INPE
UNIDADE DE PESQUISA DO MCTIC

MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA,
INOVAÇÕES E COMUNICAÇÕES

PÁTRIA AMADA
BRASIL
GOVERNO FEDERAL

sid.inpe.br/mtc-m21c/2020/04.03.20.59-TDI

INTEGRATION OF LIDAR AND HYPERSPECTRAL DATA FOR FOREST DISTURBANCE CHARACTERIZATION AND ABOVEGROUND BIOMASS ESTIMATION IN THE BRAZILIAN AMAZON

Catherine Torres de Almeida

Doctorate Thesis of the Graduate Course in Remote Sensing, guided by Drs. Lênio Soares Galvão, and Luiz Eduardo Oliveira e Cruz de Aragão, approved in March 04, 2020.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34R/4299LFP>

INPE
São José dos Campos
2020

PUBLISHED BY:

Instituto Nacional de Pesquisas Espaciais - INPE
Gabinete do Diretor (GBDIR)
Serviço de Informação e Documentação (SESID)
CEP 12.227-010
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/7348
E-mail: pubtc@inpe.br

sid.inpe.br/mtc-m21c/2020/04.03.20.59-TDI

INTEGRATION OF LIDAR AND HYPERSPECTRAL DATA FOR FOREST DISTURBANCE CHARACTERIZATION AND ABOVEGROUND BIOMASS ESTIMATION IN THE BRAZILIAN AMAZON

Catherine Torres de Almeida

Doctorate Thesis of the Graduate Course in Remote Sensing, guided by Drs. Lênio Soares Galvão, and Luiz Eduardo Oliveira e Cruz de Aragão, approved in March 04, 2020.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34R/4299LFP>

INPE
São José dos Campos
2020

Cataloging in Publication Data

Almeida, Catherine Torres de.

Al64i Integration of LiDAR and hyperspectral data for forest disturbance characterization and aboveground biomass estimation in the Brazilian amazon / Catherine Torres de Almeida. – São José dos Campos : INPE, 2020.

xxv + 128 p. ; (sid.inpe.br/mtc-m21c/2020/04.03.20.59-TDI)

Thesis (Doctorate in Remote Sensing) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2020.

Guiding : Drs. Lênio Soares Galvão, and Luiz Eduardo Oliveira e Cruz de Aragão.

1. Hyperspectral Remote Sensing. 2. Laser Scanning. 3. Data Fusion. 4. Tropical Forest. 5. Secondary Successions. I.Title.

CDU 528.8:630*221(811)

Aluno (a): ***Catherine Torres de Almeida***

Título: "INTEGRATION OF LIDAR AND HYPERSPECTRAL DATA FOR FOREST DISTURBANCE CHARACTERIZATION AND ABOVEGROUND BIOMASS ESTIMATION IN THE BRAZILIAN AMAZON"

Aprovado (a) pela Banca Examinadora em cumprimento ao requisito exigido para obtenção do Título de *Doutor(a)* em

Sensoriamento Remoto

Dra. Liana Oighenstein Anderson

Liana Oightenstein Anderson

Presidente / CEMADEN / São José dos Campos - SP

( ) Participação por Video - Conferência

(X) Aprovado ( ) Reprovado

Dr. Lenio Soares Galvão

Orientador(a) / INPE / SJCampos - SP

( ) Participação por Video - Conferência

(X) Aprovado ( ) Reprovado

Dr. Luiz Eduardo Oliveira e Cruz de Aragão

Orientador(a) / INPE / São José dos Campos - SP

( ) Participação por Video - Conferência

(X) Aprovado ( ) Reprovado

Dr. Jean Pierre Henry Balbaud Ometto

Membro da Banca / INPE / São José dos Campos - SP

( ) Participação por Video - Conferência

(X) Aprovado ( ) Reprovado

Este trabalho foi aprovado por:

( ) maioria simples

(X) unanimidade

São José dos Campos, 04 de março de 2020

Aprovado (a) pela Banca Examinadora em cumprimento ao requisito exigido para obtenção do Título de *Doutor(a)* em

Sensoriamento Remoto

Dr. Paulo Maurício Lima de Alencastro Graça

Convidado(a) / INPA / Manaus - AM

( ) Participação por Video - Conferência

(X) Aprovado ( ) Reprovado

Dr. Michael Maier Keller

Liana Oighenstein Anderson

Convidado(a) / USDA Forest Service / Estados Unidos - USA

(X) Participação por Video - Conferência

(X) Aprovado ( ) Reprovado

Este trabalho foi aprovado por:

( ) maioria simples

(X) unanimidade

São José dos Campos, 04 de março de 2020

To Patricia Torres and Nelson Almeida, the best parents I could have.

To José Ricardo de O. Nascimento Jr., the best company for all moments.

ACKNOWLEDGMENTS

This thesis would not be possible without the support of many people, to whom I am immensely grateful.

To my advisors, Lênio Soares Galvão and Luiz Eduardo Oliveira e Cruz de Aragão, for the incredible opportunity to work with this topic and for the valuable suggestions.

To the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for granting the scholarship (grant number 140502/2016-5).

To the CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil, Finance Code 001) for institutional support.

To the INPE (Instituto Nacional de Pesquisas Espaciais) and to the postgraduate studies in Remote Sensing, for the excellence in education and research.

To the Project Environmental Satellite Monitoring in the Amazon Biome (MSA-BNDES) - Activity 7 “Improvement of biomass estimation methods and emission estimation models for change of land use” funded by Amazon Fund (process #14209291), for providing the airborne remote sensing data.

To the Sustainable Landscapes Brazil project supported by the EMBRAPA (Brazilian Agricultural Research Corporation), the US Forest Service, and USAID, and the US Department of State, for providing much of the field inventory data.

To the FATE (Fire-Associated Transient Emissions in Amazonia) program, the IDSM (Instituto de Desenvolvimento Sustentável Mamirauá), and the LMF/INPA (Laboratório de Manejo Florestal do Instituto Nacional de Pesquisas da Amazônia) for providing part of the field data.

To Jean Pierre Henry Balbaud Ometto, Aline Daniele Jacon, Francisca Rocha de Souza Pereira, Luciane Yumie Sato, Aline Pontes Lopes, Paulo Maurício Lima de Alencastro Graça, Camila Valéria de Jesus Silva, Jefferson Ferreira-Ferreira, and Marcos Longo, for their valuable contribution to the study of AGB modeling and elaboration of the derived paper.

To the editors and three reviewers of the paper published in the journal Remote Sensing of Environment (chapter 6 of this thesis), for the constructive suggestions.

To the members of my thesis committee, who openly accepted to evaluate and discuss this study.

To the colleagues of the TREES lab, for the interesting scientific discussions.

To the “segunda série C”, for all moments of relaxation.

To Rafael Coll Delgado and Kaio Allan Gasparini, for encouraging me to do a PhD at INPE.

To all my teachers, for sharing their knowledge and for contributing to the education of Brazil.

Finally, I am very grateful to my family for their support and encouragement.

ABSTRACT

Advancements in remote sensing technologies provide new opportunities to answer complex ecological questions in tropical forests, which play a crucial role on the stability of global biogeochemical cycles and biodiversity. Light Detection And Ranging (LiDAR) and Hyperspectral Imaging (HSI) provide complementary information that can potentially improve the characterization of tropical forests and reduce the uncertainties in estimating greenhouse gas emissions from deforestation and forest degradation. This thesis aims to explore optimal procedures for improving tropical forest disturbance characterization and aboveground biomass (AGB) modeling using integrated LiDAR and HSI data and advanced machine learning algorithms. The study area covered 12 sites distributed across the Brazilian Amazon biome, spanning a variety of environmental and anthropogenic conditions. The methods were divided into three parts: (1) classification of forest disturbance status (Chapter 5); (2) AGB modeling (Chapter 6); and (3) analysis of the AGB variability according to anthropogenic and environmental variables (Chapter 7). Firstly, four classes of forest disturbance (undisturbed forests, disturbed mature forests, and two stages of secondary forests) were identified using Landsat time series between 1984 and 2017. Several LiDAR and HSI metrics obtained over 600 sample plots were then used as input data to three machine learning models for distinguishing those classes. Secondly, georeferenced inventory data from 132 sample plots were used to obtain a reference field AGB. A great number of LiDAR and HSI metrics (45 and 288, respectively) were submitted to a correlation filtering followed by a feature selection procedure (recursive feature elimination) to optimize the performance of six regression models. Finally, the average of AGB predictions from the best multisensor models was calculated over 600 sample plots where field AGB data were not available. A multivariable linear regression model was then used to assess the extent to which the predicted AGB variability was affected by anthropogenic (disturbance type and time) and environmental (annual rainfall, climatic water deficit, and topography) factors in secondary and mature forests. Overall, the results showed that the combination of LiDAR and HSI data improved both the classification of forest disturbances and the estimation of AGB compared to using a single data source. Using multisource remote sensing data was more effective than using advanced machine learning for both classification and regression models. The LiDAR-based upper canopy cover and the HSI-based absorption bands in the near-infrared (NIR) and shortwave infrared (SWIR) spectral regions were the most influential metrics for characterizing the disturbance status and estimating AGB. Anthropogenic disturbances played the greatest effect on predicted AGB variability, reducing up to 44% the AGB of disturbed mature forests compared to the undisturbed ones. Secondary forests displayed an AGB recovery rate of 4.4 $\mathrm{Mg.ha^{-1}.yr^{-1}}$. Water deficit also affected the variability of AGB in both mature and secondary forests, suggesting a lower recovery potential in water-stressed areas. The results highlight the potential of integrating LiDAR and HSI data for improving our understanding of forest dynamics in the face of increasing anthropogenic global changes.

Keywords: Hyperspectral remote sensing. Laser scanning. Data fusion. Tropical forest. Secondary successions. Forest degradation. Carbon stock.

INTEGRAÇÃO DE DADOS LIDAR E HIPERESPECTRAIS PARA A CARACTERIZAÇÃO DE DISTÚRBIOS FLORESTAIS E A ESTIMATIVA DA BIOMASSA ACIMA DO SOLO NA AMAZÔNIA BRASILEIRA

RESUMO

Os avanços nas tecnologias de sensoriamento remoto oferecem novas oportunidades para responder a questões ecológicas complexas em florestas tropicais, que desempenham um papel crucial nos ciclos biogeoquímicos globais e na biodiversidade. O sensoriamento remoto LiDAR (*Light Detection And Ranging*) e HSI (imageamento hiperespectral) fornecem informações complementares que podem melhorar a caracterização das florestas tropicais e reduzir as incertezas na estimativa das emissões de gases de efeito estufa devido ao desmatamento e degradação florestal. Esta tese visa explorar os procedimentos ideais para melhorar a caracterização de distúrbios das florestas tropicais e a modelagem de biomassa acima do solo (AGB) através do uso de dados LiDAR e HSI integrados e algoritmos avançados de aprendizado de máquina. A área de estudo abrangeu 12 locais distribuídos no bioma Amazônia no Brasil, incluindo uma variedade de condições ambientais e antropogênicas. Os métodos foram divididos em três partes: (1) classificação do status de distúrbio florestal (Capítulo 5); (2) modelagem da AGB (Capítulo 6); e (3) análise da variabilidade da AGB segundo variáveis antropogênicas e ambientais (Capítulo 7). Primeiramente, quatro classes de distúrbios florestais (florestas não perturbadas, florestas maduras perturbadas e dois estágios de florestas secundárias) foram identificadas usando séries temporais do Landsat entre 1984 e 2017. Várias métricas de LiDAR e HSI obtidas em 600 parcelas amostrais foram usadas como dados de entrada em três modelos de aprendizado de máquina para distinguir essas classes. Em segundo lugar, dados de inventário georreferenciados de 132 parcelas amostrais foram usados para obter a AGB de referência. Um grande número de métricas LiDAR e HSI (45 e 288, respectivamente) foram submetidas a um filtro de correlação seguido de um procedimento de seleção de atributos (*Recursive Feature Elimination*) para otimizar o desempenho de seis modelos de regressão. Finalmente, a média das estimativas de AGB derivadas dos melhores modelos multisensores foi calculada em 600 parcelas amostrais onde os dados de AGB de campo não estavam disponíveis. Um modelo de regressão linear multivariável foi então usado para avaliar até que ponto a variabilidade da AGB é afetada por fatores antropogênicos (tipo e tempo de distúrbio florestal) e ambientais (precipitação anual, déficit hídrico climático e topografia) em florestas secundárias e maduras. No geral, os resultados obtidos nos três capítulos mostraram que a combinação dos dados LiDAR e HSI melhorou a classificação dos distúrbios florestais e a estimativa da AGB em comparação ao uso de uma única fonte de dados. O uso de dados de sensoriamento remoto de várias fontes foi mais eficaz do que as técnicas avançadas de aprendizado de máquina para os modelos de classificação e regressão. A cobertura superior do dossel baseada em dados LiDAR e as bandas de absorção baseadas em dados HSI nas regiões espectrais de infravermelho próximo e infravermelho de ondas curtas foram as métricas mais influentes para caracterizar o status de perturbação e estimar a AGB. Os distúrbios antropogênicos tiveram o maior efeito na variabilidade da AGB derivada de dados multisensores, reduzindo em até 44% a AGB de florestas maduras perturbadas em comparação com as não perturbadas. As florestas secundárias apresentaram uma taxa de recuperação de AGB de 4,4 Mg.ha^{-1}.ano^{-1}. O déficit hídrico também afetou a variabilidade da AGB em florestas maduras e secundárias, sugerindo um menor potencial de recuperação em áreas sob alto estresse hídrico. Os resultados destacam o potencial da integração de dados LiDAR e HSI para melhorar nosso entendimento da dinâmica florestal diante das crescentes mudanças globais antropogênicas.

Palavras-chave: Sensoriamento remoto hiperespectral. Perfilamento a laser. Fusão de dados. Foresta tropical. Sucessões secundárias. Degradação florestal. Estoque de carbono.

LIST OF FIGURES

Page

LIST OF TABLES

Page

LIST OF ACRONYMS AND ABBREVIATIONS

3D	Three-dimensional
AGB	Aboveground Biomass
AIC	Akaike Information Criterion
ALI	Advanced Land Imager
AM	Amazonas state
ANOVA	Analysis Of Variance
AVIRIS	Airborne Visible Infrared Imaging Spectrometer
BLA	Brazilian Legal Amazon
CART	Classification and Regression Trees
CB	Cubist
CHM	Canopy Height Model
CHRIS	Compact High Resolution Imaging Spectrometer
CI	Confidence Interval
CV	Cross-validation
CWD	Climatic Water Deficit
DBH	Diameter at Breast Height
DEM	Digital Elevation Model
DF	Disturbed mature Forest
DSCI	Simpson Structural Complexity Index
DSM	Digital Surface Model
DTM	Digital Terrain Model
ENSO	El Niño Southern Oscillation
EO-1	Earth Observing-1
EVI	Enhanced Vegetation Index
ETM+	Enhanced Thematic Mapper Plus
FE	Feature Extraction
FLONA	National Forest (*Floresta Nacional* in Portuguese)
FS	Feature Selection
GPS	Global Positioning System
GV	Green Vegetation

HSCI	Shannon Structural Complexity Index
HSI	Hyperspectral Imaging
IHS	Intensity-Hue-Saturation
IMU	Inertial Measurement Unit
kNN	k-Nearest Neighbors
LAD	Leaf Area Density
LAI	Leaf Area Index
LiDAR	Light Detection And Ranging
LM	Linear Models
LMR	Linear Model with Ridge Regularization
LULC	Land Use and Land Cover
LVIS	Land, Vegetation, and Ice Sensor
MAP	Mean Annual Precipitation
ML	Maximum Likelihood
MNF	Minimum Noise Fraction
MSI	Multispectral Imaging
MT	Mato Grosso state
NDVI	Normalized Difference Vegetation Index
NIR	Near-Infrared Radiation
NP	Non-Photosynthetic vegetation/soil
NPV	Non-Photosynthetic Vegetation
OA	Overall Accuracy
PA	Pará state
PCA	Principal Component Analysis
PD	Point Density
PLSR	Partial Least Square Regression
PPI	Pixel Purity Index
PRI	Photochemical Reflectance Index
QQ	Quantile-Quantile
R^2	Coefficient of Determination
RADAR	RAdio Detection And Ranging
RBF	Radial Basis Function
RF	Random Forest
RFE	Recursive Feature Elimination

RMSE	Root Mean Squared Error
RO	Rondônia state
SF	Secondary Forest
$SF_{1\text{-}15yr}$	Initial-to-intermediate Secondary Forest
$SF_{16\text{-}32yr}$	Advanced Secondary Forest
SGB	Stochastic Gradient Boosting
SMA	Spectral Mixture Analysis
SVM	Support Vector Machine
SVR	Support Vector Regression
SWIR	Shortwave Infrared Radiation
SZA	Solar Zenith Angle
TCH	Top of Canopy Height
UF	Undisturbed Forest
USA	United States of America
VNIR	Visible and Near-Infrared Radiation
WD	Wood Density

SUMMARY

1 INTRODUCTION

Remote sensing is an essential tool for forest mapping, monitoring, and modeling, especially in large-scale studies. The availability of remotely sensed data from different sensors and platforms, spanning a wide range of spatial, spectral, radiometric, and temporal resolutions, has enabled various applications on forest resources. Since the launching of the first remote sensing system specifically designed for natural resource monitoring (the Landsat-1 satellite almost 50 years ago), remote sensing technology has been rapidly advancing (BOYD; DANSON, 2005). Such advances provide new opportunities to answer complex ecological questions from local to regional scales.

