

RESSALVA

Atendendo solicitação do(a)
autor(a), o texto completo desta
tese será disponibilizado
somente a partir de 08/11/2023.

RORAI PEREIRA MARTINS NETO

**EXTRACTION OF STRUCTURAL VARIABLES USING LIDAR
DATA COMBINED WITH HYPERSPECTRAL IMAGES FOR
CLASSIFICATION OF UPPER CANOPY TREE SPECIES IN
BRAZILIAN ATLANTIC FOREST**



Presidente Prudente
2021



UNIVERSIDADE ESTADUAL PAULISTA
CAMPUS DE PRESIDENTE PRUDENTE
FACULDADE DE CIÊNCIAS E TECNOLOGIA
Programa de Pós-Graduação em Ciências Cartográficas

RORAI PEREIRA MARTINS NETO

**EXTRACTION OF STRUCTURAL VARIABLES USING LIDAR DATA
COMBINED WITH HYPERSPECTRAL IMAGES FOR
CLASSIFICATION OF UPPER CANOPY TREE SPECIES IN
BRAZILIAN ATLANTIC FOREST**

Doctoral thesis presented to the Graduate Program of Cartographic Sciences (PPGCC) at São Paulo State University (UNESP), School of Technology and Sciences, campus Presidente Prudente, SP, Brazil, for the partial fulfilment of the requirements for the grade of Doctor in Cartographic Sciences.

Supervisor: Dr. Antonio M. G. Tommaselli
Co-Supervisors: Dr. Nilton N. Imai
Dr. Hassan C. David
Dr. Milto Miltiadou

Presidente Prudente
2021

Ficha catalográfica elaborada pela Seção Técnica de Aquisição e Tratamento da Informação - Diretoria Técnica de Biblioteca e Documentação - UNESP, Campus de Presidente Prudente

M345e Martins Neto, Rorai Pereira.
Extraction of structural variables using LiDAR data combined with hyperspectral images for classification of upper canopy tree species in Brazilian Atlantic Forest / Rorai Pereira Martins Neto. - 2021
167 f. : il.

Orientador: Antonio Maria Garcia Tommaselli
Coorientadores: Nilton Nobuhiro Imai, Hassan Camil David, Milto Miltiadou
Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Tecnologia, Presidente Prudente, 2021
Inclui bibliografia

1. Remote sensing. 2. Tropical forest. 3. Forest modelling. 4. Machine learning. 5. Tree species identification. I. Tommaselli, Antonio Maria Garcia. II. Imai, Nilton Nobuhiro. III. David, Hassan Camil. IV. Miltiadou, Milto. V. Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Tecnologia, Presidente Prudente. VI. Título.

Alessandra Kuba Oshiro Assunção
CRB-8/9013

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: Extraction of structural variables using LiDAR data combined with hyperspectral images for classification of upper canopy tree species in Brazilian Atlantic Forest

AUTOR: RORAI PEREIRA MARTINS NETO

ORIENTADOR: ANTONIO MARIA GARCIA TOMMASELLI

COORIENTADOR: HASSAN CAMIL DAVID

COORIENTADOR: NILTON NOBUHIRO IMAI

COORIENTADORA: MILTO MILTIADOU

Aprovado como parte das exigências para obtenção do Título de Doutor em CIÊNCIAS CARTOGRÁFICAS, área: Aquisição, Análise e Representação de Informações Espaciais pela Comissão Examinadora:

Prof. Dr. ANTONIO MARIA GARCIA TOMMASELLI (Participação Virtual)
Departamento de Cartografia / Faculdade de Ciências e Tecnologia de Presidente Prudente

Prof. Dr. EDSON APARECIDO MITISHITA (Participação Virtual)
Departamento de Geomática / Universidade Federal do Paraná

Prof. Dr. VERALDO LIESENBERG (Participação Virtual)
Centro de Ciências Agroveterinárias / Universidade do Estado de Santa Catarina

Dra. EIJA HONKAVAARA (Participação Virtual)
Finnish Geospatial Research Institute / Finnish Geospatial Research Institute

Prof. Dr. ALUIR PORFIRIO DAL POZ (Participação Virtual)
Departamento de Cartografia / Faculdade de Ciências e Tecnologia de Presidente Prudente



Presidente Prudente, 08 de novembro de 2021

*To my beloved ones
who always believed in me*

CURRICULUM VITAE

Rorai Pereira Martins Neto

- Birth** 18/01/1991 – Curitiba – Paraná – Brazil
- 1997 – 2004** Elementary School
Colégio Padre João Bagozzi
- 2005 – 2007** High School
Curso e Colégio Acesso
- 2009 – 2013** Bachelor in Forestry
Universidade Federal do Paraná – UFPR
- 2014 – 2015** Master in Forestry
Universidade do Estado de Santa Catarina – CAV/UDESC
Dissertation title: “*Extração de variáveis dendrométricas em árvores de Pinus taeda L. a partir de dados TLS e ALS*”
- 2015 – 2016** Substitute Professor
Universidade Federal de Santa Maria – UFSM/FW
Substitute Professor in Geosciences and Statistics area for undergraduate courses of Forestry and Agronomy
- 2017 – 2021** PhD in Cartographic Sciences
Universidade Estadual Paulista – FCT/UNESP

ACKNOWLEDGMENTS

I would like to sincerely thank the many people who made this work possible.

First, I would like to thank my supervisor, mentor, and role model, Prof Dr. Antonio Tommaselli, who helped me complete this work through his dedication, confidence, experience, exchange of knowledge, help, and friendship. It was a great professional and personal privilege to be your student.

Thank you to my co-supervisors Prof Dr Nilton Imai, Dr. Hassan David, and Dr. Milto Miltiadou. Prof Dr Nilton Imai, for all your constructive discussions, exchange of ideas, and new insights that have enriched my work. Dr. Hassan David from the Brazilian Forest Service, for his longstanding friendship, supervision, advice, and help with forest modelling and statistics. And, to Dr. Milto Miltiadou from the Cyprus University of Technology for all her assistance in understanding and processing full-waveform data in DASOS software, her feedback in writing the thesis papers, and her help in generating ideas for future directions of this work.

Dr. Eija Honkavaara from Finnish Geodetic Institute for her expertise in LiDAR and hyperspectral data and for always helping me to improve submitted papers.

To the company ENGEMAP for using their LiDAR aerial survey in the study area, and to Thiago Tiedtke dos Reis, Survey Engineer, for his willingness to solve any doubts regarding the data. To Baltazar and Valtinho for helping with field surveys and with tree species recognition.

I am grateful to the Graduate Program in Cartographic Sciences professors at PPGCC for their dedication to passing on their knowledge during my PhD, and to the many university employees for doing their utmost to keep the working environment possible despite COVID19 and for their assistance in resolving bureaucratic issues.

My personal thanks, to my family for their unconditional support in my achievements. Dealing with distance and absence was never easy, but we always knew it was for something bigger. Thank you for always believing in my potential and in my dreams.

Many thanks to the great friends that life has given me: André, Cadu, David, Emenize, Gabriella Aragão, Kessye, Luana, Luani, Luís Paulo, Magda Zanon, Nathália, Regina, Thiago Meyer, and Vinícius, who understood my absence from their lives but always received me with open arms during my sporadic appearances. Special thanks to my friends Tiago de Conto and Tainá Mamede for their help with programming in R and forest ecology respectively and for

all the summers we spent together. To Eric Glunt for always being willing to review the English to improve the manuscript. And to the group "Quarentenados" for the companionship, friendship, and exchange of experiences during the pandemic.

To Prof Dr Christel Lingnau from Federal University of Paraná, who, since my first degree, has always believed in me and in my capacity to go further, encouraging me to pursue both a master's and a doctorate degree. Part of who I am today I owe to your teachings and guidance. And to Prof. Veraldo Liesenberg from Santa Catarina State University, for his willingness to always answer my questions when they arose.

