometric overview of the journal of business research between 1973 and 2014. **Journal of Business Research**, v. 68, p. 2645-2653, 2015.

MINH, D. H. T. et al. SAR tomography for the retrieval of forest biomass and height: Cross-validation at two tropical forest sites in French Guiana. **Remote Sensing of Environment**, v. 175, p. 138-147, 2016.

MINSKY, M.; PAPERT, S. **Perceptrons**. MIT Press, Cambridge, MA. 1969.

MITCHARD, E. T. A. et al. Mapping tropical forest biomass with radar and spaceborne LIDAR in Lopé National Park, Gabon: Overcoming problems of high biomass and persistent cloud. **Biogeosciences**, v. 9, p. 179–191, 2012.

- MITCHARD, E. T. A. et al. Measuring biomass changes due to woody encroachment and deforestation/degradation in a forest – savanna boundary region of central Africa using multi-temporal L-band radar backscatter. **Remote Sensing of Environment**, v. 115, p. 2861–2873, 2011.
- MOREL, A. C. et al. Estimating aboveground biomass in forest and oil palm plantation in Sabah, Malaysian Borneo using ALOS PALSAR data. **Forest Ecology and Management**, v. 262, p. 1786-1798, 2011.
- NAIDOO, L. et al. L-band Synthetic Aperture Radar imagery performs better than optical datasets at retrieving woody fractional cover in deciduous, dry savannahs. **International Journal of Applied Earth Observation and Geoinformation**, v. 52, p. 54-64, 2016.
- NAVARRO, J. A. et al. Integration of UAV, sentinel-1, and sentinel-2 data for mangrove plantation aboveground biomass monitoring in Senegal. **Remote Sensing**, v. 11, p. 77, 2019.
- NEUENSCHWANDER, A.; PITTS, K. The ATL08 land and vegetation product for the ICESat-2 Mission. **Remote Sensing of Environment**, v. 221, p. 247-259, 2019.
- NGUYEN, L. V. et al. Estimation of tropical forest structural characteristics using ALOS/PALSAR-2 SAR data. **Advances in Remote Sensing**, v. 5, n. 2, p. 131–144, 2016.
- NINGTHOUJAM, R. K.; JOSHI, P. K.; ROY, P. S. Retrieval of forest biomass for tropical deciduous mixed forest using ALOS PALSAR mosaic imagery and field plot data. **International Journal of Applied Earth Observation and Geoinformation**, v. 69, p. 206-216, 2018.
- OON, A.; AZHAR, B. Assessment of ALOS-2 PALSAR-2 L-band and Sentinel-1 C-band SAR backscatter for discriminating between large-scale oil palm plantations and smallholdings on tropical peatlands. **Remote Sensing Applications: Society and Environment**, v. 13, p. 183-190, 2019.
- OZDEMIR, I.; KARNIELI, A. Predicting forest structural parameters using the image texture derived from worldview-2 multispectral imagery in a dryland forest, Israel. **International Journal of Applied Earth Observation and Geoinformation**, v. 13, n. 5, p. 701-710, 2011.
- PÉLLICO NETTO, S.; BRENA, D. A. **Inventário Florestal**. v.1, Curitiba, PR. 316 p. 1997.
- PODEST, E.; SAATCHI, S. Application of multiscale texture in classifying JERS-1 radar data over tropical vegetation. **International Journal of Remote Sensing**, v.2 3, p. 1487-1506, 2002.
- PONZONI, F. J.; SHIMABUKURO, Y. E.; KUPLICH, T. M. **Sensoriamento Remoto da Vegetação**. Editora Oficina de Textos. Ed. 2. 176p. 2012.
- POPE, K. O.; REY-BENAYAS, J. M.; PARIS, J. F. Radar remote sensing of forest and wetland ecosystems in the Central American tropics. **Remote Sensing of Environment**, v. 48, n. 2, p. 205–219, 1994.

PRICE, D. J. S. **Little Science, Big Science**. Columbia University Press, New York. 1963.

PULLIAINEN, J. T.; KURVONEN, L.; HALLIKAINEN, M. T. Multitemporal behavior of L-and C-band SAR observations of boreal forests. **IEEE Transactions on Geoscience and Remote Sensing**, v. 37, p. 927-937, 1999.