From global satellites to local drones, improvements in spatial and spectral resolutions have allowed a more detailed characterization of forests, such as the detection of individual trees (FERRAZ et al., 2016) and the identification of forest species (BALDECK et al., 2015). The development of hyperspectral imaging (HSI) systems, acquiring data in hundreds of narrow and contiguous spectral bands, has generated high-resolution reflectance spectra on a per-pixel basis (GOETZ et al., 1985). HSI sensors are capable of retrieving information on the biochemical composition of canopies in order to better understand forest ecosystem functioning (KOKALY et al., 2009). Moreover, the emergence of the Light Detection And Ranging (LiDAR) technology has produced three-dimensional measurements of forests, allowing the quantification of important structural attributes, such as the canopy height, leaf area density, and aboveground biomass (AGB) (LEFSKY et al., 2002a). However, LiDAR systems currently capture limited spectral information, which creates difficulties to distinguish structurally similar forests with distinct species composition or under stress conditions. In this regard, HSI sensors can complement the information produced by LiDAR instruments.

Unfortunately, along with the advancement of remote sensing technologies, the last decades have been marked by increasing anthropogenic pressure on forests and, consequently, by a growing concern on the imminent risks of climate change (IPCC, 2019). In this context, tropical forests are especially relevant because they support the greatest biodiversity in the world (DIRZO; RAVEN, 2003) and are fundamental for the global carbon cycle (BONAN, 2008). However, the great heterogeneity and complexity of tropical forests pose a challenge in obtaining accurate information on their composition and structure through conventional remote sensing approaches. The data integration from different sensors, especially the advanced LiDAR and HSI instruments, is an attractive alternative to improve the characterization of tropical forests and reduce the uncertainties in estimating greenhouse gas emissions from deforestation and forest degradation (TORABZADEH et al., 2014). In spite of the potential of the LiDAR and HSI complementary information, just a few studies have integrated these multisensor data for the characterization of tropical forests (ASNER et al., 2015; CLARK et al., 2011; VAGLIO LAURIN et al., 2014). None of them has been conducted over tropical forests of the Brazilian Amazon, considering also the different environmental and anthropogenic conditions of the region.

Therefore, the main objective of this study is to evaluate the potential of the combination of LiDAR and HSI data for characterizing forest disturbance status and estimating the AGB of the Brazilian Amazon.

The specific objectives of this research are to:

- Analyze how metrics derived from HSI and LiDAR data vary as a function of different types of anthropogenic forest disturbances and AGB classes.
- Compare the performance of LiDAR and HSI data used alone and in combination to classify forest disturbance status and estimate AGB.
- Test the performance of different prediction methods, including advanced machine learning techniques.
- Estimate AGB under different environmental and anthropogenic conditions from the best combination of remote sensing data source (LiDAR, HSI, or their integration) and prediction method.
- Evaluate the effect of anthropogenic and environmental factors on the estimated AGB.

To achieve these objectives, the thesis is structured into eight chapters:

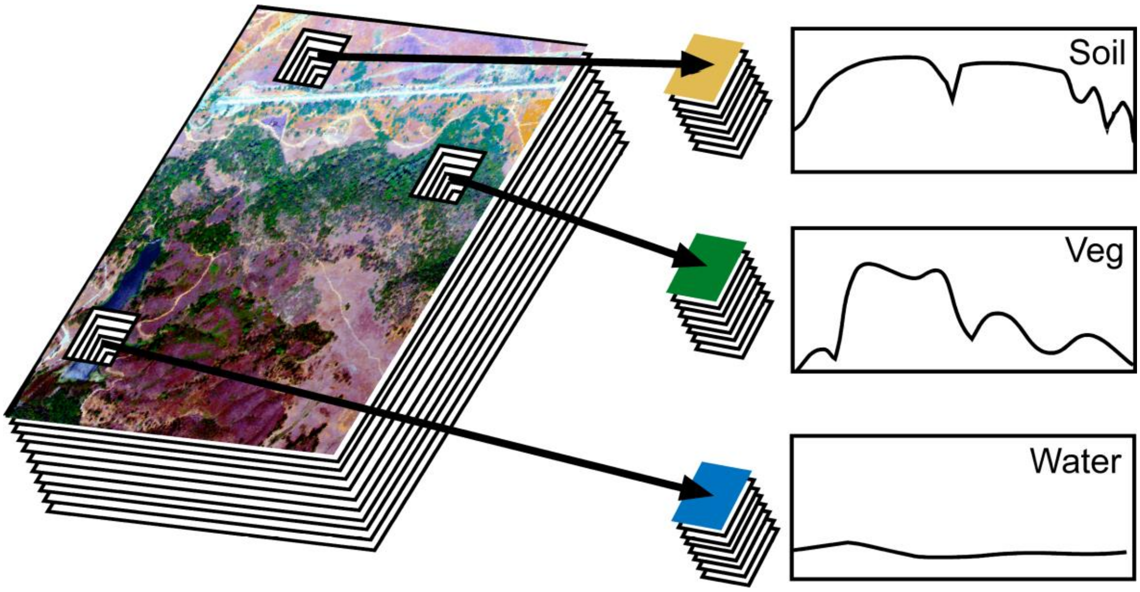
- Chapter 1 (this chapter) presents a brief introduction to the scope of the work and the objectives to be achieved.
- Chapter 2 presents a literature review on the main characteristics of HSI and LiDAR remote sensing technologies and the overall framework of multisensor data integration. The chapter also reports how LiDAR and HSI data integration has been applied to forest research.
- Chapter 3 describes the LiDAR and HSI data acquisition over the study area and related metrics.
- Chapter 4 presents the general methodology used to reach the goals of the thesis, which are detailed in the subsequent chapters (5 to 7) in the format of scientific articles.
- Chapter 5 is dedicated to investigating the integration of LiDAR and HSI data for classifying tropical forest disturbance status (undisturbed forests, disturbed forests, and two stages of secondary successions).
- Chapter 6 explores the optimal procedures for estimating AGB based on different data sources (LiDAR, HSI, and their combination) and regression algorithms (statistical linear models and advanced machine learning models).
- Chapter 7 uses the estimated AGB, derived from the best models obtained in the previous chapter, to investigate how environmental (climate and topography) and anthropogenic (disturbance type and time) factors affect the AGB in the Brazilian Amazon.
- Finally, Chapter 8 presents the concluding remarks obtained from the integrated analysis of all results from the three predecessor chapters.

2 LITERATURE REVIEW

2.1 Hyperspectral Imaging (HSI)

HSI is a passive remote sensing technique characterized by the simultaneous acquisition of images in a large number of narrow and contiguous spectral bands (GOETZ et al., 1985). In reality, this concept is much more related to the ability of the sensors to measure narrow bands (bandwidth) than to the number of bands itself (GALVÃO et al., 2012). For instance, the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) with 224 bands (400-2500 nm) and the Compact High Resolution Imaging Spectrometer (CHRIS/PROBA) with 62 bands (410-1000 nm) are both HSI instruments because they have narrow (10 nm bandwidth) and contiguous bands in each spectral range of operation. HSI is also known as imaging spectroscopy or imaging spectrometry, because it combines conventional imaging technologies with spectroradiometry, resulting in a multidimensional spatial-spectral image. Since the hyperspectral images are acquired using spectral resolution close to that observed in non-imaging laboratory spectrometers, their resultant spectra allow adequate measurements of most absorption bands that appear in different land covers (e.g., soil, vegetation, and water) on a per-pixel basis (Figure 2.1).

Figure 2.1 - Hyperspectral imaging (HSI) concept.



Source: Shippert (2003).

The ability of HSI to extract more accurate and detailed information compared to other passive remote sensors makes it suitable for a wide variety of applications. Several studies have shown the potential of hyperspectral data for forest applications, such as the classification of land cover (CLARK; KILHAM, 2016) or tree species (BALDECK et al., 2015; FÉRET; ASNER, 2013); identification of physiological responses to stress (SANCHES et al., 2014); estimation of biochemical variables (KOKALY et al., 2009); detection of burned areas (SCHEPERS et al., 2014); study of the canopy phenology (DE MOURA et al., 2017; GALVÃO et al., 2011), among others.

However, the high dimensionality of the HSI data also entails some challenges, such as the need for increased storage capacity and greater complexity in image processing. In addition, narrow and contiguous spectral bands may present great data redundancy, causing multicollinearity problems in several analytical procedures. To minimize these problems, different strategies of Feature Extraction (FE) and Feature Selection (FS) have been proposed to enhance information on the object or phenomenon of interest, while decreasing the data dimensionality and redundancy (BAJWA; KULKARNI, 2011). While the FE techniques generally include hyperspectral data transformation by principal component analysis (PCA) or minimum noise fraction (MNF) and the use of the first few components in the subsequent analysis, the FS approach uses different algorithms (e.g., correlation-based FS) to select the best variables for a given purpose. FS is preferable over FE because the selected spectral attributes are more easily interpretable from a physical point-of-view than the statistically transformed data (DAMODARAN et al., 2017). Different HSI metrics can be used as input variables for classification and regression modeling. They can include the reflectance of the bands; the PCAs and MNFs; the absorption band parameters calculated from the continuum removal method; several narrowband vegetation indices; and the endmember fractions retrieved from linear spectral mixture models (Table 2.1).

Table 2.1 - Examples of metrics that can be retrieved from HSI data and used for classification and regression modeling.

Metrics	**Description**	**Advantages**	**Limitations**	**Examples**
Spectral bands	Reflectance image of a narrow spectral channel	Contains the original information	A large number of bands can lead to redundancy	Reflectance at red and NIR bands
PCA	Linear combination of the original bands, which preserves the variance contained in the data and yields decorrelated components	Reduces redundancy	Difficult interpretation	First *n* PCA
MNF	Linear combination of the original bands to reduce dimensionality and minimize noise, yielding decorrelated components	Reduces redundancy and noise	Difficult interpretation	First *n* MNF
Continuum-removed absorption features	The continuum removal is a normalization technique to filter out absorption features	Easily interpretable, based on knowledge of the relation between the structure and chemical composition of targets at specific wavelengths	Requires high signal-to-noise ratio	Depth, width, area, and asymmetry of the absorption band
Vegetation indices	Arithmetic combination between bands, such as band ratio or normalized band difference	Can reduce solar and terrain illumination effects, while enhancing compositional information	Requires identification of appropriate bands	NDVI, EVI, PRI
Endmember fractions	The model assumes that the pixel response is a linear combination of a set of endmembers (e.g., green vegetation, soil and shade) for retrieving their abundance-fraction images on a per-pixel basis	Allows to evaluate the proportion of a particular endmember of interest	Difficulty of endmember selection representative of all land covers in the scene	Fractions of green vegetation, NPV, soil and shade

Source: Adapted from NUMATA (2012) and PU (2012).

2.2 Light Detection and Ranging (LiDAR)

LiDAR, also known as laser scanning, is an active remote sensing technique based on the emission and reception of laser pulses. The basic measurement of a LiDAR device is the time elapsed between the emission of a laser beam and the arrival of its reflection at the sensor. Multiplying this time interval by the speed of light and then dividing by two (to consider round trip) results in the distance between the sensor and the target (LEFSKY et al., 2002a). The laser beam can hit multiple objects located at different distances from the sensor. Thus, the nearest point will cause a faster returning pulse and the farthest point will take a longer return, allowing the distinction of elevations. LiDAR-derived distance measurements, accompanied by sensor position and laser beam direction information, enable obtaining the three-dimensional (3D) coordinates of each imaged point on the earth's surface (CHEN, 2014). For this purpose, LiDAR systems aboard mobile platforms (e.g., aircraft) need to integrate laser sensor information with a differential GPS and an Inertial Measurement Unit (IMU). The GPS records the platform's 3D position, while the IMU registers its orientation (roll, pitch, and yaw) (LARGE; HERITAGE, 2009).

LiDAR systems can be characterized according to:

(1) Platform: terrestrial (fixed or mobile), airborne, and spaceborne;

(2) Spectral wavelength: Most LiDAR systems operate in a single range of the electromagnetic spectrum, commonly in the green or near-infrared region. Systems applied to vegetation studies usually use sensors that operate in the near-infrared, region of the highest reflectance of vegetation (LEFSKY et al., 2002a). Some systems operating in two or more spectral ranges have been tested (WEI et al., 2012). There are already prototypes of terrestrial hyperspectral LiDAR sensors that emit a “supercontinuum laser” spanning the 420-1680 nm range (HAKALA et al., 2012);

(3) Footprint: the projection size of the laser beam on the ground, being classified as small (less than 1 m in diameter), medium (~ 10-30 m in diameter) and large (diameters greater than 50 m) (LU et al., 2012);

(4) Type of return record: discrete-return or full-waveform. In discrete-return LiDAR, data is often referred to as a “point cloud”, where each point contains

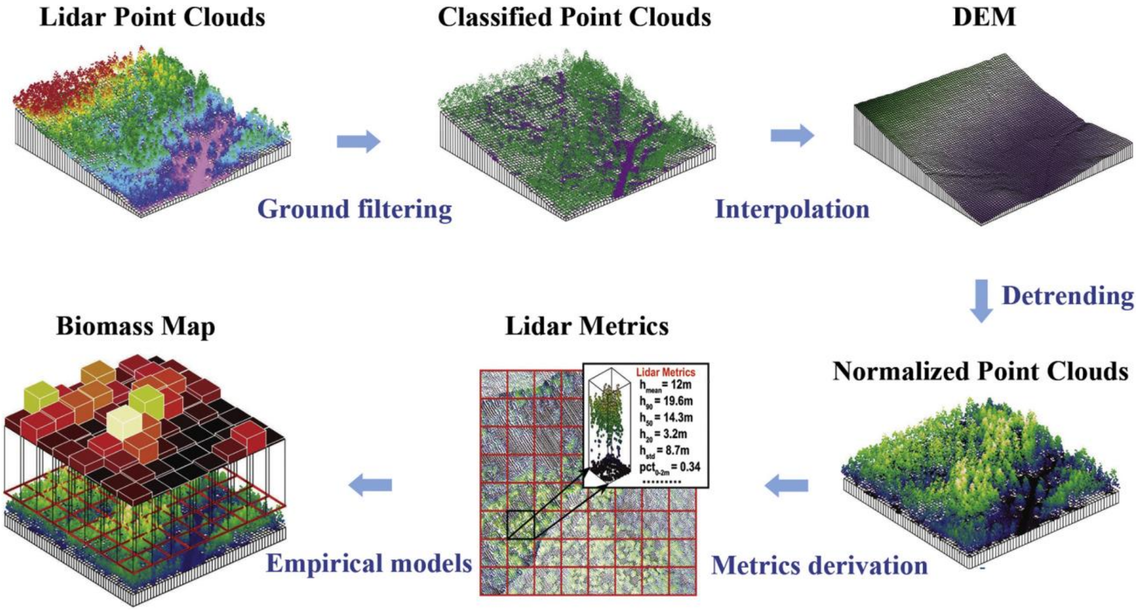
information such as 3D coordinates, signal intensity, and return type (first, last, or intermediate). In full-waveform LiDAR devices, the distribution of the entire return signal is recorded as a function of time (GIONGO et al., 2010).