Last, but not least, I thank my colleagues, interns, and friends at the PPGCC, especially Marcela Machado, Erika Saito, Guilherme Nascimento, Lucas Dias, Lara Lopes, Mariana Batista, Tayná, Samara, Fabrício Prol, Daniel Arana, Fernando Baixo, Guilherme Santos and Luís Macanoni for their assistance in the scope of cartographic sciences, reception, advice, good times, conversations, and happy hours. In addition, I thank my friends from Presidente Prudente: Marcella Rocha and Letícia, and all those who indirectly was part of my trajectory during my doctorate.

FINANCIAL ACKNOWLEDGMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. This study was also supported by: Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP—Grant: 2013/50426-4) and Academy of Finland (Grant: 273806), which jointly funded the international project Unmanned Airborne Vehicle - based 4D Remote Sensing for Mapping Rain Forest Biodiversity and Its Change in Brazil (UAV_4D_Bio).



All photos were taken by the author. They are from different species of Lapacho (ipês) in the city of Presidente Prudente - SP, Brazil.

“Nature has fixed no limits on our hopes”

Björk

ABSTRACT

The main goal of this thesis was to explore LiDAR data to estimate stand and diversity variables in a remnant of Seasonal Semideciduous Forest, also known as inland Atlantic Forest, one of the most degraded phytophysiognomies of the Atlantic Forest. In addition, the fusion of LiDAR data with hyperspectral images for classification of upper canopy species was also approached. LiDAR data were delivered as peak returns (PR) and full-waveform (FWF) and at two flight heights (900 m and 2000 m), and the hyperspectral images were collected by a light-weight camera. Despite interventions over the years, the study area, the *Ponte Branca* forest remnant, has a good state of conservation, with the presence of pioneer species to climax. PR LiDAR metrics were extracted from the 900 m height flight. Variable selection methods and machine learning techniques were tested to estimate stand and diversity variables. The new variables generated by transforming the metrics with Principal Component Analysis (PCA) and the technique of artificial neural networks (ANN) were the combination that best estimated the variables. The comparison of the attributes extracted by a voxel-based approach, available in the DASOS software, was performed for PR and FWF data. The results with data from two flight heights in the estimation of forest variables were also compared. FWF data showed less biased estimates than PR data. Eight species present in the upper canopy were classified using isolated and combined data: PR, FWF LiDAR, and hyperspectral images. The LiDAR PR metrics transformed by PCAs combined with vegetation indices obtained by hyperspectral images provided the best accuracy in the classification of tree species, with an overall accuracy of 62.8%. LiDAR data, especially FWF LiDAR, presented a great potential for the characterization of tropical forests, especially using the voxel-based approach as implemented in software DASOS. Combining data from PR LiDAR with vegetation indices from hyperspectral images improved the accuracy of classification.

Keywords: tropical forests, forest modelling, forest inventory, diversity, airborne laser scanner, peak returns, full-waveform analysis, software DASOS, machine learning, R programming, hyperspectral remote sensing, tree species identification.

RESUMO

O objetivo principal desta tese foi explorar os dados LiDAR para estimar variáveis de inventário florestal e de diversidade em um remanescente de Floresta Estacional Semidecidual, também conhecida como Mata Atlântica de interior, uma das fitofisionomias mais degradadas da Mata Atlântica. Além disso, a fusão de dados LiDAR com imagens hiperespectrais para classificação de espécies do dossel superior também foi abordada. Os dados LiDAR foram entregues como retornos de pico (PR) e forma de onda completa (FWF) e em duas alturas de voo (900 m e 2000 m), e as imagens hiperespectrais foram coletadas por uma câmera leve de baixo custo. Apesar das intervenções ao longo dos anos a área de estudo, o remanescente florestal Ponte Branca apresenta bom estado de conservação, com presença de espécies pioneiras até clímax. Foram extraídas métricas LiDAR com os PR no voo de 900 m de altura. Métodos de seleção de variáveis e técnicas de aprendizado de máquina foram testados para estimar parâmetros de inventário florestal e de diversidade. As novas variáveis geradas pela transformação das métricas por meio da Análise de Componentes Principais (PCA) e a técnica das redes neurais artificiais (ANN) foi a combinação que melhor estimou os parâmetros florestais. A comparação dos atributos extraídos por uma abordagem baseada em voxel, disponível no *software* DASOS, foi para para dados de PR e FWF. A comparação das estimativas de variáveis florestais com dados obtidos nas duas diferentes alturas de voo também foi realizada. Os dados FWF estimaram as variáveis com menos tendência que os dados de PR. Oito espécies presentes no dossel superior foram classificadas usando os seguintes dados: PR LiDAR, FWF LiDAR e imagens hiperespectrais, usados isoladamente e combinados entre si. As métricas PR LiDAR transformadas pelas PCA combinados com os índices de vegetação obtidos por imagens hiperespectrais, proporcionaram a melhor precisão na classificação das espécies arbóreas, com uma acurácia global de 62,8%. Os dados LiDAR, especialmente os FWF LiDAR, apresentaram um grande potencial para a caracterização de florestas tropicais, especialmente usando a abordagem baseada em voxel implementada no *software* DASOS. A combinação de dados PR LiDAR com os índices de vegetação das imagens hiperespectrais melhorou a precisão da classificação.

Palavras-chave: florestas tropicais, modelagem florestal, inventário florestal, diversidade, varredura laser aerotransportada, retornos de pico, análise da forma de onda completa, *software* DASOS, aprendizado de máquina, programação em R, sensoriamento remoto hiperespectral, identificação de espécies arbóreas.

LIST OF FIGURES

CHAPTER 1

- Figure 1 – Natural occurrence of Brazilian Atlantic Forest. Adapted from SOS Mata Atlântica; Inpe, (2019) 5
- Figure 2 – Remnant of Atlantic Forest in São Paulo state. Adapted from SOS Mata Atlântica; Inpe, (2019) 7

CHAPTER 2

- Figure 1 – Location of the study area in the State of São Paulo and aerial image (RGB) of *Ponte Branca* forest remnant where T1 to T8 represent the transects and P1 to P7 the plots. Adapted from Berveglieri et al., (2018). 18
- Figure 2 – Species accumulation curve with a boxplot of each plot. The dotted lines represent the confidence interval of 95%. 20
- Figure 3 – Diameter distribution for all trees and for species with higher IVIs. 24
- Figure 4 – (a) Canopy height model of *Ponte Branca* forest remnant. (b) Frequency distribution of pixels referring to tree heights. 25

CHAPTER 3

- Figure 1 – Location of the study area and the canopy height model representing the tree heights inside the *Ponte Branca* forest remnant. 36
- Figure 2 – Airborne laser scanner (ALS) point cloud representing the vertical structure of a plot of the *Ponte Branca* forest remnant. (a) Three-dimensional view of the plot. (b) Cross section of the same plot. 39
- Figure 3 – Correlogram between the 54 extracted LiDAR metrics. 47
- Figure 4 – (a) First five principal components (PCs) and the percentage of variance explained by each one. (b) Projection of Table 5. PCs and the LiDAR metrics. 49
- Figure 5 – (a) RMSE in percentage for the regression methods tested for the seven estimated forest variables; (b) Bias in the percentage of each modelled variable for each regression method tested. 51
- Figure 6 – Relative importance of PCs for each forest variable modelled by the best-selected regression method. 54

CHAPTER 4

- Figure 1 – Representation of a modelled waveform. The light blue dots are the LiDAR pulses, and the dark blue line is the Gaussian fitted model. Adapted from Ducic et al. (2006). 78