QUINLAN, J. R. Learning with continuous classes. In N. Adams, & L. Sterling (Eds.). **Proceedings of the Fifth Australian Joint Conference on Artificial Intelligence**. Hobart, Tasmania. Singapore: World Scientific, 1992, p. 343-348.

QUINLAN, J. R. **C4.5: Programs for Machine Learning. Morgan Kaufmann. In 5th Australian Joint Conference on Artificial Intelligence**. Singapore. World Scientific. p. 343-348, 1993.

RANSON, K. J.; SUN, G. Mapping biomass of a northern forest using multifrequency SAR data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 32, n. 2, 1994.

RIPLEY, B. et al. **Support functions and datasets for Venables and Ripley's MASS**. R Foundation for Statistical Computing, Vienna, Austria. Available online: <http://www.stats.ox.ac.uk/pub/MASS4/> (accessed on 12.12.2018)

RODRIGUES, S. P. et al. Mapping patient safety: a large-scale literature review using bibliometric visualization techniques. **BMJ Open**, v. 4, n. 3, 2014.

ROMANSKI, P.; KOTTHOFF, L. **Package "Fselector"**. 18p. 2018. Disponível em: <<https://cran.r-project.org/web/packages/FSelector/FSelector.pdf>>. Acesso em: 27 de jan. 2020.

ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, n. 6, p. 386-408, 1958.

RUIZ, L. A. et al. Analysis of the Influence of Plot Size and LiDAR Density on Forest Structure Attribute Estimates. **Forests**, v. 5, n. 5, p.936-951, 2014.

SAATCHI, S. et al. Distribution of aboveground live biomass in the Amazon basin. **Global Change Biology**, v. 13, p. 816-837, 2017.

SAATCHI, S. S. et al. Benchmark map of forest carbon stocks in tropical regions across three continents. **Proceedings of the National Academy of Sciences**, v. 108, n. 24, p. 9899-9904, 2011a.

SAATCHI, S. S. et al. Impact of spatial variability of tropical forest structure on radar estimation of aboveground biomass. **Remote Sensing of Environment**, p. 2836-2849, 2011b.

SAMUEL-ROSA, A. et al. **Pedometrics - pedometric tools and techniques**. 2014. R package version 0.1-7. Disponível em: <https://r-forge.r-project.org/R/?group_id=1887>. Acesso em: 10 de nov. 2019.

SANQUETTA, C. R. et al. Comparison of data mining and allometric model in estimation of tree biomass. **BMC Bioinformatics**, v. 16, n. 1, p. 247, 2015.

SANQUETTA, C. R. et al. **Inventários florestais: planejamento e execução**. 3^o edição. 406 p. Curitiba. 2014.

SANQUETTA, C. R. et al. On the use of data mining for estimating carbon storage in the trees. **Carbon Balance Management**, v. 8, n. 6, 2013.

SANTI, E. et al. The potential of multifrequency SAR images for estimating forest biomass in Mediterranean areas. **Remote Sensing of Environment**, v. 200, p. 63-73, 2017.

SANTORO, M.; CARTUS, O. Research Pathways of Forest Above-Ground Biomass Estimation Based on SAR Backscatter and Interferometric SAR Observations. **Remote Sensing**, v. 10, n. 4, p. 2-23, 2018.

SANTOS, J. Savanna and tropical rainforest biomass estimation and spatialization using jers-1 data. **International Journal of Remote Sensing**, p. 1217-1229, 2002.

SANTOS, J. R.; GAMA, F. F.; DA CONCEIÇÃO BISPO, P. Estimating forest biomass by remote sensing radar data in brazil. **Drewno**, v. 57, n. 192, p.120-132, 2014.

SCHLIEP, L.; HECHENBICHLER, K.; LIZEE, A. **Package “kkn”**. 2016. Disponível em: < <https://cran.r-project.org/web/packages/kkn/kkn.pdf> >. Acesso em: 30 de nov. 2019.

SCHREIER, G. **SAR Geocoding: Data and Systems**. Wichmann. 1993.

SENTINEL-2_TEAM. **Sentinel-2 User Handbook**. European Space Agency. 2015.

SHIMADA, M. et al. New global forest/non-forest maps from ALOS PALSAR data (2007–2010). **Remote Sensing of Environment**, v. 155, p. 13-31, 2014.

SHIMADA, M. Orthorectification and slope correction of SAR data using DEM and its associated evaluation. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, n. 3, p. 657-671, 2010.