According to the characteristics described above, the most commonly used LiDAR systems for forestry applications are airborne, operating in near-infrared, with small-footprint and discrete-return (VAUHKONEN et al., 2014). This type of LiDAR will be addressed in the following paragraphs since it was the one used in this study.

Metrics derived from LiDAR data can be used to model forest canopy structural properties such as diameter at breast height (DBH), basal area, stem volume, and biomass. Two approaches have been used to estimate forest properties from LiDAR data: individual tree-based and area-based approaches (GOLDBERGS et al., 2018). If the LiDAR point density is sufficient to obtain multiple returns from the same tree, it is possible to extract attributes by the individual tree through segmentation procedures for canopy delineation. The area-based approach consists of obtaining metrics from the LiDAR points divided into a regular grid (STRAUB et al. al., 2009).

Obtaining LiDAR metrics requires pre-processing of raw data, usually involving filtering, interpolation, and detrending processes (Figure 2.2). Firstly, the LiDAR point cloud is classified into ground and non-ground points. Then, the ground points are interpolated into a digital terrain/elevation model (DTM or DEM). The DTM is subtracted from the raw point cloud, yielding a detrended point cloud that has a ground elevation of zero and captures canopy vertical structure. The resulting normalized point cloud can be used to compute area-based metrics and can also be rasterized into a Canopy Height Model (CHM). These intermediate LiDAR products are essential for empirical models of biomass estimation (ZHAO et al., 2018).

Figure 2.2 - Workflow for deriving canopy structural attributes from raw airborne discrete-return LiDAR data.



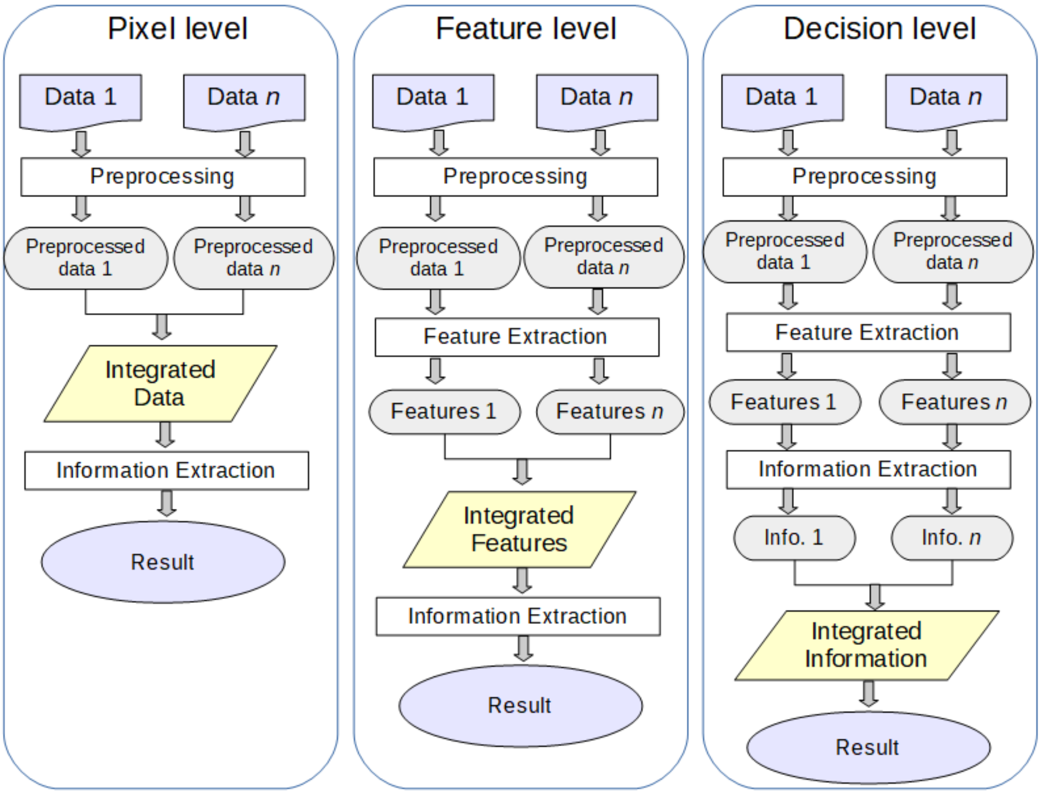
Source: Xiao et al. (2019).

2.3 Multisensor data integration

The increasing number of remote sensors is generating a massive volume of data with different spatial, spectral, radiometric, and temporal resolutions. Multisensor data integration (also called multisensor data fusion or combination) takes advantage of this increased data availability to produce more detailed information than each source can produce separately. Therefore, the purpose of multisensor data integration is to improve the quality of information for decision making (POHL; VAN GENDEREN, 1998; ZHANG, 2010).

Image integration can be divided into three processing levels according to the stage at which the fusion takes place (Figure 2.3): pixel, feature, and decision level (POHL; VAN GENDEREN, 1998). The pixel-level consists of combining the physical parameters recorded in two or more images to generate a synthetic image. This level is mainly applied to optical images (ZHANG, 2010) and requires only preprocessing steps, including radiometric and geometric corrections and resampling to a common pixel size. The advantage of this level of integration is the preservation of the original information. The disadvantage is the large data dimensionality, which makes processing slow. The integration of redundant information contained in pixels can be accomplished by various methods, which are commonly applied in the fusion of multispectral and panchromatic images to improve the spatial resolution. Some of these methods are those based on algebraic operations (multiplication, Brovey transform, high pass filter), spatial transformations (IHS, PCA, Gram-Schmidt), and pyramid decomposition (Wavelet transform) (ZHANG; YANG, 2012).

Figure 2.3 - Processing levels of image integration.



Source: Author's production.

In feature-level integration, each original data source is submitted to feature extraction/selection techniques in both the spatial domain (e.g., lines, intersections, texture) and/or spectral domain (e.g., indices, linear combinations, derivatives). The various features derived from the different data sources are later combined to replace the original data in information extraction procedures. Feature-level integration has the advantage of compressing the amount of data while retaining the relevant information of the variable or phenomenon of interest. Compared to pixel-level integration, feature-level enables higher processing speed (ZHANG; YANG, 2012).

In decision-level integration, each data source is processed independently for information extraction. The results obtained are then combined from decision rules to generate final information. An example of an application for decision-level integration is the classification of regions of interest (e.g., buildings, streets, tree species, and land cover) from two or more data sources. Each data source is used to generate one or more class maps, which are then combined through voting strategies in an attempt to reduce misclassification (MURA et al., 2015).

According to Zhang (2010), dividing the image integration approaches in three processing levels (pixel, feature, and decision) does not fit all the integration possibilities. In practice, it is common to proceed with the combination of different levels. Importantly, some data fusion methods (e.g., pan-sharpening) were developed for similar data types. Thus, using these techniques on data sources of different nature, such as active and passive sensors, may not provide the best results. Integrating data from different sensors require some caution from the user. For instance, large differences in spatial resolution, measured physical quantities, data acquisition geometry, light sources, and acquisition periods, can complicate the implementation and validation of the integration methods (TORABZADEH et al., 2014). For successful integration of data from multiple remote sensors, the quality of each data source must be ensured; the radiometric, atmospheric and geometric corrections must be well performed; and the choice of the fusion method and level should be appropriate to the data type and objectives of the analysis.

2.4 Prediction methods based on multisensor data

The use of multisensor data, supported by advanced methods of analysis, may overcome some of the problems faced with single datasets, improving the quality of the requested information (KOCH, 2010). Two major approaches have been used for information extraction based on remote sensing data: physical methods, based on the physical laws between the electromagnetic radiation and the characteristics of a target; and empirical methods, based on statistical relationships between remote sensing and reference data

(LIANG et al., 2012). Due to the simplicity of empirical methods, they have been widely used for multisensor integration than physical approaches (TORABZADEH et al., 2014). Empirical methods include classification techniques, used to predict categories, and regression models, used to predict continuous variables. Many algorithms have been developed for both classification and regression models, such as Linear Models (LM), Support Vector Machine (SVM), Stochastic Gradient Boosting (SGB), Random Forest (RF) and Cubist (CB). A brief overview of each model used in this thesis is introduced below.

LM are parametric methods developed for regression analysis, which account for linear relationships between response and predictors. LM (multivariable regression with ordinary least squares), associated with some technique of feature selection (e.g., stepwise), is the most common method applied to AGB estimation (LU et al., 2014). As a parametric technique, it requires assumptions such as linearity, residual normality, homoscedasticity, and independence (OSBORNE; WATERS, 2002). Furthermore, conventional LM may generate spurious results due to multicollinearity. Regularization methods such as ridge regression are valuable for addressing this issue, reducing the impact of redundant variables by shrinking their coefficients (DUZAN; SHARIFF, 2015).

SVM is a non-parametric machine learning technique widely used for classification purposes (MOUNTRAKIS et al., 2011). This method is also effective for regression tasks, being commonly referred to as SVR (Support Vector Regression) (BASAK et al., 2007). The main idea behind SVM is to transform a nonlinear problem into linear by mapping the input data into a high-dimensional feature space, using a kernel function. The radial basis function (RBF) kernel has been widely used, since it generally performed better than other kernels, such as linear and polynomial. The required parameters for the RBF-SVM are the *cost*, which controls the complexity of the boundary between support vectors, and the *sigma*, which is a smoothing parameter. The range of values for the sigma parameter can be estimated, for instance, using the *sigest* function from the R package *kernlab* (KARATZOGLOU et al., 2004). The SVM method has proven its robustness to dimensionality, outliers in the training data, and the generalization ability, for both classification and regression tasks (MONNET et al., 2011).

RF is an ensemble learning method that combines predictions of multiple Classification and Regression Trees (CART) (BREIMAN, 2001). Each tree is independently created from a bootstrap sample of the original data (a bagging approach). Moreover, each node of the tree is split using a specified number of randomly selected features (*mtry*). RF has become popular in remote sensing applications due to its promising predictive capabilities for high-dimensional datasets. Furthermore, RF is insensitive to multicollinearity, data noise, outliers, and overfitting (BELGIU; DRAGUT, 2016).

SGB uses a boosting ensemble method for combining predictions of several classification or regression trees. Simple trees are fitted sequentially using the loss function gradient from the prior tree to increase emphasis on observations modeled poorly. At each iteration, a random subsample of the training dataset (without replacement) is used as input (FRIEDMAN, 2002). Instead of developing single complex trees, relatively small trees are combined by averaging their weighted predictions. The SGB involves parameters for controlling the learning process: (i) the number of boosting iterations (*n.trees*); (ii) the number of nodes per tree (*interaction.depth*); (iii) the learning rate (*shrinkage*), which penalizes the importance of each consecutive iteration; and (iv) the minimum terminal node size (*n.minobsinnode*) (ELITH et al., 2008). Several advantages of the SGB algorithm have been highlighted, including its low sensitivity to outliers, great ability to deal with unbalanced training datasets, and its robustness in dealing with interaction among predictors (FRIEDMAN, 2002). Promising SGB results have been reported for remote sensing classification (CHIRICI et al., 2013; GODINHO et al., 2016; LAWRENCE et al., 2004) and regression purposes (CARREIRAS et al., 2012; FILIPPI et al., 2014; MANQI et al., 2014). For instance, Chirici et al. (2013) found a superiority of the SGB method over CART and RF for mapping forest fuel types using both LiDAR and multispectral data. SGB has also been used for estimating AGB with RADAR (CARREIRAS et al., 2012), LiDAR (MANQI et al., 2014), and HSI (FILIPPI et al., 2014).

CB is a rule-based tree model, which produces linear regression models instead of simple values in the terminal nodes of trees, based on the M5 model tree (RULEQUEST, 2018). In contrast to RF and SGB, CB does not retrieve one final model but a set of rules associated with sets of multivariable models. CB can also use a boosting-like scheme called committees, in which subsequent trees are created using adjusted versions to the training set outcome. Predictions from all the committees are averaged to produce the final prediction (JOHN et al., 2018). In addition, the predictions generated by the model rules can be adjusted using nearby points from the training set data (defined by the parameter *neighbors*). CB is a viable method for AGB estimation across different sites and scales (BLACKARD et al., 2008; JOHN et al., 2018; MANQI et al., 2014).

2.5 LiDAR and HSI data integration for forest applications

Recently, the use of multisensor data for forest applications has gained interest. This is especially true for the integration of passive and active remote sensors because of their complementary characteristics. The integration of LiDAR and HSI remote sensing technologies has been used in several forest applications, such as land cover mapping (ZHANG et al., 2016); tree species classification (DALPONTE et al., 2008; SOMMER et al., 2015); biomass modeling (CLARK et al., 2011; FASSNACHT et al., 2014); and estimation of biochemical and physiological properties (BROADBENT et al., 2014; THOMAS et al., 2008). Some examples of the main forest applications of LiDAR and HSI data integration will be described below.

2.5.1 Mapping land use/land cover and tree species

Most forest studies using LiDAR and HSI data integration have prioritized the mapping of land use/land cover or tree species (TORABZADEH et al., 2014). Information on land cover type and forest species composition are essential for the management and monitoring of natural resources. For instance, this information can be applied to mapping trees of economic (e.g., timber species) and ecological (e.g., invasive species or functional plant groups) importance. In addition, classification of land cover types and species may be useful as a preliminary step in quantifying biophysical, biochemical, and physiological characteristics by stratifying image attributes by species, species groups, or land cover types.

Individual tree species discrimination from remote sensing data requires high spatial and/or spectral resolution. Thus, hyperspectral sensors, especially airborne sensors that have high spatial resolution, have already been used for the classification of forest species, functional groups, or land cover types in tropical (CLARK et al., 2005), subtropical (YANG et al., 2009), and temperate (BOSCHETTI et al., 2007; PLOURDE et al., 2007) regions. However, the information provided by HSI is limited to the bi-dimensional plane, which restricts a characterization of the canopy vertical structure. This structural information may assist in distinguishing species with similar spectral behavior. Thus, the interest in integrating HSI with structural data provided by LiDAR for forest composition studies has grown over the last decade.

Table 2.2 summarizes the results of some studies that have evaluated the integration of HSI and LiDAR data for the classification of tree species or land cover. Most of them showed considerably greater accuracy in the classification performed with both data sources, respective to the classification using only HSI data. However, the results depend on several factors, such as sensor characteristics, data preprocessing, selected metrics for analysis, fusion level, classification method, number and type of classes, and characteristics of the studied vegetation. Some studies have considered different combinations of these factors to identify the best conditions for accurate classification of tree species and land cover. For instance, Ghosh et al. (2014) evaluated the effect of the HSI spatial resolution on tree species mapping, with and without the integration of LiDAR data. For this purpose, three HSI data sources were used: two airborne HyMap datasets with 4 m and 8 m resolutions, and one spaceborne Hyperion dataset with 30 m spatial resolution. The authors also tested different features and two classification models (SVM and RF). From the spatial resolution analysis, the 8 m resolution dataset generally produced the best results, probably because the images were not as heterogeneous as in the 4 m resolution, nor as homogeneous as in the 30 m resolution. From the classifiers, RF had better results than SVM when spectral indices were used as input data. SVM was more appropriate than RF when reflectance data were used. The HSI input data that produced the best classification accuracy were the MNF components. The integration of LiDAR height did not produce a significant improvement in the accuracy of tree species classification.

Table 2.2 - Examples of studies on forest species or land cover classification based on the HSI and LiDAR data integration.