Figure 2 – LiDAR pulses emitted at different scanning angles intercepting the voxelized space. Wang and Glennie, (2015).....	80
Figure 3 – Location of the study area. (a) Map of Brazil with emphasis on the state of São Paulo. (b) West of the state of São Paulo. (c) <i>Ponte Branca</i> forest remnant with the location of the surveyed plots in the different successional stages found. Adapted from Berveglieri et al. (2018); Martins-Neto et al. (2021).	81
Figure 4 – Representation of LiDAR datatypes. The blue dots are the discrete peak returns peaks, and the solid black line is the FWF data that is digitized at intervals of 1 ns as denoted by the red dotted lines.....	84
Figure 5 – Flowchart of LiDAR data processing to extract the forest variables.	85
Figure 6 – Point clouds at different heights of flights and division of the area into voxels of equal size and voxels with different intensities.	87
Figure 7 – (a) Point cloud representation of a plot with an established center and a radius that contains the plot area. (b) Plot after the voxelization process and feature vector extraction from a cuboid three-dimensional window, with the movement of the top window to the first non-empty voxel of the middle column. Adapted from Miltiadou et al. (2020).	89
Figure 8 – Pixel value computed from the thickness information in a column of voxels. Adapted from Miltiadou et al., 2019.	92
Figure 9 – Maps containing the distance between the first and the last non-empty voxel for PR and FWF data with different noise filtering thresholds at flight heights of 900 m and 2000 m. Details of waveform samples at different thresholds and flight heights are also shown.....	93
Figure 10 – Percentage of importance of the attributes extracted by DASOS in the best scenarios for the estimation of the four forest variables.....	97

CHAPTER 5

Figure 1 – Canopy height models of <i>Ponte Branca</i> forest remnant. (a) All tree heights. (b) Lower stratum. (c) Middle stratum. (d) Upper stratum.	118
Figure 2 – Targets located near the overflown area. In red the radiometric targets, and in blue the GCPs targets.	122
Figure 3 – PR LiDAR data processing flowchart.....	123
Figure 4 – FWF LiDAR data processing flowchart.	124
Figure 5 – Hyperspectral images flowchart. Adapted from Moriya et al. (2017) and Näsi et al. (2015).	125

Figure 6 – Top view of a voxelised tree, with the ITC (in green), the centroid calculated from the ITC (in blue) and the 3D window with 5 voxels in X and Y.....	129
Figure 7 – Individual tree crowns samples (in green) for each species identified in the field and delineated in hyperspectral orthomosaic (R: 780.49 nm; G: 650.96 nm; B: 535.09 nm). The yellow dots are the centroids calculated from the ITCs.	130
Figure 8 – (a) Percentage of variance explained by each PC. (b) Projection of PR LiDAR metrics in each PCs and their respective contribution.	133
Figure 9 – Mean spectra of each tree species with standard error bars.	134
Figure 10 – Importance of variables for classification of tree species in scenario "d".	137
Figure B1 – Spectra obtained from each tree for each species, from hyperspectral orthomosaics.	150

LIST OF TABLES

CHAPTER 2

Table 1 – Phytosociological parameters of the tree species sampled in the <i>Ponte Branca</i> forest remnant.	21
--	----

CHAPTER 3

Table 1 – Statistics about forest variables calculated from field data.	37
Table 2 – Light detection and ranging (LiDAR) metrics extracted from normalized point clouds.	42
Table 3 – Summary of the architectures adopted for modelling using artificial neural networks (ANNs).	44
Table 4 – Summary of the parameters adopted for modelling using epsilon regression type-support vector machine (ϵ -SVM).	46
Table 5 – Selection of the best model for each estimated forest variable.	50
Table A1 – Statistics of forest variables calculated from each of 15 surveyed plots.	70
Table A2 – The best multiple linear regression models fitted for the 7 forest variables using the 5 PCs and 15 uncorrelated LiDAR metrics as input data.	71

CHAPTER 4

Table 1 – Descriptive statistics of measured and derived forest variables.	82
Table 2 – Technical specifications of Riegl LMS-Q680i equipment (RIEGL, 2012).	84
Table 3 – Attributes extracted in DASOS.	90
Table 4 – Accuracy assessment of forest variables estimated with PR and FWF data from different flight heights. The underlined values are those with the smallest bias for each variable estimated in each of the tested scenarios.	96

CHAPTER 5

Table 1 – Characteristics of selected species in the field.	119
Table 2 – Wavelengths (λ) used in each of the Rikola camera bands and their respective FWHM.	121
Table 3 – Summary of flight campaigns for images acquisition.	122
Table 4 – Vegetation indices (VIs) calculated from hyperspectral orthomosaics.	130
Table 5 – Different scenarios tested with the three different datasets with the number of features used in each test.	131
Table 6 – Classification results for each tested scenario. The best results are highlighted.	136
Table 7 – Confusion matrix for the classification of the 8 tree species in the best scenario.	136

LIST OF SYMBOLS AND ABBREVIATIONS

AD	Absolute density
AGB	Aboveground biomass
AIC	Akaike information criterion
ALS	Airborne laser scanner
ANN	Artificial neural network
AnPe	<i>Anadenanthera peregrina</i>
APG	Angiosperm phylogeny group
ApLe	<i>Apuleia leiocarpa</i>
AsPo	<i>Aspidosperma polyneuron</i>
BA	Basal area
BBA	Bundle block adjustment
BRDF	Bidirectional reflectance distribution function
CHM	Canopy height model
cm	Centimeter
CoLa	<i>Copaifera langsdorffii</i>
D	Simpson diversity index
DBH	Diameter at breast height
DEN	Density
DN	Digital numbers
DR	Discrete return
DSM	Digital surface model
DTM	Digital terrain model
EOPs	Exterior orientation parameters
EVA	Ethylene vinyl acetate
FWF	Full-waveform
FWHM	Full width at half maximum
GCP	Ground control points
GNSS	Global navigation satellite system
GSD	Ground sample distance
H	Height
H'	Shannon-Weaver index
ha	Hectare
HeAp	<i>Helietta apiculata</i>
Hm	Average height
HyCo	<i>Hymenaea courbaril</i>
IOPs	Interior orientation parameters
ITC	Individual tree crowns
IVI	Importance value index
J	Pielou equability evenness index
km	Kilometers
LiDAR	Light detection and ranging
LOOCV	Leave-one-out cross-validation
m	Meters
MCARI	Modified Chlorophyll Absorption in Reflectance Index
MDBH	Mean diameter at breast height
Mg	Megagram
ms	Miliseconds

N	Number of trees
ND	Normalized Difference 682/553
NDVI	Normalized Difference Vegetation Index
NDVI _h	Normalized Difference 780/550 Green NDVI hyper
nl	Noise level
nm	Nanometers
ns	Nanoseconds
NTS	Number of tree species
OA	Overall accuracy
OLS	Ordinary least-squares multiple regression
PA	Producer's accuracy
PC	Principal component
PCA	Principal component analysis
PR	Peak returns
PRI	Photochemical Reflectance Index
PSRI	Plant Senescence Reflectance Index
PSSR	Pigment Specific Simple Ratio
PtPu	<i>Pterodon pubescens</i>
QMD	Quadratic mean diameter
R ²	Coefficient of determination
RD	Relative density
RDo	Relative dominance
RE	Red edge
RENDVI	Red Edge Normalized Difference Vegetation Index
REP	Red edge position
RF	Relative frequency
RF	Random forest
RMSE	Root-mean-square error
s	Seconds
SIPI	Structure Insensitive Pigment reflectance index
SVM	Support vector machine
SWIR	Short-wave infrared
SyRo	<i>Syagrus romanzoffiana</i>
TB&C	Tropical Biomass & Carbon
TIN	Triangular irregular network
UA	User's accuracy
UAV	Unmanned aerial vehicle
VI	Vegetation index
VNIR	Visible-near infrared
Voxel	Volume element
ε-SVM	Epsilon support vector machine
κ	Cohen's Kappa index
λ	Wavelengths
μm	Micrometers