SILVA, C. A. et al. Combined effect of pulse density and grid cell size on predicting and mapping aboveground carbon in fast-growing Eucalyptus forest plantation using airborne LiDAR data. **Carbon Balance and Management**, v. 12, n. 13, p. 1-16, 2017.

SMALL, H. Visualizing science by citation mapping. **Journal of the American Society for Information Science and Technology**, v. 50, p. 799-813, 1999.

SMALL, D. Flattening gamma: Radiometric terrain correction for SAR imagery. **IEEE Transactions of Geoscience and Remote Sensing**, v. 49, n. 8, p. 3081-3093, 2011.

SOUZA, D. V. **Aprendizado de máquina para predição de biomassa e volume comercial de árvores em florestas tropicais**. 190 f. Tese (Doutorado em Engenharia Florestal) – Setor de Ciências Agrárias, Universidade Federal do Paraná, Curitiba, 2020.

SOUZA, G. S. A. et al. Multi-sensor prediction of Eucalyptus stand volume: A support vector approach. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 156, p.135-146, 2019.

SPANNER, M. A. et al. Remote Sensing temperate coniferous forest leaf area index. The influence of canopy closure, understory and background reflectance. **International Journal of Remote Sensing**, v. 11, n. 1, p. 95-111, 1990.

STICKLER, C. M. et al. The potential ecological costs and cobenefits of REDD: a critical review and case study from the Amazon region. **Global Change Biology**, v. 15, p. 2803-2824, 2009.

TAN, P. N.; STEINBACH, M.; KUMAR, V. **Introduction to Data Mining**. [S.l.]: Editora Ciência Moderna, 2009.

THAPA, R. B. et al. Potential of high-resolution ALOS-PALSAR mosaic texture for aboveground forest carbon tracking in tropical region. **Remote Sensing of Environment**, v. 160, p. 122-133, 2015.

UGOLINI, D. et al. Temporal trend, geographic distribution, and publication quality in asbestos research. **Environmental Science and Pollution Research**, v. 22, p. 6957-6967, 2015.

URBAZAEV, M. et al. Estimation of forest aboveground biomass and uncertainties by integration of field measurements, airborne LiDAR, and SAR and optical satellite data in Mexico. **Carbon Balance and Management**, v. 13, n. 5, p. 1-20, 2018.

VAN ECK, N. J.; WALTMAN, L. Software survey: VOSviewer, a computer program for bibliometric mapping. **Scientometrics**, v. 84, p. 523-538, 2010.

VERRELST, J. et al. Machine learning regression algorithms for biophysical parameter retrieval: opportunities for Sentinel-2 and -3. **Remote Sensing of Environment**, v. 118, p. 127-139, 2012.

VENABLES, W. N.; RIPLEY, B. D. **Modern Applied Statistics with S**. Fourth edition. Springer. 2002.

WANG, B. et al. An overview of climate change vulnerability: a bibliometric analysis based on Web of Science database. **Natural Hazards**, v. 74, p. 1649-1666, 2014.

WARING, H. R.; RUNNING, S. W. **Forest Ecosystems**. Analysis at Multiples Scales. 3rd edition. Academic Press. San Diego, USA. 2007.

WAY, J. et al. The effect of changing environmental conditions on microwave signatures of forest ecosystems: preliminary results of the March 1988 Alaskan aircraft SAR experiment. **International Journal of Remote Sensing**, v. 11, p. 1119-1144, 1990.

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data mining practical machine learning tool and techniques**. 3^oed. Elsevier. 2011. 629 p.

WOODHOUSE, I. H. **Introduction to microwave remote sensing**, Boca Raton: Taylor & Francis Group CRC Press, 2006. 370 p.

YANG, L. et al. Global trends of solid waste research from 1997 to 2011 by using bibliometric analysis. **Scientometrics**, v. 96, p. 133-146, 2013.

ZEIDLER, J.; WEGMANN, M.; DECH, S. **Spatio-temporal robustness of fractional cover upscaling: a case study in semi-arid Savannah's of Namibia and Western Zambia**. Earth Resources and Environmental Remote Sensing/GIS Applications III, p. 85380S, 2012.

ZHOU, Z.; GOH, Y. M.; LI, Q. Overview and analysis of safety management studies in the construction industry. **Safety Science**, v. 72, p. 337-350, 2015.