Reference	**Classes**	**Classification model**	**HSI results**		**HSI+LiDAR results**	
			Accuracy	kappa	Accuracy	kappa
Dalponte et al. (2008)	19 tree species + 4 land cover	ML, SVM, kNN	-	0.88	-	0.89
Jones et al. (2010)	6 broadleaf and 5 conifer species	SVM	72.3	0.60	73.5	0.60
Naidoo et al. (2012)	8 savanna species	RF	80.3	0.76	87.7	0.84
Heinzel, Koch (2012)	Pine, spruce, oak, and beech	SVM	64.7	-	88.0	-
Dalponte et al. (2012)	7 species + non-forest	SVM, RF	74.1	0.66	83.0	0.77
	5 forest types + non-forest		79.3	0.72	91.7	0.88
	Coniferous + broadleaf + non-forest		95.8	0.94	96.3	0.94
	Forest + non-forest		98.8	0.98	99.6	0.99
Ghosh et al. (2014)	Beech, Douglas fir, oak, red oak, pine	SVM, RF	86.0	-	86.0	0.83
Sommer et al. (2015)	8 broadleaf and 5 conifer species	RF	77.0	-	91.4	0.89
Geerling et al. (2007)	5 floodplain vegetation	ML	74.4	0.63	80.6	0.71
	8 floodplain vegetation		57.8	0.51	63.5	0.57
Koetz et al. (2008)	9 land cover	SVM	69.2	0.65	75.4	0.72
Bigdeli et al. (2015)	15 land cover	SVM, kNN, ML, fuzzy-kNN, fuzzy- ML	-	-	95.3	0.93
Wang and Glennie (2015)	9 land cover	ML, SVM	85.8	0.82	92.6	0.91
Zhang et al. (2016)	11 vegetation types	SVM, kNN, RF	84.6	0.81	91.1	0.89

In addition to the HSI spatial resolution, the density of LiDAR points per m^2 may also interfere in the classification results of forest species. In this context, Dalponte et al. (2012) used two LiDAR acquisitions (low and high point density) to evaluate the effect of point density in the classification of hardwood and coniferous species. They also tested the effect of the spectral resolution of passive sensors, using a hyperspectral airborne sensor (Aisa Eagle) and a GeoEye-1 multispectral satellite sensor. The influence of the number and type of classes, which ranged from detailed (e.g., tree species) to broad vegetation (e.g., coniferous and broadleaf or forest and non-forest), was also evaluated. Finally, two classifiers (SVM and RF) were used to obtain the maps of species, vegetation types and land cover, by integrating LiDAR height metrics and selected spectral bands. The results showed that the hyperspectral data were more effective than the multispectral data, both integrated with LiDAR, for the classification of species and broad classes. Multispectral data considerably reduced the classification accuracy of tree species and forest types, while maintained good accuracy for the more generalized classes. LiDAR-derived height increased the classification accuracy when combined with both multispectral and hyperspectral data. High LiDAR point density provided more information for species-level classification than low point density. This greater level of information was related to the possibility of obtaining more metrics from higher density data, which increased the accuracy compared with the use of only maximum height. In this analysis, SVM was better than RF under all tested conditions.

The choice of classifier may impact the outcome. As noted by Ghosh et al. (2014), several classifiers may yield similar results depending on the input data. An alternative to improve classification accuracy is to use decision-level integration of the results provided by different classifiers. Based on this strategy, Bigdeli et al. (2015) proposed the decision-level integration of LiDAR and HSI data for mapping 15 land cover classes. They also compared the performance of two groups of classifiers using crisp (SVM, ML, and kNN) and fuzzy (kNN and ML) approaches. In crisp classification, each pixel is associated with a class, whereas in fuzzy classification each pixel is related to a degree of pertinence to the classes. These results showed that LiDAR and HSI data integration improved the classification accuracy compared to the use of single-data alone (LiDAR or HSI). Both crisp and fuzzy decision-level fusion schemes produced greater accuracy than any single classifier. However, the best accuracy was obtained from integrating the fuzzy classifiers.

Studies that applied the integration of LiDAR and HSI data for the classification of land cover or forest species were mostly conducted over temperate or boreal ecosystems. Further studies are needed to fill knowledge gaps in tropical forests, which have greater structural complexity and biodiversity.

2.5.2 Estimation of biophysical attributes: aboveground biomass (AGB)

Remote sensing can be used to estimate biophysical attributes, such as canopy height, leaf area index (LAI), fractional cover, and AGB. Empirical relationships between the property of interest and remote sensing metrics are commonly used for this purpose (TORABZADEH et al., 2014). However, the accuracy of the empirical models based on passive optical data is generally limited due to the saturation of some reflectance-derived metrics over high-density forests (FANG et al., 2012). Three-dimensional information obtained from LiDAR active sensors can improve canopy structure characterization and increase the accuracy of the estimates, especially over dense forests (MAN et al., 2014). Therefore, the integration of HSI and LiDAR data is currently being used to improve the biophysical attribute estimates provided by LiDAR, especially because of the capacity of HSI data to provide information on species composition, senescence, and stress (CLARK et al., 2011; SWATANTRAN et al., 2011). Biophysical attributes such as basal area (ANDERSON et al., 2008) and LAI (THOMAS et al., 2011) had benefited from multisensor integration. However, several studies using the combination of HSI and LiDAR data for biophysical attributes estimation have focused on AGB. The results of some of them are summarized below.

Different approaches can be used to integrate LiDAR with passive optical data (FENG et al., 2017). One approach is to use optical images for vegetation classification and then to establish LiDAR-based AGB models on the different vegetation types. Chen et al. (2012) used this approach with LiDAR and aerial photography data and observed improvements in the performance of the AGB model. Another alternative is to use both LiDAR and optical sensor metrics directly as predictors in AGB models. Several studies have used this approach with improved results in biomass estimation (ANDERSON et al., 2008; VAGLIO LAURIN et al., 2014; LUO et al., 2017a). For instance, Vaglio Laurin et al. (2014) found that the combined use of LiDAR metrics with HSI reflectance bands improved the AGB estimates (R^2 = 0.70) when compared to the sole use of the LiDAR height metrics (R^2 = 0.64). Luo et al. (2017a) used Partial Least Square Regression (PLSR) models to estimate aboveground, belowground and total biomass in northwest China from LiDAR metrics and HSI vegetation indices. The results showed that LiDAR data had greater biomass prediction power when compared to vegetation indices. However, compared to single-LiDAR estimates, the combination of both data sources improved the biomass estimates, producing gains of 5.8%, 2.2%, and 2.6% for belowground, aboveground, and total biomass, respectively. Anderson et al. (2008) combined LVIS LiDAR data with AVIRIS hyperspectral data to estimate basal area, stem diameter, and AGB in an experimental mixed temperate forest area. The results showed that the data fusion improved the estimation of the three variables compared to models that used only LiDAR or only HSI. The improvements in R^2 ranged between 8 and 25%, while the decrease in the estimation error varied between 5 and 25%.

Other AGB studies showed that the addition of hyperspectral data did not significantly improve the LiDAR estimates (CLARK et al., 2011; FASSNACHT et al., 2014; LATIFI et al., 2012). To evaluate the improvement in AGB and stem density models, Latifi et al. (2012) extracted various height and intensity metrics from LiDAR data. They combined LiDAR data with four types of hyperspectral metrics: 125 HyMap bands, the first 25 PCA components, the first 25 MNF components, and six spectral indices. These features were submitted to an evolutionary genetic algorithm for the selection of the most parsimonious variables. To estimate the biophysical attributes of interest (AGB and stem density), LiDAR height metrics were more effective than the HSI metrics. Only a few HyMap metrics contributed to the quantification of these attributes.

Several factors may explain the difference in multisensor integration performance, such as the different modeling methods, sensor specifications, vegetation type of the study area, and field data sampling. Fassnacht et al. (2014) evaluated the performance of different AGB estimation methods (stepwise linear regression, SVM, RF, Gaussian and kNN processes). They tested for each method the influence of the data type (HSI, LiDAR or both) and field sample size in two study areas located in Chile and Germany.

The authors observed that the most important factor for improving the AGB estimation was the data type, with LiDAR being superior to HSI. Unlike other studies, the integration of LiDAR with HSI data did not improve model performance. In general, the prediction method was more important than the sample size. The RF method displayed the best AGB performance. The results suggest that the appropriate choice of the prediction method may be more effective for increasing performance than obtaining more field samples. However, similar to land cover classification studies, more research is necessary to confirm their findings over other study areas such as the complex tropical ecosystems of the Amazon.

2.5.3 Estimation of biochemical and physiological attributes

The leaf content of biochemical constituents, such as nitrogen, water, and photosynthetic pigments, provides an indicator of the physiological state of the vegetation. This information can be used to monitor the spatial and temporal dynamics of nutrient cycling, including the vegetation stress to a particular limiting condition or the photosynthetic ability of the local vegetation. The different biochemical constituents of plants selectively interact with radiation, absorbing energy more intensely at some specific wavelengths. Thus, remote sensing data with high spectral resolution allow the identification of absorption bands that can be used to estimate the presence and concentration of biochemical constituents (NIU; YAN, 2012).

Most studies on the biochemical properties of vegetation from spectral features were performed at the leaf level or over small canopies under controlled laboratory conditions (BLACKBURN, 2002). Few studies have examined the HSI applicability in estimating biochemical constituents at the canopy level over complex forests (CURRAN et al., 1997, SMITH et al., 2003). One difficulty in obtaining remote biochemical estimates at the canopy/landscape levels is the interference of canopy structure on the spectral measurements. Canopy structure can mask the biochemical spectral features, preventing their detection. Also, differences in viewing-illumination geometry produce different degrees of shadows for the sensors, affecting the remote detection of canopy composition (ASNER; MARTIN, 2008).

Empirical models, radiative transfer models, and reflectance inversion models have been used to describe the effects of viewing-illumination, atmospheric transmissivity, and canopy architecture on the reflectance spectra and canopy biochemical attributes detected by the sensors (ASNER et al., 1998; ASNER; MARTIN, 2008; BROGE, LEBLANC, 2000; ZARCO-TEJADA et al. 2001). Asner and Martin (2008) evaluated the relationship between canopy reflectance and biochemical properties by simulating different conditions of canopy structure and viewing-illumination geometry. The study was supported by field spectral and biochemical data acquired over 162 forest tropical species in Australia. First, the authors evaluated the relationship between biochemical and spectral data at leaf-level using PLSR analysis. They found a good correlation (0.79-0.91) between the leaf reflectance and biochemical constituents such as chlorophyll, carotenoid, and water content. Using field data as a reference, the spectral reflectance curve of the tree species was then simulated with geometric-optical radiative transfer models. The progressive increase in canopy structural characteristics in simulated crowns had little effect on pigment and water estimation but affected the prediction of nitrogen and phosphorus content. The two factors that most negatively affected the prediction of biochemical constituents were the LAI and viewing-illumination geometry. To circumvent these limitations, the authors suggested integrating HSI with LiDAR data to filter out high LAI pixels containing only the illuminated fraction of the canopy. In practice, this strategy reduces the negative influence of shadows on the estimates.

Thus, the combination of HSI and LiDAR data can assist in the pixel sampling strategy to look for portions of the canopies where the variation in the vegetation structure and viewing-illumination geometry affects less the prediction of biochemical parameters. In this approach, the contribution of LiDAR data to biochemical modeling is indirect, because only HSI data are used as independent variables in regression models. Because the vegetation structural information is related, to some extent, to biochemical information, another strategy is to use LiDAR metrics directly in the prediction models (THOMAS et al., 2008). Unfortunately, despite the potential of using both HSI and LiDAR data for the estimation of biochemical and physiological parameters, only a few studies of this nature have been conducted to date.

Thomas et al. (2008) extracted height metrics from LiDAR data and spectral indices from HSI data to predict chlorophyll and carotenoid concentrations in Canadian boreal forests. The authors integrated the multisource metrics by dividing the LiDAR height metric per the derivative chlorophyll index. The integrated data were more efficient in the estimation of photosynthetic pigments compared to the exclusive use of LiDAR or HSI metrics, producing an R^2 value higher than 0.90. The mapping of total chlorophyll concentration in the study area revealed a spatial pattern indicative of the composition of different tree species.

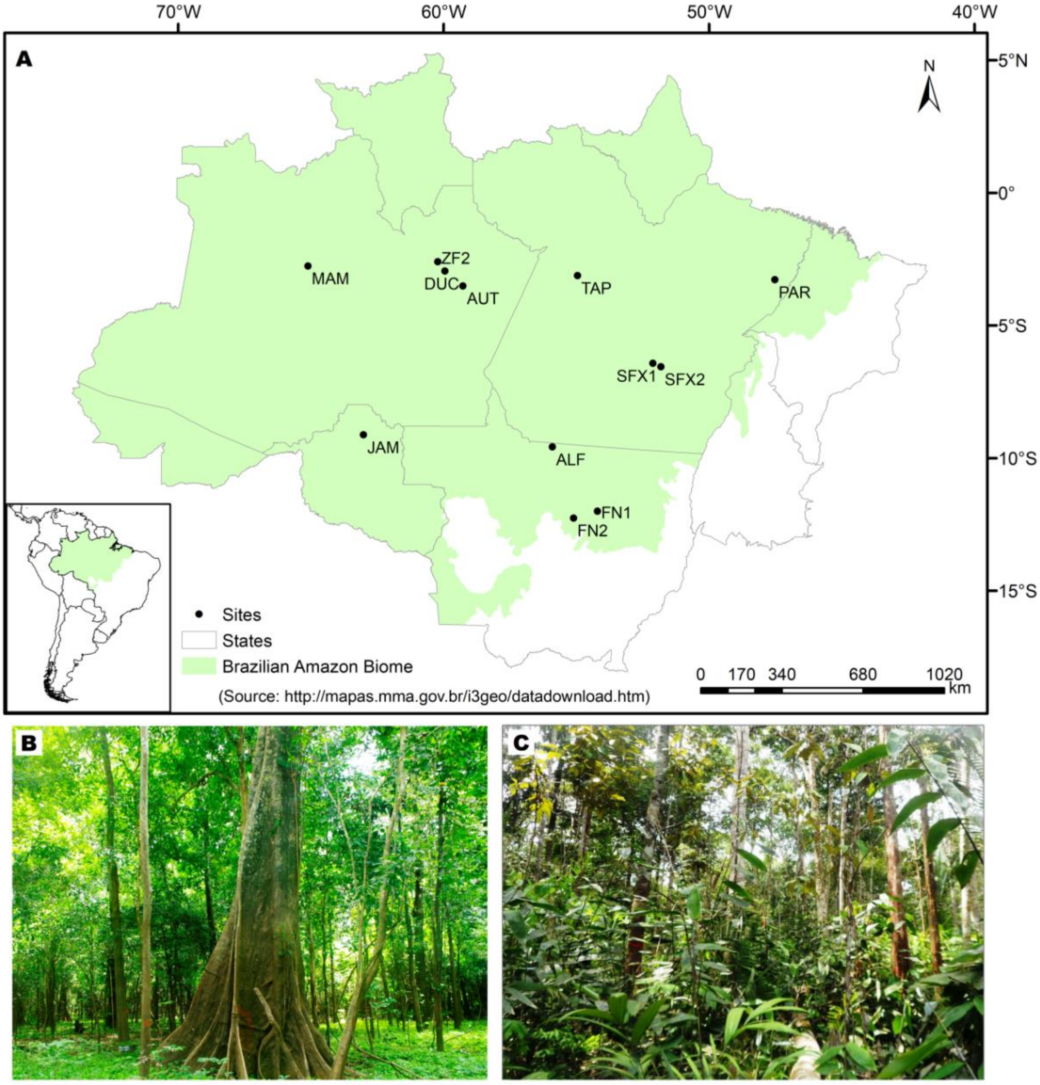
Other studies used HSI data as independent variables in regression models after applying masks derived from LiDAR height (BLACKBURN, 2002) or a combination of LiDAR height and NDVI thresholds (ASNER et al., 2015). The masks isolated canopy pixels, removing areas of clearings, shadows, water, and exposed soils from the analysis. The results found by Blackburn (2002) showed that the HSI metrics, without the application of the mask, were not related to the photosynthetic pigment concentration when considering the data set containing both coniferous and hardwood species. However, considering only the hardwoods, a relationship was observed between the position of the red-edge wavelength and the pigment concentration. The use of the LiDAR-derived mask did not produce significant improvements in the estimation of pigment concentration in hardwood species. However, it allowed the estimation of pigment concentration per unit of leaf mass for conifer species.

Asner et al. (2015) also evaluated other biochemical constituents besides photosynthetic pigments. The use of the mask before modeling produced the best performance for estimating photosynthetic pigments, nitrogen, phosphorus, iron, and carbon, with R^2 values ranging from 0.39 to 0.58. The lower performance of the estimates found in this study area may be related to the greater complexity of tropical forests compared to tree plantations or boreal forests. Higher biodiversity and structural complexity make the field sampling a more difficult process, especially considering the sampling restriction to the solar illuminated foliage. Although further studies are needed to assess the effectiveness of these masking strategies, the results indicate the potential use of HSI and LiDAR data for biochemical estimates, even over highly complex forests.