CONTENTS

CHAPTER 1 – INTRODUCTION

1 OVERVIEW AND MOTIVATION	5
2 HYPOTHESIS AND OBJECTIVES	10
3 CONTENTS AND CONTRIBUTIONS	11
4 INTERNATIONAL COOPERATION	12
5 REFERENCES	13
CHAPTER 2 – STRUCTURE AND TREE DIVERSITY OF AN INLAND ATLANTIC FOREST – CASE STUDY OF <i>PONTE BRANCA</i> FOREST REMNANT, BRAZIL	
1 SCOPE	17
2 MATERIALS AND METHODS	17
2.1 STUDY AREA.....	17
2.2 FOREST INVENTORY	18
2.3 VEGETATION ANALYSIS.....	19
3 RESULTS AND DISCUSSION	20
4 CONCLUSIONS	29
5 REFERENCES	30
CHAPTER 3 – IDENTIFICATION OF SIGNIFICATIVE LIDAR METRICS AND COMPARISON OF MACHINE LEARNING APPROACHES FOR ESTIMATING STAND AND DIVERSITY VARIABLES IN HETEROGENEOUS BRAZILIAN ATLANTIC FOREST	
1 INTRODUCTION	32
2 MATERIALS AND METHODS	35
2.1 FIELD SURVEY	35
2.2 LIDAR DATA COLLECTION	38
2.3 LIDAR DATA PROCESSING	38
2.4 INPUT DATA SELECTION.....	41
2.5 REGRESSION TECHNIQUES SETTINGS	43
2.6 EVALUATION AND PERFORMANCE OF TESTED MODELS.....	46
3 RESULTS	47
3.1 CORRELATION ANALYSIS AND PCAs.....	47
3.2 MODEL PERFORMANCE AND EVALUATION	50
3.3 IMPORTANCE OF INPUT METRICS	53
4 DISCUSSION	55
5 CONCLUSIONS	60
6 REFERENCES	61
APPENDIX A – CHAPTER 3	69
CHAPTER 4 – EXTRACTING TROPICAL FOREST VARIABLES BY VOXEL APPROACH – HOW DO LIDAR DATA TYPE AND FLIGHT HEIGHT AFFECT ESTIMATES?	
1 INTRODUCTION	73
2 BACKGROUND ON FULL-WAVEFORM LIDAR DATA	77
3 MATERIALS AND METHODS	80

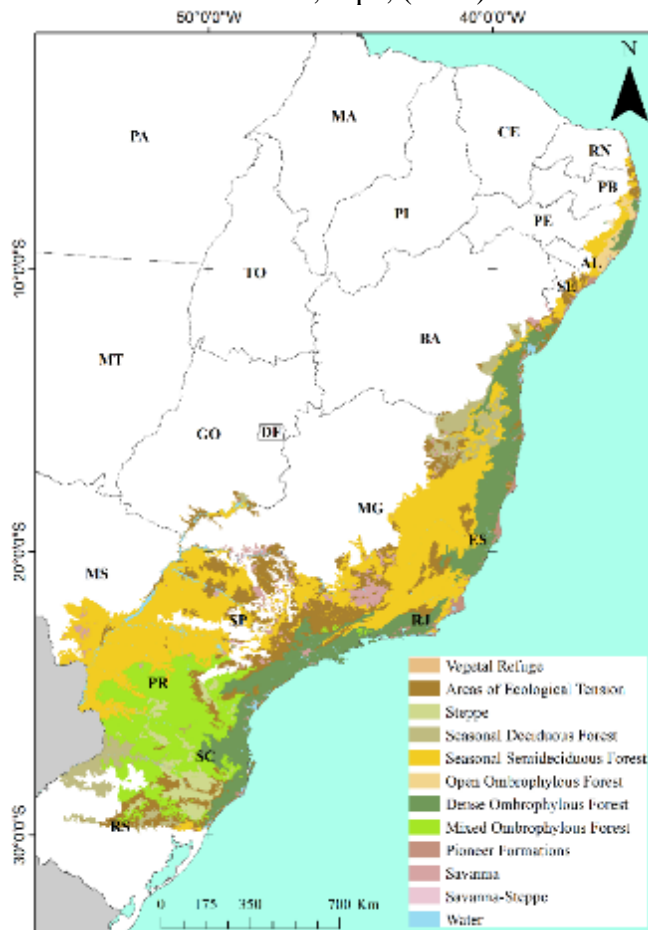
3.1 STUDY AREA.....	80
3.2 FIELD PLOT AND VARIABLES ACQUISITION.....	82
3.3 ABOVEGROUND BIOMASS ESTIMATION	82
3.4 LIDAR DATA SURVEY	83
3.5 LIDAR DATA PROCESSING	85
3.5.1 Voxelization	86
3.5.2 Extraction of feature vectors.....	88
3.6 MODELING FOREST VARIABLES.....	90
3.7 MODEL GOODNESS-OF-FIT	91
4 RESULTS.....	92
4.1 SELECTING NOISE LEVEL THRESHOLD.....	92
4.2 FOREST VARIABLES ESTIMATION.....	95
5 DISCUSSION	99
6 CONCLUSION.....	105
7 REFERENCES	107
CHAPTER 5 – CLASSIFICATION OF TREE SPECIES IN THE UPPER CANOPY OF ATLANTIC FOREST INTEGRATING HYPERSPECTRAL IMAGES AND LIDAR DATA	
1 INTRODUCTION.....	116
2 DATASETS ACQUISITION.....	117
2.1 STUDY SITE AND SAMPLES SELECTION.....	117
2.2 REMOTE SENSING DATA.....	120
2.2.1 LiDAR data.....	120
2.2.2 Hyperspectral imagery data	120
3 METHODOLOGY	122
3.1 LIDAR DATA PROCESSING	122
3.2 HYPERSPECTRAL IMAGES PROCESSING.....	124
3.3 FEATURES EXTRACTION	127
3.4 AUTOMATIC CLASSIFICATION.....	131
4 RESULTS.....	132
4.1 PRINCIPAL COMPONENT ANALYSIS AND SPECTRAL FEATURES	132
4.2 TREES SPECIES CLASSIFICATION	135
5 DISCUSSION	137
6 CONCLUSIONS.....	140
7 REFERENCES	142
APPENDIX B – CHAPTER 5	149
CHAPTER 6 – FINAL REMARKS.....	151

CHAPTER 1 – INTRODUCTION

1 OVERVIEW AND MOTIVATION

The Brazilian Atlantic Forest domain is the second largest tropical forest in America, with high biodiversity rates and a unique biota with a very rich endemism, which are second only to those of the Amazon forest (CARVALHO JÚNIOR et al., 2008; FIASCHI; PIRANI, 2009). The Atlantic domain occurs along the Brazilian coast (Figure 1), extending far inland in some areas of south and south-eastern Brazil (FIASCHI; PIRANI, 2009). It embraces a huge variety of formations and forest ecosystems with very different structures and floristic compositions, conditioned by the region's edaphoclimatic and topographic characteristics (SOS MATA ATLÂNTICA; INPE, 2019). More than 70% of the Brazilian population lives in the Atlantic forest domain; as a result, it is the most threatened biome in Brazil (SCARANO; CEOTTO, 2015).

Figure 1 – Natural occurrence of Brazilian Atlantic Forest. Adapted from SOS Mata Atlântica; Inpe, (2019)

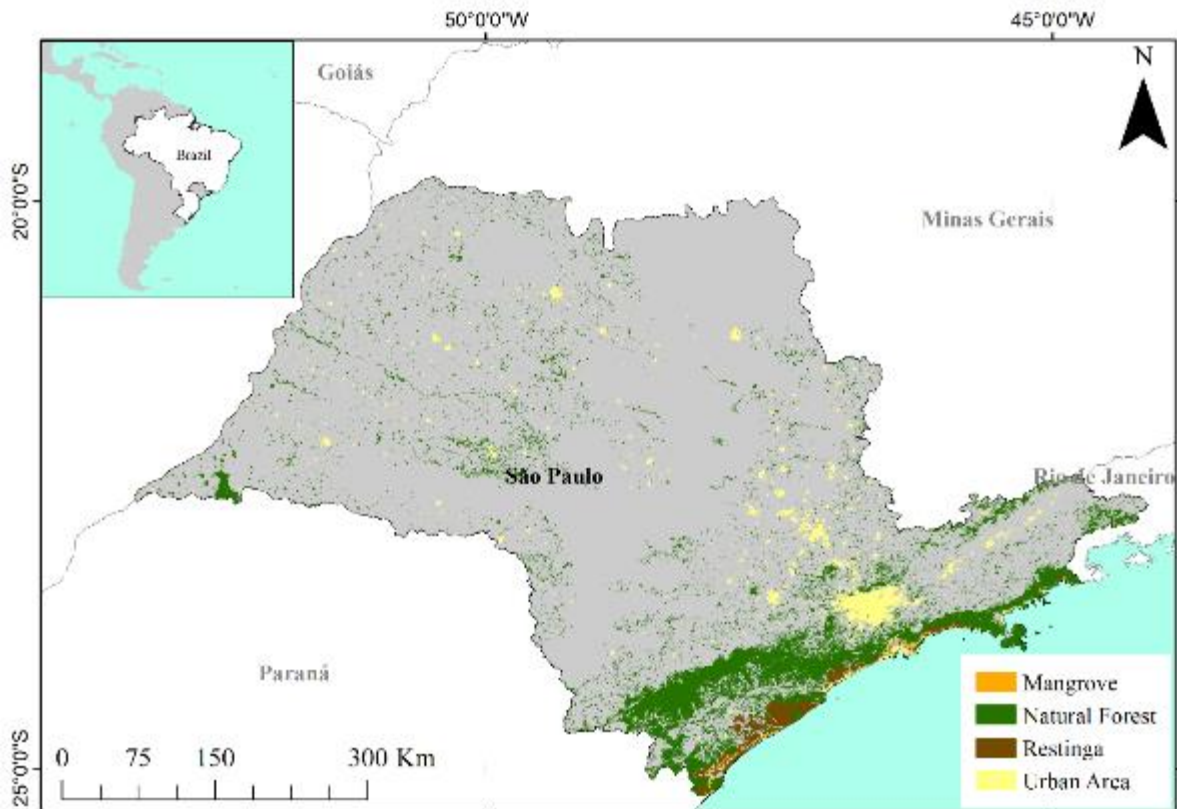


Logging, disordered urban growth, agricultural encroachment and industrialization have contributed to the deforestation and fragmentation of the Atlantic forest, reducing the original coverage to only 11.6% of the original natural environment. The remaining forest is distributed in small areas (DEAN, 1996; HADDAD et al., 2015; HARGREAVES, 2008; RIBEIRO et al., 2009) in the secondary successional stage that have been modified from their original vegetation cover (FONSECA; RIBEIRO; CARVALHO, 2013). There is biological evidence that mature forests with climax species are being replaced and dominated by pioneer species, in a process called “secondarization” (JOLY; METZGER; TABARELLI, 2014; LEAO et al., 2014; SCARANO; CEOTTO, 2015).

The current forest remnants are insufficient to maintain biodiversity, which has led to borderline situations, such as isolation of fauna and flora populations, genetic impoverishment, and growing edge effects (GUERRA et al., 2013; SCARANO; CEOTTO, 2015). Thus, the Atlantic forest is considered one of the world’s priorities for biodiversity conservation, with approximately 14000 vascular plant species distributed in 208 families, of which approximately 8000 are classified as endemic (WERNECK et al., 2011). Currently, an average of 170 unpublished species are identified each year (FIASCHI; PIRANI, 2009; MYERS et al., 2000). Therefore, this biome is a biodiversity hotspot because it has already lost more than 75% of its original coverage and still shows a high degree of endemism (MYERS et al., 2000; WILLIAMS et al., 2011).

Currently, the state of São Paulo, Brazil, has remnants of approximately 16,3% of the original area of the Atlantic Forest in the state (Figure 2). These remnants are located mainly along the coast and in the nearby areas, which are represented by dense ombrophilous forest and mixed ombrophilous forest, respectively (SOS MATA ATLÂNTICA; INPE, 2019). In the western region of the state, called *Pontal do Paranapanema*, there are a few remnants of the seasonal semideciduous forest: those in the *Morro do Diabo* state park and the *Mico-Leão-Preto* (Black Lion Tamarin - *Leontopithecus chrysopygus*) ecological station. Even though these remnants are small, they are recognized for their importance in the strategies for conservation of the Atlantic forest due to their extreme biological importance for endangered specimens of flora and fauna (MMA; IBAMA; ICMBIO, 2007).

Figure 2 – Remnant of Atlantic Forest in São Paulo state. Adapted from SOS Mata Atlântica; Inpe, (2019)



The seasonal semideciduous forest, which is also known as the inland Atlantic Forest, is influenced by dual climatic seasonality. In tropical regions, the dry and rainfall periods are well defined and have an average annual temperature of 21°C, but in the subtropical region, there is a short period of drought with a temperature that decreases to below 15°C in winter (DITT, 2002). Due to the decrease in precipitation and humidity in winter, the presence of epiphytes is not very noticeable, but these weather conditions significantly increase the number of lianas (RODERJAN et al., 2002). These climatic features affect the upper canopy layer by inducing dormancy; 20 to 50% of the trees in the forest are deciduous and lose their leaves in seasons with unfavorable conditions (CAMPANILI; SCHÄFFER, 2010).

For an understanding of these native forest remnants, structural analysis of these fragments is necessary. It includes various approaches such as floristic composition (estimates of similarity, richness, and diversity of species), determination of parameters related to horizontal and vertical structures of vegetation, and additional information related to the size and health of tree trunks, infestation of vines and identification of rare species (DE SOUZA; SOARES; SOARES, 2013).

Information about the structure of forests is obtained through forest inventories. This information is essential to infer about the development, degradation, and successional stage of a given forest fragment (VAUGHN; MOSKAL; TURNBLOM, 2012). However, field inventories are costly, expensive, and restricted to small areas or sampling techniques. Furthermore, in tropical forests, these activities are hampered by the high understory density, presence of lianas and vines, and aerial roots in these environments, resulting in difficulties for walking through the forest. Besides, measuring some tree variables like canopy area and tree heights is troublesome because of the variation in tree height and overlapping of tree crowns (MARTINS-NETO et al., 2021).

The scarcity of efficient methods for determining the forest structure limits the relationship between the forest structure and its functional characteristics at compatible spatial scales (LEFSKY et al., 2002; PARKER; RUSS, 2004; POPESCU; WYNNE; NELSON, 2002). Remote Sensing data offer the possibility of providing information about forests at different spatial and temporal scales with high geometric precision and are very useful for extracting information about forest cover, forest biomass, and forest structure (SHEN et al., 2018; TAUBERT et al., 2021).

Among the available remote sensing technologies, LiDAR (Light detection and ranging) data is a powerful tool for monitoring forest environments. The ability of the pulses to penetrate through small openings in the vegetation canopy allows obtaining three-dimensional information on the forest structure and terrain (LIM et al., 2003; SHAN; TOTH, 2018), which are used to estimate important biophysical parameters of forests (HYYPÄÄ et al., 2008; VAUHKONEN et al., 2012).

LiDAR systems can be classified according to their return type as discrete return (DR) and full waveform (FWF). Discrete return point clouds have been widely used in forestry studies, in which metrics are extracted representing the vertical structural, such as descriptive statistics of heights, intensity reflected by the object, and the number of returns for each emitted pulse. With these metrics it is possible to extract parameters of the vegetation, such as diameter at breast height, aboveground biomass, tree density and volume, and also for the classification of tree species (FALKOWSKI et al., 2009; MALTAMO; NÆSSET; VAUHKONEN, 2014; SHI et al., 2018). However, discrete return systems record information of the number of returns for each emitted pulse, limiting detection to surfaces sufficiently separated in space (CAO et al., 2016). With the technology advancement, FWF systems were developed to record and digitize the amount of energy returned to the sensor, after being backscattered by objects present

on the surface. The returned signal is sampled at equal time intervals (LIM et al., 2003; MALLET; BRETAR, 2009). More information is recorded in FWF than in DR systems, and the resulting waveform contains the properties of all elements that intersect the emitted beam path. This allows a better interpretation of the physical structure and geometric backscatter properties of the objects intercepted, presenting new opportunities for forestry studies that can improve the characterization of the forest structure (MALLET; BRETAR, 2009; PIROTTI, 2011; REITBERGER; KRZYSZEK; STILLA, 2006; SHAN; TOTH, 2018).