3 STUDY AREA AND REMOTE SENSING DATA ACQUISITION

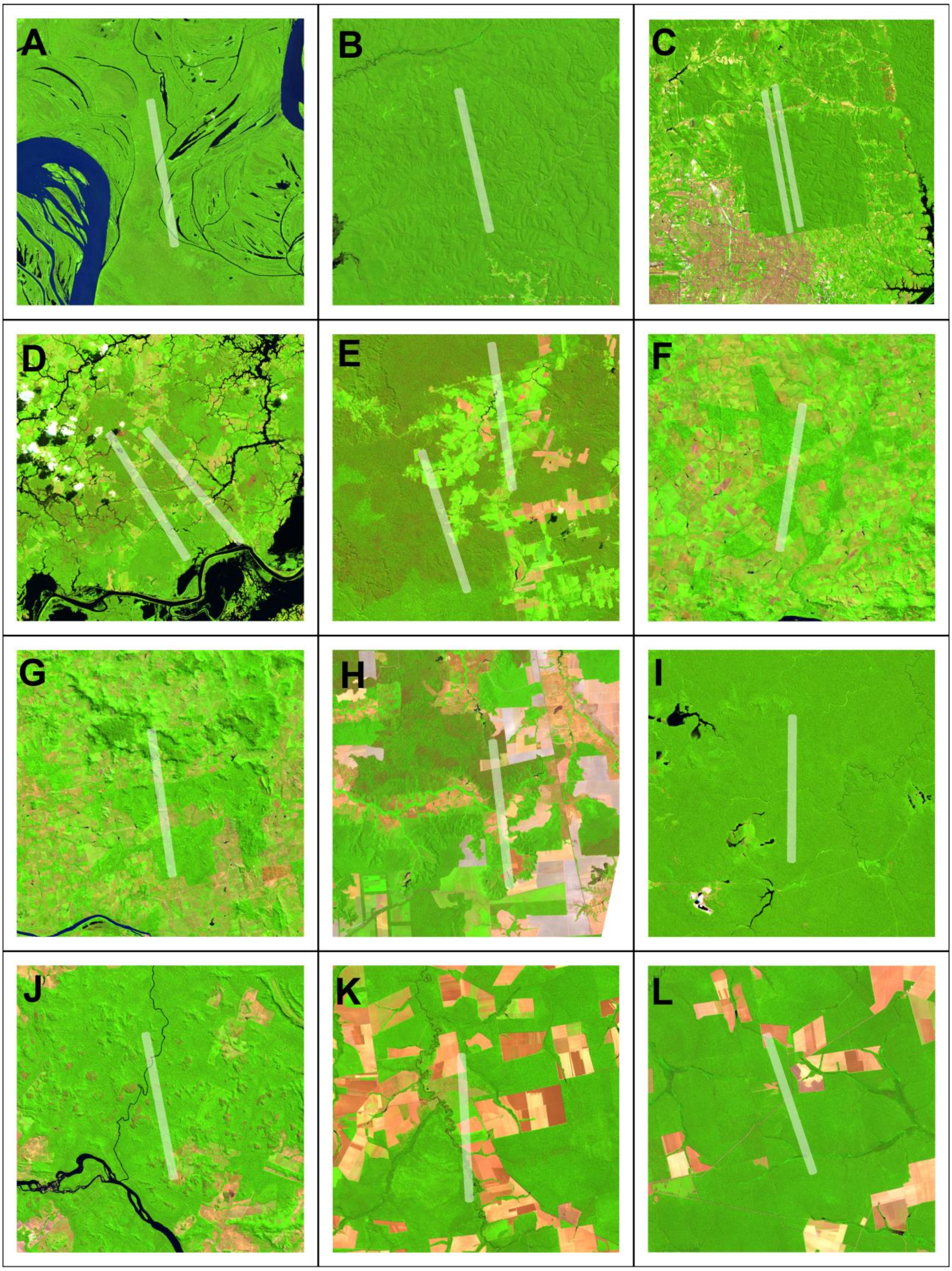
This study was conducted at 12 sites across the Brazilian Amazon biome, distributed in the states of Amazonas, Pará, Rondônia, and Mato Grosso (Figure 3.1). At each site, airborne LiDAR and HSI data were collected in transects of approximately 12.5 x 0.3 km. Most sites were represented by a single transect, while the sites AUT, DUC, and TAP were covered by two transects each (Figure 3.2). In this chapter, the characteristics of the study area and the LiDAR and HSI data will be described.

Figure 3.1 - (A) Distribution of the studied sites in the Brazilian Amazon Biome. Examples of sampled forests are shown in (B) for a seasonally flooded undisturbed mature forest (MAM site) and in (C) for a *terra firme* forest degraded by understory fire (AUT site).



Source: Author's production.

Figure 3.2 - Landscape (area of 25 x 25 km) around each site represented by the HSI and LiDAR superposed flight lines (transects): (A) MAM, (B) ZF2, (C) DUC, (D) AUT, (E) TAP, (F) SFX1, (G) SFX2, (H) PAR, (I) JAM, (J) ALF, (K) FN1, and (L) FN2. The images are OLI/Landsat-8 color composites with bands 6 (red), 5 (green) and 4 (blue), from 2016-2017.



Source: Author's production.

3.1 Study area characterization

The study sites encompass a wide variety of anthropogenic, climatic, geological, and edaphic conditions. Regarding the forest disturbance conditions, the MAM site, located within the Mamirauá Sustainable Development Reserve (a conservation unit of the Amazonas state), is covered by undisturbed flooded forests (Figure 3.2A). The forests of both ZF2 and DUC sites, located in Manaus (AM), are predominantly undisturbed (Figure 3.2B and 3.2C). The DUC site is mainly located within the Adolpho Ducke reserve, an area of 100 km^2 established in 1962. However, the site also encompasses some secondary forests, mostly in old succession stages, situated at the northern borders of the reserve. In the AUT site (Figure 3.2D), located in Autazes (AM), some small areas were cleared close to highways and rivers, becoming mostly secondary successions at the early stages of vegetation regrowth. The area also has a history of forest fires under the effect of the El Niño Southern Oscillation (ENSO) in 1998/99, 2010 and 2015/16. Few relatively undisturbed areas are found at this site. Similarly, the TAP site, in the municipality of Belterra (PA), accounts for a few undisturbed areas (Figure 3.2E). The site is located between the boundaries of the FLONA Tapajós, covering the community of São Jorge, which contains some secondary forests. Some areas in the FLONA were submitted to selective logging. Besides, extensive fires also affected the area, especially in 2015/16. The SFX1 and SFX2 sites (Figure 3.2F and 3.2G) are located in the São Félix do Xingu municipality (PA) and are mainly composed of forest fragments degraded by recurrent understory fires. The sites also encompass few early secondary successions. The PAR site (Figure 3.2H), located in Paragominas (PA), comprises very degraded forests by conventional logging operations followed by large fire events in 1992, 1998, 2006, and 2016. Some areas of secondary forests are also observed at this site. The JAM site (Figure 3.2I), located in the Itapuã do Oeste municipality (RO) at the FLONA Jamari, is a conservation unit of sustainable use, where reduced-impact logging is authorized by forest concession (SFB, 2020). The ALF site (Figure 3.2J) is located between the Alta Floresta and Novo Mundo municipalities, in the state of Mato Grosso (MT). The site is mainly covered by mature forests, either undisturbed or disturbed by understory fires and fragmentation, with the occurrence of few secondary forests. The sites FN1 (Figure 3.2K) and FN2 (Figure 3.2L) are located in the Feliz Natal municipality (MT), a transitional region between ombrophilous and seasonal forests. The region is composed of large deforested areas. Except for some gallery forests, the remaining mature forests over the site are degraded by conventional logging and/or major fire events.

The climate of the Brazilian Amazon biome is classified as type A (equatorial) according to the Köppen-Geiger classification (KOTTEK et al., 2006). Rainfall gradient ranges from wetter conditions on the MAM, ZF2, DUC, and AUT sites (rainy equatorial Af climate) to drier conditions on the PAR, SFX1, SFX2, ALF, FN1, and FN2 sites (dry and wet tropical Aw climate). The climate of sites TAP and JAM is classified as Am (tropical monsoon), presenting intermediate conditions between the rest of the sites. Overall, prolonged dry seasons (three to five months) are usually observed toward the eastern Amazon. The long-term (1973-2013) annual rainfall reported for the Brazilian Legal Amazon (BLA) is approximately 2100 mm (ALMEIDA et al., 2017). In the studied sites, annual rainfall ranges from 1800 mm at the FN2 site to more than 3000 mm at the MAM site. The mean annual temperature over the BLA is 26.5 ºC (ALMEIDA et al., 2017), varying in the study sites from 24.6 ºC at SFX2 to 27.0 ºC at AUT.

Concerning geological and edaphic conditions, the Amazon is commonly classified into regions with similar substrate origin and soil fertility (QUESADA et al., 2011). From the sites used in this study, MAM, ZF2, DUC, AUT, and TAP are part of the so-called Central Amazonia region, comprised of old sedimentary substrates and low soil fertility. On the other hand, the sites PAR, SFX1, SFX2, JAM, ALF, FN1, and FN2 are located over the Brazilian Shield composed of pre-Cambrian rocks with related high fertility soils. The predominant soil types are Acrisols and Ferralsols, with Gleysols occurring in the seasonal floodplain of the MAM site (QUESADA et al., 2011). From a topographic point of view, all sites are considered as lowlands having altitudes lower than 500 m. The AUT and MAM sites present the lowest altitude (< 50 m), while the southeastern sites (SFX1, SFX2, ALF, FN1, and FN2) show the highest values ranging from 200 to 500 m (Table 3.1).

Table 3.1 - Description of the study sites.

Brazilian state	**Site**	**Latitude**	**Longitude**	**Altitude**	**MAT**	**MAP**	**CWD**	**Forest type**	**Forest status**
		(º)	(º)	(m)	(ºC)	(mm.yr^{-1})	(mm.yr^{-1})		
Amazonas (AM)	MAM	-2.76	-65.10	36.7	26.7	3406	0.0	SFO	UF
	ZF2	-2.60	-60.21	61.6	26.4	2356	-60.7	TFO	UF
	DUC	-2.95	-59.94	86.2	26.5	2308	-127.3	TFO	UF, SF
	AUT	-3.51	-59.26	25.7	27.0	2293	-109.2	TFO	UF, DF, SF
Pará (PA)	TAP	-3.12	-54.95	123.0	25.8	1848	-317.0	TFO	UF, DF, SF
	SFX1	-6.43	-52.11	205.3	24.9	1981	-213.4	TFO	DF, SF
	SFX2	-6.56	-51.81	289.1	24.6	1964	-208.5	TFO	DF, SF
	PAR	-3.28	-47.52	128.8	25.9	1915	-512.8	TFO	DF, SF
Rondônia (RO)	JAM	-9.12	-63.01	93.6	25.2	2388	-231.7	TFO	UF, DF
Mato Grosso (MT)	ALF	-9.58	-55.90	254.0	26.6	2216	-328.5	TFO	UF, DF, SF
	FN1	-12.00	-54.20	320.0	24.7	1815	-454.6	TFT	DF, SF
	FN2	-12.26	-55.10	338.7	24.7	1807	-446.4	TFT	UF, DF, SF

Altitude are from LiDAR-based DTM. MAT (Mean Annual Temperature) and MAP (Mean Annual Precipitation) are from WorldClim version 2 (FICK; HIJMANS, 2017). CWD (Climatic Water Deficit) is based on Chave et al. (2014). Abbreviations: SFO, Seasonally Flooded Ombrophilous forest; TFO, *Terra Firme* (unflooded) Ombrophilous forest; TFT, *Terra Firme* (unflooded) Transitional forest (ecotone between ombrophilous and seasonal forests); UF, Undisturbed Forest; DF, Disturbed Mature Forest (submitted to fragmentation, fire, or selective logging); SF, Secondary Forest.

3.2 Airborne LiDAR data

Airborne discrete-return LiDAR data were acquired between January 2016 and April 2017 using the Trimble HARRIER 68i system at an average height of 600 m above ground and a scan angle of 45°. The LiDAR sensor recorded multiple returns with a minimum point density of four points.m^{-2} and a small footprint of approximately 30 cm. The horizontal accuracy varied among sites from 0.035 m to 0.185 m, while the vertical accuracy ranged from 0.07 m to 0.33 m. The raw point cloud of each site was preprocessed by first identifying and removing isolated noisy points with the *lasnoise* function, from the LASTools software (ISENBURG, 2018). The parameters *step_xy*, *step_z* and *isolated* were set to 10, 5 and 5, respectively. Ground points were filtered (*GroundFilter* function with cellsize of 10, tolerance of 0.05 and 10 iterations) and then interpolated (*TINSurfaceCreate* function) into a digital terrain model (DTM) with a 1 m spatial resolution, using the FUSION/LDV software (MCGAUGHEY, 2014). To obtain the height above ground of each point, the DTM was subtracted from point elevations (function *Clipdata*, FUSION/LDV). The normalized point clouds were clipped according to the spatial extent of samples (function *PolyClipData*, FUSION/LDV) to further calculate the LiDAR metrics at the plot level.

Several LiDAR metrics have been proposed as potential predictors of canopy structural attributes such as AGB (LU et al., 2014; ZHANG Z. et al., 2017). Here, we tested a variety of area-based LiDAR metrics (Table 3.2) related to height distribution (height statistics such as mean, standard deviation, and percentiles), canopy cover (proportion of returns and Leaf Area Density), structural complexity (Shannon and Simpson diversity indices), and topography (terrain roughness).

Table 3.2 - Metrics calculated from LiDAR data.

Metrics	**Description**
Height	
H.max	Maximum height (m).
H.mean	Mean height (m) of first returns above 2 m.
H.pX	Xth (05, 10, 20, 25, 30, 40, 50, 60, 70, 75, 80, 90, or 95th) percentile of height distribution of first returns above 2 m.
H.sd	Height standard deviation (m) of first returns above 2 m.
H.cv	Height coefficient of variation (%) of first returns above 2 m.
H.skew	Skewness of height distribution of first returns above 2 m.
H.kurt	Kurtosis of height distribution of first returns above 2 m.
Canopy cover	
PD_{a_b}	Number of first returns between a height interval a_b (2_10, 10_20, or 20_30) divided by the number of all first returns.
PD_h	Number of first returns above a height h (2, 6, 10, 14, 18, 22, 26, or 30) divided by the number of all first returns.
PD_{1st}	Number of first returns above 2m divided by the number of all returns above 2m.
LAD_{a_b}	Leaf Area Density (m^2 m^{-3}) between the height interval a_b (2_10, 10_20, or 20_30).
LAD_h	Leaf Area Density (m^2 m^{-3}) above the height h (2, 6, 10, 14, 18, 22, 26, or 30).
Structural complexity	
HSCI	Shannon Structural Complexity Index, calculated from the LAD profile.
DSCI	Simpson Structural Complexity Index, calculated from the LAD profile.
Topography	
Roughness	Mean terrain roughness from a 10-m DTM.

Height metrics were calculated from the first returns that were considered to belong to the tree canopy, i.e., points above a 2-m height (NÆSSET; GOBAKKEN, 2008). We used only the first returns because they are more related to canopy surface structure (THOMAS et al., 2006) and are more stable across different LiDAR acquisition settings, such as the point density (SINGH et al., 2016) and flying altitude (NÆSSET, 2009).

Two types of canopy cover-related metrics were calculated. The first consists of point densities (PD) at different height intervals (e.g., the proportion of returns above 2 m or between 2 and 10 m) or for different return types (PD_{1st}, the proportion of first returns related to all returns). The second is based on the Leaf Area Density (LAD) profile, which corrects the LiDAR point density from occlusion effects (BOUVIER et al., 2015). The LAD profile was calculated with the LAD function of the *lidR* package (ROUSSEL; AUTY, 2018), with a height bin of 2 m and an extinction coefficient k of 0.695. The constant k was based on the study by Stark et al. (2012) in central Amazon. Canopy cover-related metrics were also derived using just the first returns, except the PD_{1st} metric, which also considered the number of all canopy returns in its formulation.