Another remote sensing technology that has shown potential for studies in forests is hyperspectral imaging. Hyperspectral sensors offer large amounts of continuous-narrow spectral bands, which give detailed spectral signatures of trees directly related to the chemical, structural and physiological properties of foliage and canopies. As tropical forests are very complex, hyperspectral sensors provide valuable information, mainly for the identification of species and determination of forest succession (FERREIRA et al., 2016; LIESENBERG; BOEHM; GLOAGUEN, 2009; MIYOSHI et al., 2020; SANCHEZ-AZOFEIFA et al., 2017; SOTHE et al., 2019).

The recent availability of sensors at orbital level, such as the full wave form LiDAR sensor GEDI (Global Ecosystem Dynamics Investigation), and hyperspectral imaging sensors like EnMAP (Environmental Mapping and Analysis Program), PRISMA (Hyperspectral Precursor of the Application Mission) and HISUI (Hyperspectral Imager Suite), allow the knowledge of the vegetation structure at regional and continental scales. However, in fragmented landscapes, such as the Brazilian Atlantic Forest, where a large part of the remaining forests is small, it is necessary to use sensors with higher spatial resolutions that better capture the structure of these forest remnants. Thus, with data from sensors such as LiDAR and RGB cameras, multi and hyperspectral on-board on aircraft and unmanned aerial vehicles (UAV), it is possible to obtain information about the structure of these small forest remnants at scales compatible with their size.

In Brazil, the use of high-resolution remote sensing data such as LiDAR point cloud and UAV images are mainly focused on extracting information from homogeneous forest plantations. In the case of native forests, studies are mainly focused on the Amazon, due to the availability of LiDAR data provided by the Sustainable Landscapes Brazil program (<https://www.paisagenslidar.cnptia.embrapa.br/webgis/>) and the Amazon Fund (<http://www.fundoamazonia.gov.br/en/home/>). Thus, there is a lack of data that can describe the vegetation in other Brazilian forest typologies, mainly the more fragmented ones.

As mentioned before, the Atlantic Forest is the most degraded biome in Brazil, and few studies describe the structure and composition of the Atlantic Forest remnants in the interior of the state of São Paulo, since most of these conservation units are located on the coast (Figure 2). Thus, there is a potential use of LiDAR data and hyperspectral images for extracting variables to understand the composition and structure of forest environments. These data will be used to obtain information from fragmented areas, e.g., the remnants of the Brazilian Atlantic Forest, and will serve for a better knowledge of these degraded fragments and best practices of sustainable management and conservation purposes.

2 HYPOTHESIS AND OBJECTIVES

There are two hypothesis on this doctoral thesis: (i) the use of peak return and full waveform airborne LiDAR allows the estimation of variables that describe the structure and diversity of a small and heterogeneous Atlantic forest remnant, that has been disturbed over time; and (ii) the combination of LiDAR data with hyperspectral images improves the accuracy of tree species classification in complex forests. Based on these hypothesis, the following objectives were established:

- From field data, characterize the forest remnant in terms of structure, floristic composition, and diversity.
- Compare four machine learning approaches with different numbers of peak return LiDAR metrics as input data to estimate seven stand and diversity variables.
- Analyze the full waveform and verify if it is worth using it compared to the peak return data to estimate more complex variables.
- Check if different flight heights and different return types affect the estimations of forests variables.
- Test the automatic classification of tree species present in the upper canopy using LiDAR data and hyperspectral images isolated and combined to check improvements in classification accuracy.

3 CONTENTS AND CONTRIBUTIONS

This thesis is structured into six chapters, and each one has a specific objective to meet the main project issues. In **Chapter 1**, an introduction about the Atlantic Forest, its conservation status, and the need for further studies was shown. The potential of remote sensing to describe the structure of these forest remnants is discussed.

Chapter 2 describes the study area and the field forest inventories. The objective was to obtain the vertical and horizontal structure, floristic composition, and diversity of the studied forest remnant. The data obtained in this chapter served as parameters for comparison with the estimations using LiDAR data and classification using LiDAR data and hyperspectral images.

LiDAR data were explored as peak returns in **Chapter 3**. Several metrics related to the distribution of heights, intensity, and pulse returns were extracted. Input data selection methods were tested to choose the best input dataset to estimate stand and diversity variables of the studied remnant. In addition, several machine learning modeling techniques were tested to understand how they deal with the problem of low sample sufficiency and a large number of predictor variables.

Some variables were harder to estimate due to the difficulty of finding relationships with LiDAR metrics, especially in tropical forests which present a more complex structure. Thus, in **Chapter 4**, the full waveform LiDAR data were explored using a voxel-based processing technique. The objective was to compare whether full-waveform analysis improves variable estimations in tropical forests. The influence of flight height on the estimations was also analyzed and how the voxel-based analysis behaves with peak returns and full-waveform LiDAR data.

In **Chapter 5**, the classification of upper canopy species was performed, testing different datasets: hyperspectral images obtained with a lightweight camera, peak returns LiDAR data, and full-waveform LiDAR data. These datasets were used separately and combined to understand how the spectral information from the hyperspectral images with the geometric information of the LiDAR data influence the accuracy of tree species classification. Finally, in **Chapter 6**, a summary of thesis results, conclusions and future recommendations are presented.

There is an extensive literature about the application of LiDAR remote sensing in boreal and temperate forests, as well the use of hyperspectral sensors for species classification. Studies in Brazilian forests related to the themes of extracting inventory and diversity variables using

LiDAR are scarce for other biomes. Regarding tree species classification, studies are even scarcer, due to the great diversity of species that Brazilian forests have. Thus, in the thesis chapters, we always seek to base our results and discussions in tropical forests, Brazil, or elsewhere, whenever possible.

With this study we look forward to show the potential of using high-resolution remote sensing data from different sources for the characterization and extraction of structural information from small remnants of Brazilian native forests. Due to their size, these forests are often neglected when using sensors with medium and low spatial resolution.

Several well-established techniques in the literature were used to process peak return LiDAR data, and some were adapted to select the best set of attributes for estimating forest parameters. A new processing technique is presented using a voxel-based approach for analyzing the full waveform. We expect that FWF data will present new possibilities for the characterization of tropical forests.

Finally, we expect the combination of geometric data from LiDAR with spectral information from hyperspectral images will improve species classification in high complexity forests. The results of this thesis will serve for conservation decision-making and management plans. This data integration and extraction of the best set of attributes to estimate forest structural variables and species classification brings innovation and new processing perspectives for the structural understanding of small fragments of native forests.

4 INTERNATIONAL COOPERATION

This PhD thesis is a result of an international project entitled “Unmanned Airborne Vehicle-based 4D Remote Sensing for Mapping Rain Forest Biodiversity and its Change in Brazil” (UAV4D-BIO), jointly funded by São Paulo Research Foundation, Brazil (FAPESP – Grant: 2013/50426-4) and the Academy of Finland (AKA - Grant: 273806). The objective of UAV-4D-BIO was to develop technologies for biodiversity and its change mapping. The results obtained here served to understand better the studied forest remnant and the presentation of methodologies that can be tested in other forest remnants and forest typologies.

5 REFERENCES

CAMPANILI, M.; SCHÄFFER, W. B. **Mata Atlântica: manual de adequação ambiental**. Brasília: MMA/SBF, 2010.

CAO, L.; GAO, S.; LI, P.; YUN, T.; SHEN, X.; RUAN, H. Aboveground biomass estimation of individual trees in a coastal planted forest using full-waveform airborne laser scanning data. **Remote Sensing**, [s. l.], v. 8, n. 9, p. 729, 2016.

CARVALHO JÚNIOR, O. A. De; COELHO, M. A. N.; MARTINS, É. de S.; GOMES, R. A. T.; COUTO JUNIOR, A. F.; OLIVEIRA, S. N. De; SANTANA, O. A. Mapeamento da vegetação na floresta atlântica usando o classificador de árvore de decisão para integrar dados de sensoriamento remoto e modelo digital de terreno. **Revista Brasileira de Geofísica**, [s. l.], v. 26, n. 3, p. 331–345, 2008.

DE SOUZA, A. L.; SOARES, C. P. B.; SOARES, C. P. B. **Florestas nativas: estrutura, dinâmica e manejo**. [s.l: s.n.].

DEAN, W. A ferro e fogo: a história e a devastação da Mata Atlântica brasileira. In: **A ferro e fogo: a história e a devastação da Mata Atlântica brasileira**. [s.l: s.n.]. p. 484–484.

DITT, E. H. **Fragmentos florestais no Pontal do Paranapanema**. São Paulo: Annablume, 2002. . Acesso em: 12 nov. 2019.

FALKOWSKI, M. J.; EVANS, J. S.; MARTINUZZI, S.; GESSLER, P. E.; HUDAK, A. T. Characterizing forest succession with lidar data: An evaluation for the Inland Northwest, USA. **Remote Sensing of Environment**, [s. l.], v. 113, n. 5, p. 946–956, 2009.

FERREIRA, M. P.; ZORTEA, M.; ZANOTTA, D. C.; SHIMABUKURO, Y. E.; DE SOUZA FILHO, C. R. Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data. **Remote Sensing of Environment**, [s. l.], v. 179, p. 66–78, 2016.

FIASCHI, P.; PIRANI, J. R. Review of plant biogeographic studies in Brazil. **Journal of systematics and evolution**, [s. l.], v. 47, n. 5, p. 477–496, 2009.

FONSECA, S. N.; RIBEIRO, J. H. C.; CARVALHO, F. A. Estrutura e diversidade da regeneração arbórea em uma floresta secundária urbana (Juiz de Fora, MG, Brasil). **Floresta e Ambiente**, [s. l.], v. 20, n. 3, p. 307–315, 2013.

GUERRA, T. N. F.; RODAL, M. J. N.; E SILVA, A. C. B. L.; ALVES, M.; SILVA, M. A. M.; DE ARAÚJO MENDES, P. G. Influence of edge and topography on the vegetation in an Atlantic Forest remnant in northeastern Brazil. **Journal of forest research**, [s. l.], v. 18, n. 2, p. 200–208, 2013.

HADDAD, N. M.; BRUDVIG, L. A.; CLOBERT, J.; DAVIES, K. F.; GONZALEZ, A.; HOLT, R. D.; LOVEJOY, T. E.; SEXTON, J. O.; AUSTIN, M. P.; COLLINS, C. D. Habitat fragmentation and its lasting impact on Earth's ecosystems. **Science advances**, [s. l.], v. 1, n. 2, p. e1500052, 2015.

HARGREAVES, P. Phytosociology in Brazil. **J Plant Sci Biotechnol**, [s. l.], v. 2, p. 12–20, 2008.

HYYPPÄ, J.; HYYPPÄ, H.; LECKIE, D.; GOUGEON, F.; YU, X.; MALTAMO, M. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. **International Journal of Remote Sensing**, [s. l.], v. 29, n. 5, p. 1339–1366, 2008.

JOLY, C. A.; METZGER, J. P.; TABARELLI, M. Experiences from the Brazilian Atlantic Forest: ecological findings and conservation initiatives. **New Phytologist**, [s. l.], v. 204, n. 3, p. 459–473, 2014.

LEAO, T. C.; FONSECA, C. R.; PERES, C. A.; TABARELLI, M. Predicting extinction risk of Brazilian Atlantic Forest angiosperms. **Conservation Biology**, [s. l.], v. 28, n. 5, p. 1349–1359, 2014.

LEFSKY, M. A.; COHEN, W. B.; HARDING, D. J.; PARKER, G. G.; ACKER, S. A.; GOWER, S. T. Lidar remote sensing of above-ground biomass in three biomes. **Global ecology and biogeography**, [s. l.], v. 11, n. 5, p. 393–399, 2002.

LIESENBERG, V.; BOEHM, H.-D. V.; GLOAGUEN, R. The contribution of CHRIS/PROBA data for tropical peat swamp landscape discrimination purposes. In: 2009 IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM 2009, **Anais...**: IEEE, 2009.

LIM, K.; TREITZ, P.; WULDER, M.; ST-ONGE, B.; FLOOD, M. LiDAR remote sensing of forest structure. **Progress in physical geography**, [s. l.], v. 27, n. 1, p. 88–106, 2003.

MALLET, C.; BRETAR, F. Full-waveform topographic lidar: State-of-the-art. **ISPRS Journal of photogrammetry and remote sensing**, [s. l.], v. 64, n. 1, p. 1–16, 2009.

MALTAMO, M.; NÆSSET, E.; VAUHKONEN, J. Forestry applications of airborne laser scanning. **Concepts and case studies. Manag For Ecosys**, [s. l.], v. 27, p. 2014, 2014.

MARTINS-NETO, R. P.; TOMMASELLI, A. M. G.; IMAI, N. N.; DAVID, H. C.; MILTIADOU, M.; HONKAVAARA, E. Identification of Significant LiDAR Metrics and Comparison of Machine Learning Approaches for Estimating Stand and Diversity Variables in Heterogeneous Brazilian Atlantic Forest. **Remote Sensing**, [s. l.], v. 13, n. 13, p. 2444, 2021.

MIYOSHI, G. T.; IMAI, N. N.; GARCIA TOMMASELLI, A. M.; ANTUNES DE MORAES, M. V.; HONKAVAARA, E. Evaluation of hyperspectral multitemporal information to improve tree species identification in the highly diverse atlantic forest. **Remote Sensing**, [s. l.], v. 12, n. 2, p. 244, 2020.

MMA; IBAMA; ICMBIO. **Plano de manejo da Estação Ecológica Mico - Leão - Preto**, 2007. Disponível em: <http://www.icmbio.gov.br/portal/images/stories/imgs-unidades-coservacao/eseq_mico_leao_preto.pdf>. Acesso em: 6 nov. 2019.

MYERS, N.; MITTERMEIER, R. A.; MITTERMEIER, C. G.; DA FONSECA, G. A.; KENT, J. Biodiversity hotspots for conservation priorities. **Nature**, [s. l.], v. 403, n. 6772, p. 853, 2000.

PARKER, G. G.; RUSS, M. E. The canopy surface and stand development: assessing forest canopy structure and complexity with near-surface altimetry. **Forest Ecology and Management**, [s. l.], v. 189, n. 1–3, p. 307–315, 2004.

PIROTTI, F. Analysis of full-waveform LiDAR data for forestry applications: a review of investigations and methods. **iForest-Biogeosciences and Forestry**, [s. l.], v. 4, n. 3, p. 100, 2011.

POPESCU, S. C.; WYNNE, R. H.; NELSON, R. F. Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. **Computers and electronics in agriculture**, [s. l.], v. 37, n. 1–3, p. 71–95, 2002.

REITBERGER, J.; KRZYSZEK, P.; STILLA, U. Analysis of full waveform lidar data for tree species classification. **International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences**, [s. l.], v. 36, n. Part 3, p. 228–233, 2006.

RIBEIRO, M. C.; METZGER, J. P.; MARTENSEN, A. C.; PONZONI, F. J.; HIROTA, M. M. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. **Biological conservation**, [s. l.], v. 142, n. 6, p. 1141–1153, 2009.

RODERJAN, C. V.; GALVÃO, F.; KUNIYOSHI, Y. S.; HATSCHBACH, G. G. As unidades fitogeográficas do estado do Paraná, Brasil. **Ciência & Ambiente**, [s. l.], v. 24, n. 1, p. 75–92, 2002.

SANCHEZ-AZOFEIFA, A.; ANTONIO GUZMÁN, J.; CAMPOS, C. A.; CASTRO, S.; GARCIA-MILLAN, V.; NIGHTINGALE, J.; RANKINE, C. Twenty-first century remote sensing technologies are revolutionizing the study of tropical forests. **Biotropica**, [s. l.], v. 49, n. 5, p. 604–619, 2017.

SCARANO, F. R.; CEOTTO, P. Brazilian Atlantic forest: impact, vulnerability, and adaptation to climate change. **Biodiversity and Conservation**, [s. l.], v. 24, n. 9, p. 2319–2331, 2015.