Metrics related to canopy structural complexity are based on two indices commonly used to describe species diversity in biological systems: the Shannon (H') and Simpson (D) indices (MAGURRAN, 2004). These diversity indices combine richness (number of species) and evenness (species abundance distribution) into a single measure. When applied to LiDAR data, they operate as a measure of vertical structural diversity, increasing with the vertical extent of the canopy and with a more equal distribution of point density or leaf area density across the profile (STARK et al., 2012). While the Shannon index is more strongly influenced by richness (in that case, canopy height), the Simpson index gives more weight to evenness (i.e. the homogeneity of canopy area profiles). Therefore, we also tested this approach for measuring structural complexity. The HSCI (Equation 3.1) and DSCI (Equation 3.2) indices used here are equivalent to the Shannon and Simpson indices, respectively. However, they were normalized by a fixed number of height bins to have a scale between 0 and 1:

$$HSCI = \frac{-\sum_{i=1}^{HB}[p_i * \ln(p_i)]}{\ln(HB)} \tag{3.1}$$

$$DSCI = \frac{1}{\sum_{i=1}^{HB}({p_i}^2) * HB} \tag{3.2}$$

where $p_i = LAD_i/\sum LAD_i$, i.e. the proportion of LAD in height bin i; and HB is the maximum number of height bins. In this study, HB was equal to 30, because we used 2 m bins between 0 and 60 m (maximum canopy height across the field plots).

Finally, we calculated the terrain roughness for characterizing the local topographic variability. Roughness was defined as the difference between the highest and lowest altitude in a 3 × 3 moving window (WILSON et al., 2007). To avoid extreme localized roughness values, we averaged the 1-m DTM to obtain a 10-m DTM, which served as input data in the analysis.

3.3 Airborne HSI data

Airborne hyperspectral data were collected between September and October 2017 using the AISAFenix sensor (Specim, Spectral Imaging, Ltd.) at an average height of 800 m above ground. To reduce variations in viewing-illumination geometry, we oriented the flight lines simultaneously close to the N-S direction. In addition, the HSI data were preferentially collected over sunny days between 10 a.m. and 1 p.m. (local time). The mean solar zenith angle (SZA) during data acquisition was 30º with a standard deviation of 7º. The at-sensor radiance was measured in 361 bands in the spectral range of 380-2500 nm, where 87 bands were located in the VNIR (visible and near-infrared) region and 274 bands in the SWIR (shortwave infrared). Bandwidth ranged from 5.7 nm (SWIR) to 6.8 nm (VNIR). The spatial resolution was 1 m. Due to noise, we removed bands outside the range of 460-2330 nm and around the two major spectral intervals of atmospheric water vapor absorption (1400 and 1900 nm), reducing the number of bands to 232. We used the Atmospheric/Topographic Correction for Airborne Imagery tool (ATCOR-4; version 6.3) to convert the radiance images into atmospherically-corrected surface reflectance data. Water vapor estimates were based on the 940-nm absorption feature. Data provided by a GPS onboard the aircraft were used for geometric correction of the scenes.

In addition to the 232 reflectance bands (R_λ, which λ is the wavelength band center in nm), we calculated several metrics from the HSI data: 30 vegetation indices (Table 3.3), 20 continuum-removal absorption parameters, and 6 sub-pixel metrics based on the linear spectral mixture analysis (SMA). These metrics explored the potential information associated with vegetation properties at the main spectral regions: visible region (460-690 nm), mainly associated with pigments; red-edge interval (690-760 nm), sensitive to changes in chlorophyll; near-infrared (NIR: 760-1300 nm), expressing scattering of radiation by canopy constituents and having absorption bands due to leaf water at selected wavelengths (980 and 1200 nm); and SWIR (1500-2330 nm), having absorption bands due to lignin-cellulose and nitrogen.

Table 3.3 - Vegetation indices calculated from the AISAFenix reflectance data.

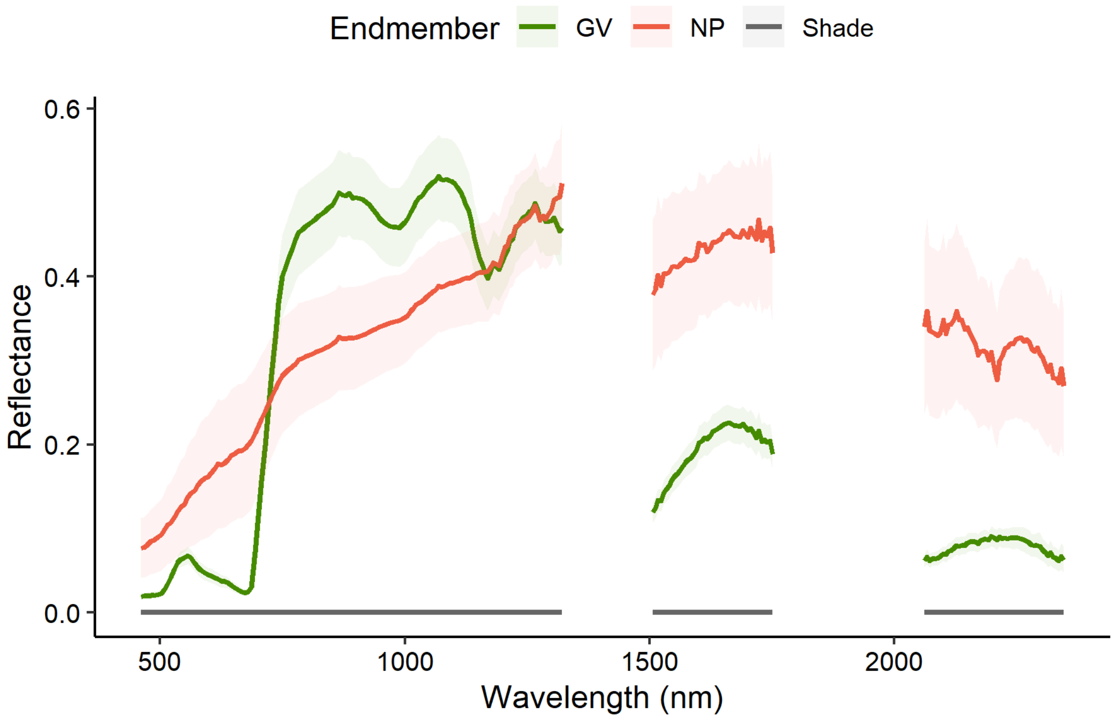
Abbr.	**Vegetation Index**	**Equation**	**Reference**
ARI1	Anthocyanin Reflectance Index 1	$(1/R_{549}) - (1/R_{701})$	Gitelson et al. (2006)
ARI2	Anthocyanin Reflectance Index 2	$[(1/R_{549}) - (1/R_{701})] * R_{797}$	Gitelson et al. (2006)
CAI	Cellulose Absorption Index	$0.5\ (R_{2039} + R_{2199}) - R_{2100}$	Nagler (2000)
CRI1	Carotenoid Reflectance Index 1	$(1/R_{515}) - (1/R_{549})$	Gitelson et al. (2006)
CRI2	Carotenoid Reflectance Index 2	$(1/R_{515}) - (1/R_{701})$	Gitelson et al. (2006)
D_{LAI}	Difference for Leaf Area Index	$R_{1724} - R_{969}$	le Maire et al. (2008)
DWSI1	Disease Water Stress Index 1	R_{797}/R_{1662}	Apan et al. (2004)
DWSI2	Disease Water Stress Index 2	R_{1662}/R_{549}	Apan et al. (2004)
DWSI3	Disease Water Stress Index 3	R_{1662}/R_{680}	Apan et al. (2004)
DWSI4	Disease Water Stress Index 4	R_{549}/R_{680}	Apan et al. (2004)
DWSI5	Disease Water Stress Index 5	$(R_{797} + R_{549})/(R_{1662} + R_{680})$	Apan et al. (2004)
EVI	Enhanced Vegetation Index	$2.5\ (R_{797} - R_{673})/(R_{797} + 6\ R_{673} - 7.5\ R_{474} +1)$	Huete et al. (2002)
GNDVI	Green Normalized Difference Vegetation Index	$(R_{797} - R_{549})/(R_{797} + R_{549})$	Gitelson et al. (1996)
LWVI1	Leaf Water Vegetation Index 1	$(R_{1096} - R_{983})/(R_{1096} + R_{983})$	Galvão et al. (2005)
LWVI2	Leaf Water Vegetation Index 2	$(R_{1096} - R_{1204})/(R_{1096} + R_{1204})$	Galvão et al. (2005)
ND_{Bleaf}	Normalized Difference for Leaf Biomass	$(R_{2160} - R_{1540})/(R_{2160} + R_{1540})$	le Maire et al. (2008)
ND_{chl}	Normalized Difference for Leaf Chlorophyll	$(R_{927} - R_{708})/(R_{927} + R_{708})$	le Maire et al. (2008)
NDLI	Normalized Difference Lignin Index	$[\log(1/R_{1751}) - \log(1/R_{1679})]/[\log(1/R_{1751}) + \log(1/R_{1679})]$	Serrano et al. (2002)
NDNI	Normalized Difference Nitrogen Index	$[\log(1/R_{1512}) - \log(1/R_{1679})]/[\log(1/R_{1512}) + \log(1/R_{1679})]$	Serrano et al. (2002)
NDVI	Normalized Difference Vegetation Index	$(R_{797} - R_{680})/(R_{797} + R_{680})$	Rouse et al. (1973)
NDWI	Normalized Difference Water Index	$(R_{859} - R_{1237})/(R_{859} + R_{1237})$	Gao (1996)
PRI	Photochemical Reflectance Index	$(R_{529} - R_{570})/(R_{529} + R_{570})$	Gamon et al. (1992)
PSRI	Plant Senescence Reflectance Index	$(R_{680} - R_{502})/R_{749}$	Merzlyak et al. (1999)
PWI	Plant Water Index	R_{900}/R_{969}	Peñuelas et al. (1997)
REP	Red-Edge Position	$700 + 40\ [(R_{re} - R_{701})/(R_{742} - R_{701})]$	Guyot; Baret (1988)
		$R_{re} = (R_{673} + R_{783})/2$	
RVSI	Red-Edge Vegetation Stress Index	$[(R_{714} + R_{749})/2] - R_{735}$	Merton (1998)
SR	Simple Ratio	R_{797}/R_{680}	Jordan (1969)
VI_{green}	Vegetation Index green	$(R_{549} - R_{680})/(R_{549} + R_{680})$	Gitelson et al. (2002)
VOG1	Vogelmann Index 1	R_{742}/R_{721}	Vogelmann et al. (1993)
VOG2	Vogelmann Index 2	$(R_{735} - R_{749})/(R_{714} + R_{728})$	Vogelmann et al. (1993)

Five continuum-removal absorption bands were defined from fixed wavelength edges: 461-536 nm (495-nm band), 556-749 nm (670-nm band), 893-1074 nm (980-nm band), 1097-1265 nm (1200-nm band), and 2039-2199 nm (2100-nm band). The continuum-removed spectrum was calculated by dividing the reflectance values within the absorption band by the corresponding values of a continuum line established between the edges (CLARK; ROUSH, 1984). To reduce noise in the original reflectance, the spectra were firstly smoothed using a Savitzky-Golay filter with a window size of five bands and a first polynomial order. The continuum-removed absorption bands were characterized by the depth (D_c) at the absorption center (c), the width at half depth (W_c), the band area (A_c, the sum of depths along the band), and the asymmetry (As_c, the ratio of the area left to area right of the band center) (KOKALY et al., 2009).

The fractional abundance of the green vegetation (GV), shade, and non-photosynthetic vegetation/soil (NP) endmembers were calculated using the *unmix* function from the R package *hsdar* (LEHNERT et al., 2018). To select endmembers for GV and NP, we applied sequentially the minimum noise fraction (MNF) and the pixel purity index (PPI) techniques using the Environment for Visualizing Images (ENVI; Harris Geospatial Solutions, Inc). Candidate endmembers detected by the PPI were projected over an n-dimensional scatterplot for finding the purest pixels at each site. The final GV and NP endmembers were then obtained by averaging the purest pixels of all sites (Figure 3.3). For the shade endmember, we considered a photometric shade with a uniform reflectance of zero (CLARK et al., 2011). The endmember spectrum of NP represents a mixture of bright soils and non-photosynthetic vegetation since these scene components could not be distinguished from each other in the images.

All HSI metrics were first obtained on a pixel-basis and then converted to the plot-level by calculating the average of all pixels values within the sample plot. Because shade generally relates to canopy structure, we also calculated the proportion of pixels with shade fraction below 30% (S_{0_30}), between 30 and 60% (S_{30_60}), and above 60% (S_{60}).

Figure 3.3 - Endmembers (GV= green vegetation, NP= non-photosynthetic vegetation/soil, and shade) spectra used in the spectral mixture analysis (thick lines). The colored area around the lines represents the standard deviation of the sites.

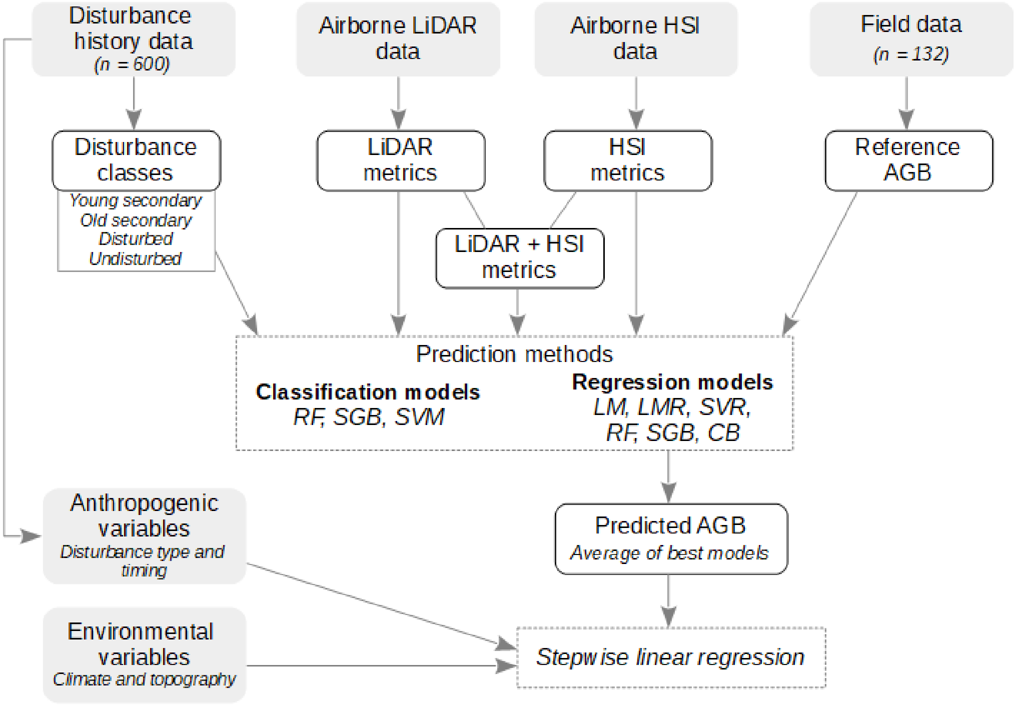


Source: Author's production.

4 GENERAL METHODOLOGY

Figure 4.1 summarizes the main methodological steps used in this thesis. In chapters 5 and 6, LiDAR and HSI metrics were used as predictors for modeling forest ecosystem properties. However, in chapter 5, we focused on a classification task, i.e. predicting a categorical property (forest disturbance status). Three classification machine learning models were tested: Random Forest (RF), Stochastic Gradient Boosting (SGB), and Support Vector Machine (SVM). These models were calibrated and validated based on the disturbance classes detected from Landsat time series analysis.

Figure 4.1 - Flowchart summarizing the main methodological steps of this thesis. Input data are represented in gray boxes, while processed data are indicated in black outline white boxes. Modeling processes are shown in dotted outline white boxes.



Source: Author's production.

In chapter 6, six regression models were used to estimate AGB from the LiDAR and HSI metrics. The reference AGB used for training and testing models was obtained from 132 available field plots. The best models were then used to predict AGB over 600 samples where we collected disturbance data. As no single best regression method was found, with more than one method showing similar prediction power, the mean AGB derived from these methods was obtained. Other studies have suggested that model averaging generally performs better than single-model predictions (EXBRAYAT et al., 2013; HU et al., 2015).

Those predicted AGB, together with environmental and anthropogenic disturbance data from the 600 samples over the Brazilian Amazon, were used in chapter 7 to analyze the major factors affecting AGB variability. For this purpose, we performed a stepwise linear regression for both mature and secondary forests using the estimated AGB as the dependent variable and the environmental/anthropogenic variables as potential predictors. Further details on the methods will be presented in the specific chapters.

5 CHARACTERIZING TROPICAL FOREST DISTURBANCE STATUS WITH LIDAR AND HYPERSPECTRAL REMOTE SENSING

5.1 Introduction

The Brazilian Amazon forest is recognized for its key role in providing local, regional and global ecosystem services, including biodiversity maintenance, climate regulation, and greenhouse gas mitigation (STRAND et al., 2018). However, these benefits have been threatened by the large extent of deforestation, forest degradation and their synergistic relation with climate change (NOBRE et al., 2016). The conversion of primary forests by deforestation in the Brazilian Amazon has been well monitored over the past three decades (INPE, 2019). In contrast, the extent of forests regenerating after deforestation and forests subjected to more subtle disturbances, such as selective logging and fire, is less well characterized (ASNER et al., 2009a; TYUKAVINA et al., 2016). Monitoring the extent of the remaining undisturbed forests, as well as the expansion of disturbed mature forests and secondary successions at different vegetation regeneration stages, is critically needed for improving conservation, management, and restoration strategies. Moreover, discriminating forest status under anthropogenic influences can help minimize uncertainties in carbon emission estimates due to forest disturbance and carbon uptake through forest recovery.

In order to better characterize different tropical forest disturbance status, the understanding of how forest structural and functional traits respond to anthropogenic disturbances is a critical goal. Remote sensing features derived from passive multispectral (MSI) or hyperspectral imaging (HSI) are generally related to chemical and compositional traits, having great potential for differentiating land use and land cover (LULC) classes. For instance, Vieira et al. (2003) showed that the combination of NDVI and ETM+/Landsat-7 reflectance of band 5 better separated different successional stages in eastern Amazonia. Da Silva et al. (2014) found an overall accuracy of 89% for mapping LULC in the Tapajós National Forest (Pará-Brazil) by using spectral and textural attributes from the ALI/EO-1 sensor. In the same study area, Galvão et al. (2009) demonstrated the potential of the hyperspectral multiangular CHRIS/PROBA data for the discrimination between primary forest and three stages of secondary successions. Thenkabail et al. (2004) established the advantages of using narrowband Hyperion data over broadband IKONOS, ETM+, and ALI data for classifying complex rainforest vegetation in southern Cameroon.

Active remote sensing, particularly LiDAR, is very suited to characterize vegetation structural traits. It has shown promising results for discriminating successional stages in tropical forests (BISPO et al., 2019; CASTILLO et al., 2012). The structural information provided by LiDAR, along with the spectral information provided by HSI, can better describe the highly heterogeneous human-modified tropical forests. Sun et al. (2019) reported that waveform LiDAR combined with hyperspectral metrics generally produced more accurate forest age maps than using a single data source in a tropical dry forest of Costa Rica. Despite the potential of this synergism in data analysis, there are no investigations on the Amazonian tropical moist forests to test the combined use of LiDAR and hyperspectral data to discriminate LULC classes in human-modified areas.

In this context, this chapter aims to test the effectiveness of LiDAR and HSI data, alone and in combination, to classify forest disturbance status (undisturbed forests, disturbed forests, and two stages of secondary forests). Several LiDAR and HSI metrics related to structural and functional characteristics were calculated and submitted to three machine learning algorithms: Random Forest (RF), Stochastic Gradient Boosting (SGB), and Support Vector Machine (SVM). Thus, the effect of using multiple data sources and different classifiers was tested to better characterize complex forests at different stages of disturbance/recovery in the Brazilian Amazon.

5.2 Material and Methods

5.2.1 Remote sensing data and reference disturbance classes

Twelve sites distributed throughout the Brazilian Amazon biome were considered in this study (Table 5.1). All study sites were surveyed with both airborne small-footprint discrete-return LiDAR and airborne high spatial resolution (1 m) HSI. LiDAR data were acquired between 2016 and 2017 by the Trimble HARRIER 68i system, whereas HSI data were obtained in 2017 by the AISAFenix sensor. The overlapping flight lines of LiDAR and HSI measured approximately 12.5 km by 0.3 km.

Table 5.1 - Sample distribution of forest disturbance classes ($SF_{1\text{-}15yr}$: initial-to-intermediate secondary forests; $SF_{16\text{-}32yr}$: advanced secondary forests; DF: disturbed mature forests; and UF: undisturbed mature forests) across the studied sites.

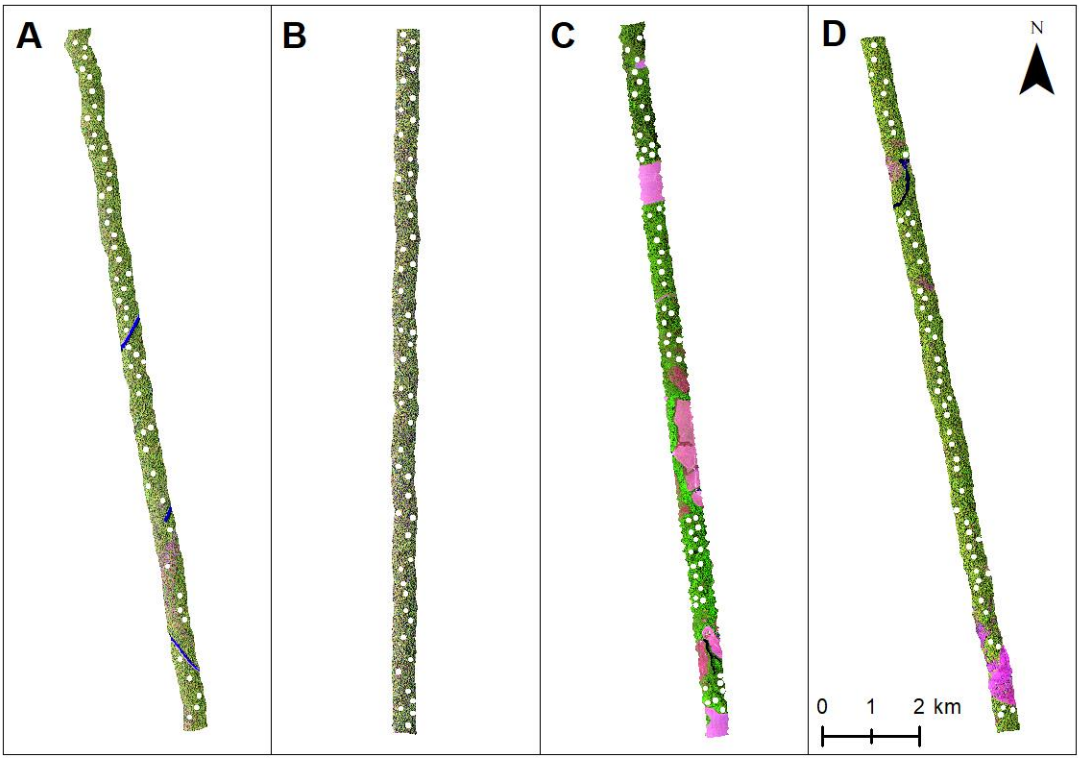
Site	**Sample distribution by class**			
	$SF_{1\text{-}15yr}$	$SF_{16\text{-}32yr}$	DF	UF
MAM	0	0	0	50
ZF2	0	0	0	50
DUC	3	18	0	29
AUT	18	4	21	7
TAP	10	4	31	5
SFX1	3	0	47	0
SFX2	2	0	48	0
PAR	8	10	32	0
JAM	0	0	35	15
ALF	1	1	18	30
FN1	6	3	41	0
FN2	0	1	44	5
Total	51	41	317	191

Besides the airborne remote sensing data, we also used Landsat time series (TM/Landsat-5, ETM+/Landsat-7, and OLI/Landsat-8) from 1984 to 2017 (more than 32 images per site) to identify the status of the reference forest disturbance over the sites. Four classes of forest disturbance were defined based on visual inspection of the Landsat images of the time series: initial-to-intermediate secondary forests ($SF_{1\text{-}15yr}$); advanced secondary forests ($SF_{16\text{-}32yr}$); disturbed mature forests (DF); and undisturbed mature forests (UF). We considered as mature forests the areas under permanent natural forest cover since 1984. Undisturbed mature forests were then defined as mature forests that showed no evidence of disturbance by fire or selective logging, while disturbed mature forests presented at least one of those disturbance types. Secondary forests or successions were defined as forests regenerating after complete deforestation. Although in some studies secondary forests are not distinguished from disturbed mature forests, we considered that these forests are sufficiently different in structure, composition, dynamics, and management to justify their distinction (PUTZ; REDFORD, 2010). Secondary forests in the Amazon are commonly separated into three successional stages based on the stand age (GALVÃO et al., 2009; MORAN et al., 2000): initial (< 5 years), intermediate (5-15 years), and advanced (> 15 years) successions. Here, due to the limited coverage of the initial successions over the study sites, we grouped the initial and intermediate successions into a broader class. Therefore, the class $SF_{1\text{-}15yr}$ consists of areas where the last deforestation event occurred between 2002 and 2016, while the $SF_{16\text{-}32yr}$ areas were deforested before 2002.

5.2.2 Sample allocation

To collect data for training and testing the classification models, a total of 600 samples (50 samples per 12 sites) were allocated in forest cover areas within the flight lines of LiDAR and HSI. The samples were distributed in a spatially balanced way along the flight lines (Figure 5.1), aiming to represent the variability of each site in terms of anthropogenic disturbances and environmental conditions. The samples were separated by at least 100 m from each other. All samples over undisturbed forests were placed by at least 300 m from the forest edges. To capture the spatial variation of forest canopies within a stand, the sample unit chosen was a square plot of 0.25 ha (50 x 50 m). Plots of 0.25 ha have adequate size to represent the structural variability of tropical forests, as shown in previous studies (GRUSSU et al., 2016; ZOLKOS et al., 2013). After tracking the 600 samples over time using the Landsat time series of images, we allocated 51 samples in the $SF_{1\text{-}15yr}$ class, 41 in the $SF_{16\text{-}32yr}$ class, 317 in the DF class, and 191 in the UF class.

Figure 5.1 - Example of sample allocation in four sites: (A) MAM, in the Amazonas state; (B) JAM, in the Rondônia state; (C) PAR, in the Pará state; and (D) ALF, in the Mato Grosso state. Samples are represented by small white squares. The flight lines are represented by the AisaFenix false-color composite with bands centered at 1601 nm (red), 900 nm (green) and 680 nm (blue).



Source: Author's production.

5.2.3 LiDAR and HSI metrics

From the LiDAR and HSI data, structural and functional metrics were derived over the sample plots to be used as predictors in machine learning models. A total of 34 area-based LiDAR metrics were considered, including metrics related to height distribution (e.g., mean, standard deviation, and percentiles of height), canopy cover (proportion of first returns and Leaf Area Density in a specific height interval), structural complexity (Shannon and Simpson diversity indices), and topography (terrain roughness). From the HSI data, we considered a total of 278 metrics: 232 reflectance bands, 30 vegetation indices, 10 continuum-removal absorption parameters (depth and width at five absorption wavelengths), and 6 sub-pixel metrics (GV, NP, Shade, S_{0_30}, S_{30_60}, and S_{60}). To reduce the number of LiDAR and HSI metrics and avoid redundancy, we eliminated highly correlated metrics (absolute Pearson's correlation greater than 0.95) and metrics with linear dependencies, using the *findCorrelation* function and the *findLinearCombos* function from the R package *caret*, respectively (KUHN, 2008). The remained metrics (Table 5.2) were used as predictors for classification models in three different datasets: a LiDAR-only dataset (20 metrics), an HSI-only dataset (42 metrics), and the combination of both data sources (62 metrics). These metrics have been described in detail in Chapter 3.

Table 5.2 - LiDAR and HSI metrics used as predictors in the classification models.

Data Source	**Metric Type**	**Selected Metrics**
LiDAR	Height statistics	H.max, H.mean, H.p05, H.p95, H.sd, H.cv, H.skew, H.kurt
	Canopy cover	PD_{1st}, LAD_{2_10}, LAD_{10_20}, LAD_{20_30}, LAD_{2}, LAD_{6}, LAD_{14}, LAD_{22}, LAD_{26}
	Structural complexity indices	DSCI, HSCI
	Topography	Roughness
HSI	Reflectance bands	R_{461}, R_{549}, R_{673}, R_{852}, R_{1181}, R_{1735}, R_{2149}, R_{2265}
	Vegetation indices	ARI1, ARI2, CAI, D_{LAI}, DWSI2, DWSI3, DWSI4, DWSI5, LWVI1, LWVI2, ND_{Bleaf}, ND_{chl}, NDLI, NDNI, NDVI, NDWI, PRI, PSRI, PWI, REP, RVSI, SR
	Continuum-removal absorption features	D_{495}, D_{980}, D_{1200}, D_{2100}, W_{495}, W_{670}, W_{980}, W_{1200}, W_{2100}
	Sub-pixel fractions	NP, S_{30_60}, S_{60}

5.2.4 Training and validation of classification models

We tested the performance of three machine learning algorithms: Random Forest (RF), Stochastic Gradient Boosting (SGB), and Support Vector Machine (SVM). These classifiers were applied to the three datasets by using the *train* function of the *caret* package. This function fitted each model and calculated a performance measure based on cross-validation (5-fold repeated 10 times) over different tuning parameters to select the optimal model from those parameters. For the RF classifier, the *mtry* parameter was tuned (from the values 2, 4, 6, 8, and 10) and the *ntree* parameter was set to 1000. For the SGB, tuning parameters were *n.trees* (50, 100, and 150) and *interaction.depth* (1, 2, and 3). The parameters *shrinkage* and *n.minobsinnode* were set to the default values (0.1 and 10, respectively). For SVM, we used the Radial Basis Function Kernel by tuning the parameters *cost* (0.5, 1, 2, and 4) and *sigma* (0.01, 0.03, and 0.07).

Regarding the performance measure used to select the optimal model, we considered the overall F1 (F1 average of the four classes). The F1 score combines precision (aka user's accuracy) and recall (aka producer's accuracy or sensitivity) by calculating its harmonic mean (Equation 5.1), thus providing a single performance measurement for a given class (SOKOLOVA; LAPALME, 2009):

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \tag{5.1}$$

In addition to the overall and by-class F1, the overall accuracy (OA) was also reported. Even though OA tends to undervalue the performance of classifiers on smaller classes, this measure is widely used and may be useful for comparison among other studies.

A two-way analysis of variance (ANOVA) followed by a Tukey test was used to assess whether there were any differences in performance measures (OA and overall/by-class F1) among the nine models (3 data sources x 3 classifiers). The eta squared (η^2), i.e. the ratio of the sum of the squares of the factor by the total sum of squares (COHEN, 1988), was calculated to examine the effect of the data source, classifier, and their interaction on overall model performance (OA and F1).

5.2.5 Variable importance

To explore the potential of each LiDAR and HSI metric to separate forest disturbance classes, we performed the Kruskal-Wallis test and calculated the eta squared based on the H statistic (Equation 5.2):

$$\eta^2[\mathrm{H}] = \frac{\mathrm{H} - \mathrm{k} + 1}{n - \mathrm{k}} \tag{5.2}$$

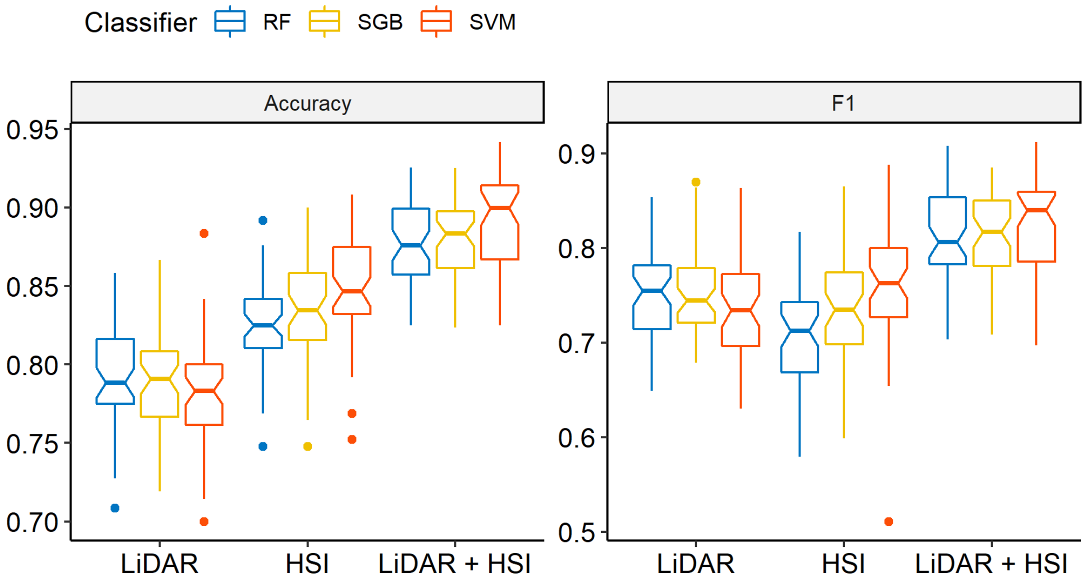
where H is the value obtained in the Kruskal-Wallis test; k is the number of disturbance classes; and *n* is the total number of observations. The non-parametric Kruskal-Wallis test was chosen because some remote sensing metrics had skewed distribution, violating the assumptions of parametric methods. In this context, the eta squared indicates the proportion of total variation in the metric explained by the forest disturbance classes, serving as a univariate measure of metric importance. We also assessed the metric's importance ranking provided by the RF and SGB procedures. All statistical analysis considered a significance level of 0.05 and was performed in the R version 3.4.0.