SHAN, J.; TOTH, C. K. **Topographic laser ranging and scanning: principles and processing**. [s.l.] : CRC press, 2018.

SHEN, X.; CAO, L.; CHEN, D.; SUN, Y.; WANG, G.; RUAN, H. Prediction of forest structural parameters using airborne full-waveform LiDAR and hyperspectral data in subtropical forests. **Remote Sensing**, [s. l.], v. 10, n. 11, p. 1729, 2018.

SHI, Y.; WANG, T.; SKIDMORE, A. K.; HEURICH, M. Important LiDAR metrics for discriminating forest tree species in Central Europe. **ISPRS journal of photogrammetry and remote sensing**, [s. l.], v. 137, p. 163–174, 2018.

SOS MATA ATLÂNTICA; INPE. **Atlas dos Remanescentes Florestais da Mata Atlântica: Relatório técnico período 2017 - 2018.**, 2019. Disponível em: <https://www.sosma.org.br/wp-content/uploads/2019/05/Atlas-mata-atlantica_17-18.pdf>. Acesso em: 6 nov. 2019.

SOTHE, C.; DALPONTE, M.; ALMEIDA, C. M. De; SCHIMALSKI, M. B.; LIMA, C. L.; LIESENBERG, V.; MIYOSHI, G. T.; TOMMASSELLI, A. M. G. Tree Species Classification in a Highly Diverse Subtropical Forest Integrating UAV-Based Photogrammetric Point Cloud and Hyperspectral Data. **Remote Sensing**, [s. l.], v. 11, n. 11, p. 1338, 2019.

TAUBERT, F.; FISCHER, R.; KNAPP, N.; HUTH, A. Deriving tree size distributions of tropical forests from LiDAR. **Remote Sensing**, [s. l.], v. 13, n. 1, p. 131, 2021.

VAUGHN, N. R.; MOSKAL, L. M.; TURNBLOM, E. C. Tree species detection accuracies using discrete point lidar and airborne waveform lidar. **Remote sensing**, [s. l.], v. 4, n. 2, p. 377–403, 2012.

VAUHKONEN, J.; ENE, L.; GUPTA, S.; HEINZEL, J.; HOLMGREN, J.; PITKÄNEN, J.; SOLBERG, S.; WANG, Y.; WEINACKER, H.; HAUGLIN, K. M. Comparative testing of single-tree detection algorithms under different types of forest. **Forestry**, [s. l.], v. 85, n. 1, p. 27–40, 2012.

WERNECK, M. de S.; SOBRAL, M. E. G.; ROCHA, C. T. V.; LANDAU, E. C.; STEHMANN, J. R. Distribution and endemism of angiosperms in the Atlantic Forest. **Natureza & Conservação**, [s. l.], v. 9, p. 188–193, 2011.

WILLIAMS, K. J.; FORD, A.; ROSAUER, D. F.; DE SILVA, N.; MITTERMEIER, R.; BRUCE, C.; LARSEN, F. W.; MARGULES, C. Forests of East Australia: the 35th biodiversity hotspot. In: **Biodiversity hotspots**. [s.l.] : Springer, 2011. p. 295–310.

CHAPTER 6 – FINAL REMARKS

Knowledge of the structure of the few remnants of the Atlantic Forest is necessary for decision making and the development of conservation and sustainable management plans to mitigate the effects of degradation over time. This work shows the potential of using high-resolution Remote Sensing to estimate structural variables in a small remnant of inland Atlantic Forest. The performance and effectiveness of peak return and full waveform LiDAR data for estimation of stand and diversity variables and the combination of these data with hyperspectral images for classification of tree species were assessed. This final chapter discusses the highlights, open problems, and recommendations for future studies based on the results achieved.

(1) The study area, the *Ponte Branca* forest remnant, was characterized in Chapter 2, and is a transition area between the Atlantic Forest and Cerrado biomes (with the presence of tree species typical of the Cerrado). The species accumulation curve presents an almost stability with 15 plots, that is, increasing the number of plots does not significantly increase the number of sampled species. Even though a larger number of surveyed plots is indicated, mainly for applications using machine learning approaches for variables estimation, as presented in the other chapters. Instead of using a few plots with large dimensions, decreasing the plot size and increasing the number of samples is more appropriate. Furthermore, the rectangular shape better captures the heterogeneity of the area than the square shape. For studies that do not aim at repeated measurements over time, the use of circular plots or variable area sampling methods (e.g., quadrant method) is also indicated.

(2) For the use of LiDAR metrics as input data in models to estimate forest variables, a pre-selection of metrics is necessary due to the high correlation and to reduce the dimensionality of the data. The transformation of the 54 peak returns LiDAR metrics related to statistics of heights, intensities and pulse returns by Principal Component Analysis (PCA) was effective for estimating seven stand and diversity variables (Chapter 3). Each Principal Component (PC) was able to synthesize relevant information from the set of LiDAR metrics. Among the estimation methods tested, artificial neural networks were the method that produced the best results. However, to achieve these results, we used complex architectures (with two and three hidden layers) that can result in overfitting. The tested architectures must be used with a larger database to check for overfitting. Furthermore, even with the best results, the variables basal area (BA), density (DEN) and the number of tree species (NTS) showed high estimation

errors due to the great variability and difficulty in estimating these variables especially in tropical forests.

(3) A voxel-based approach was used to verify a possible improvement in estimations of the BA, DEN, NTS and aboveground biomass (AGB) variables (Chapter 4) not only for the peak return LiDAR also for full waveform data. Two different flight heights (900 m and 2000 m) were also tested to understand how they influence the estimations. Low-level filtering is needed to eliminate noise signals in FWF data. The chosen low-level filtering threshold is site-specific and depends on the sensor and flight height used. Variable estimations using attributes extracted from waveforms were less biased than PR estimates. There were no significant differences regarding flight height, but lower flight heights require higher filtering thresholds as the sensor records more noisy signals. The results indicate a potential for the use of FWF data in tropical forests and the voxel-based approach available in the DASOS software, as there was an improvement in the errors in the estimates of the four variables, when compared to traditional peak return metrics. For future studies, combine the metrics related to the distribution of heights, intensity, and pulse return from peak return LiDAR data with the attributes extracted from voxelized space of FWF data, to verify improvements in estimations.

(4) The classification of tree species is a complicated task, especially in tropical forests due to the great heterogeneity even among species. In Chapter 5 we tested the combinations of peak return and full waveform LiDAR data with hyperspectral images to understand how geometric information from LiDAR and spectral information from hyperspectral images improve classification accuracy. Among the tested scenarios, the combination of vegetation indices extracted from hyperspectral images with peak return LiDAR metrics, transformed by PCA, resulted in the best classification accuracy. The delineation of the tree crowns was performed manually as the great variability in the tree crowns makes the automatic segmentation process difficult and challenging. For future studies, adapt and develop techniques for automatic segmentation of trees in tropical forests to automate the process of extracting features from individual trees.

(5) Unlike what is found in literature, full waveform LiDAR data used isolated or combined with hyperspectral images were not effective in the classification of tree species, with low accuracy (Chapter 5). We have used different sets of full waveform metrics than other studies, so new tests must be performed by changing the parameters of the software DASOS, like the size of 3D window for extracting attributes and normalizing waveform samples intensities, to verify improvements in classification.

(6) As the forest remnants of the Atlantic Forest are very fragmented and the areas are small, many of these areas have several successional stages, from areas of initial regeneration to areas of mature forest. Therefore, it is recommended to perform field campaigns simultaneously or as close as possible, to the flight campaigns, mainly due to the tree species' phenology and the forest dynamics. The temporal lag between the different collection dates, can result in both geometric and spectral differences.

(7) Finally, we suggest to apply the methods used in this thesis to other forest remnants and larger areas of tropical forest (e.g. Amazon forest) and verify if the accuracy obtained in the estimates and classification is maintained in different forests typologies of different sizes. We also encourage the application of the presented methods and algorithms to data acquired by orbital sensors and medium spatial resolution sensors to understand if they are suitable to a wider spectrum of remote sensing data available for the estimation of structural parameters of forest as well as the classification of tree species.