5.3 Results

5.3.1 Effect of remote sensing data and machine learning classifier on model performance

Figure 5.2 shows the results of the cross-validation, in terms of overall accuracy and overall F1, for each data source and classifier. Irrespective of the classifier used, the best performances were achieved with the use of multisource data for both accuracy (mean cross-validated of 0.88-0.89) and F1 (mean cross-validated of 0.81-0.83). The ANOVA results (Table 5.3) confirmed that the classification performance was mostly affected by the data source, which explained 61% of the OA variation and 31% of the F1 variation. The effect of the classifiers and their interaction with data sources was significant (p-value < 0.05) but weak ($\eta^2 \leq 0.03$) for both OA and F1. Therefore, no single classifier was better for all data sources. However, the overall performance of the SVM was slightly better when used with HSI-only data and the combined LiDAR + HSI data. This was due to the greater capacity of the SVM to discriminate the $SF_{16\text{-}32yr}$ class using HSI data and, thereafter the multisource data (Table 5.4). This is an advantage of SVM, considering that the advanced secondary forest was the most difficult class to classify (lower F1 score).

Figure 5.2 - Overall performance of models with different data sources (LiDAR, HSI, and multisource) and classifiers (RF = Random Forest, SGB = Stochastic Gradient Boosting, and SVM = Support Vector Machine).



Source: Author's production.

Table 5.3 - ANOVA results for the assessment of differences in performance measures.

Factor	**Degree of Freedom**	**Sum of Squares**	**Mean Square**	**F value**	**p-value**	**η^2**
Overall Accuracy						
Data	2	0.70	0.35	366.45	0.00	0.61
Classifier	2	0.01	0.00	3.53	0.03	0.01
Data:Classifier	4	0.02	0.01	5.11	0.00	0.02
Residuals	441	0.42	0.00			
Overall F1						
Data	2	0.57	0.29	107.07	0.00	0.31
Classifier	2	0.02	0.01	4.27	0.02	0.01
Data:Classifier	4	0.06	0.02	5.84	0.00	0.03
Residuals	438	1.17	0.00			

Table 5.4 - Cross-validated overall and by-class performance for each data source and classifier. Distinct letters in a column indicate significant differences in performance from the Tukey test.

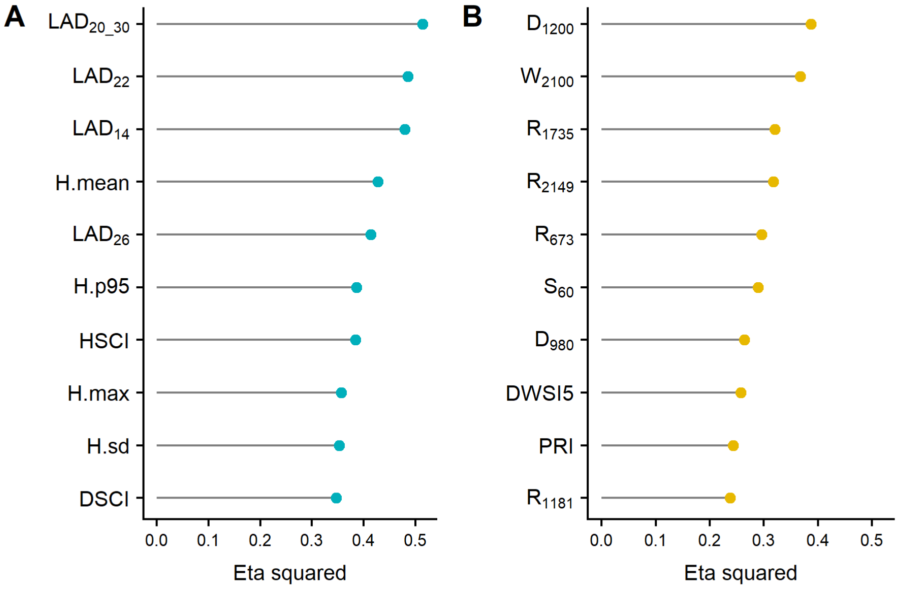
Data	**Classifier**	**Overall Accuracy**	**Overall F1**	**By-class F1**			
				$SF_{1\text{-}15yr}$	$SF_{16\text{-}32yr}$	DF	UF
LiDAR	RF	0.79 a	0.75 a	0.78 a	0.64 ab	0.83 a	0.75 a
	SGB	0.79 a	0.75 a	0.80 a	0.65 ab	0.83 a	0.75 a
	SVM	0.78 a	0.73 ab	0.75 a	0.64 ab	0.82 a	0.73 a
HSI	RF	0.82 b	0.71 b	0.61 b	0.51 c	0.87 b	0.83 b
	SGB	0.84 bc	0.74 ab	0.65 b	0.56 ac	0.88 b	0.85 bc
	SVM	0.85 c	0.76 a	0.61 b	0.68 b	0.89 b	0.87 cd
LiDAR + HSI	RF	0.88 d	0.81 c	0.79 a	0.66 b	0.91 c	0.88 de
	SGB	0.88 d	0.81 c	0.79 a	0.66 ab	0.91 c	0.88 de
	SVM	0.89 d	0.83 c	0.78 a	0.71 b	0.92 c	0.90 e

HSI-only models presented higher OA than LiDAR-only models due to the better performance of HSI data in the prevailing classes of disturbed and undisturbed mature forests. In terms of overall F1, a more suitable metric for unbalanced datasets, the HSI-only models generally performed similarly as LiDAR-only models, except for the RF classifier. RF with LiDAR data displayed a significantly greater overall F1 than the RF with HSI data. LiDAR data generally produced better discrimination of secondary forests compared to HSI data, especially for the $SF_{1\text{-}15yr}$ class. The combination of the LiDAR's ability to better discriminate the two successional stages with the HSI's ability to better discriminate the disturbed and undisturbed mature forests increased the overall F1 of hybrid models in up to 8.1% over the best single-model.

5.3.2 Importance of LiDAR and HSI metrics for class separability

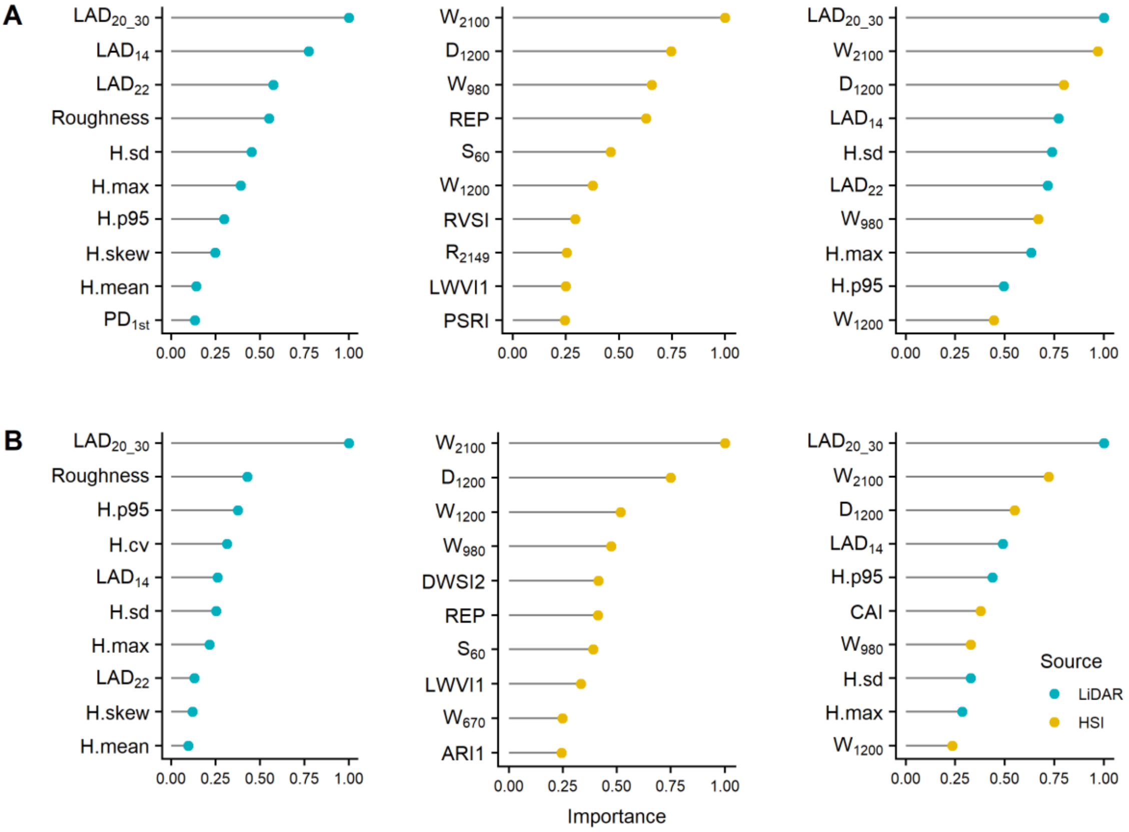
LiDAR and HSI metrics that explained most of the disturbance classes' variability (highest $\eta^2[H]$) are presented in Figure 5.3. Most of these metrics were also ranked as the most important for the RF and SGB models (Figure 5.4). Table 5.5 summarizes the main structural and functional characteristics associated with the LiDAR and HSI metrics useful for discriminating the forest disturbance classes.

Figure 5.3 - Top 10 LiDAR (A) and HSI (B) metrics ranked according to eta squared (η^2[H]).



Source: Author's production.

Figure 5.4 - Relative importance of the 10 highest ranked variables for the RF (A) and SGB (B) classifier with LiDAR, HSI, and the combined dataset.



Source: Author's production.

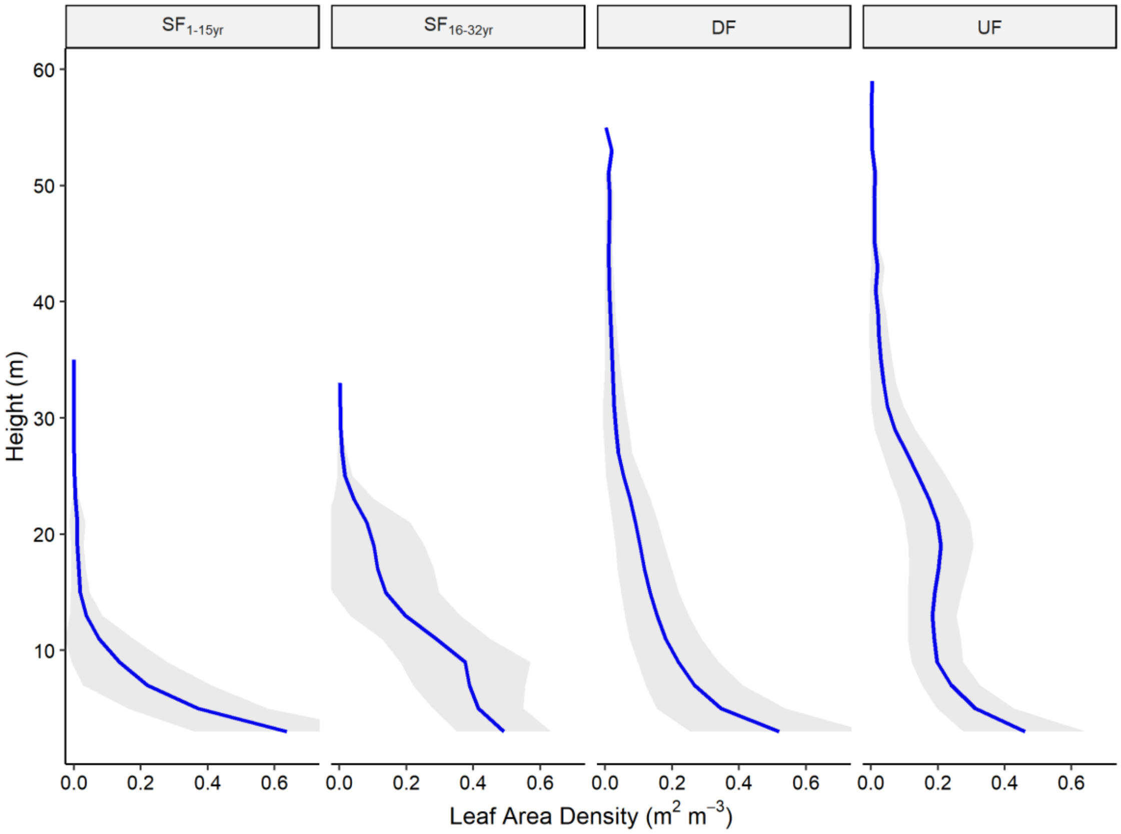
Table 5.5 - Comparison of the average structural and functional characteristics derived from LiDAR and HSI data among forest disturbance classes.

Characteristic	**Metric**	**Unit**	**Disturbance class**			
			$SF_{1\text{-}15yr}$	$SF_{16\text{-}32yr}$	DF	UF
LiDAR						
Upper canopy density	LAD_{20_30}	$m^2\ m^{-3}$	0.00 a	0.06 b	0.28 c	0.68 d
Mean canopy height	H.mean	m	6.53 a	11.59 b	16.68 c	21.51 d
Top of canopy height	H.p95	m	10.92 a	16.43 b	28.83 c	31.30 d
Structure complexity	HSCI	unitless	0.37 a	0.55 b	0.67 c	0.74 d
Canopy heterogeneity	H.sd	m	2.59 a	3.32 b	7.20 d	6.74 c
HSI						
Canopy moisture/LAI	D_{1200}	%	17.85 a	18.89 b	18.61 b	20.03 c
Canopy moisture/LAI and non-photosynthetic biochemicals	W_{2100}	nm	51.57 a	59.24 b	52.46 a	61.32 b
Photosynthetic pigments	R_{673}	%	2.71 a	2.05 c	2.23 b	1.78 d
Canopy gaps/emergent trees	S_{60}	unitless	0.01 a	0.04 b	0.08 c	0.13 d
Health	DWSI5	unitless	1.74 a	2.02 c	1.91 b	2.09 c

Distinct letters in a row indicate significant differences of the characteristic between disturbance classes from a pairwise Wilcoxon test with a Holm correction.

From the LiDAR dataset, metrics related to the upper canopy density (LAD_{20_30}, LAD_{22}, and LAD_{14}) were the most important. The upper canopy density exhibited a significant increase from the initial-to-intermediate secondary successions to the undisturbed mature forests (Table 5.5 and Figure 5.5). The LAD between the height interval of 20 and 30 m (LAD_{20_30}) presented the highest $\eta^2[H]$ (0.51) and relative importance for the RF and SGB classifiers, either in single or hybrid models. Such height interval represents the canopy height distribution of mature forests, which showed a mean canopy height (H.mean) around 20 m and a top of canopy height (H.p95) around 30 m (Table 5.5). Very tall trees (> 30 m) were rare in secondary forests. When they occurred, such tall trees may be remnants of the native vegetation. Thus, metrics related to the top of canopy height, such as H.max and H.p95, showed good separability between secondary and mature forests. Canopy structural complexity and heterogeneity, as measured by metrics such as HSCI and H.sd, were also good indicators of forest regrowth. Consequently, secondary successions had homogeneous canopies with little height variation, while mature forests had heterogeneous and complex canopies, commonly presenting several vertical strata. The terrain roughness presented the lowest η^2[H] (0.02), indicating a poor ability to explain alone the variability between the disturbance classes. However, RF and SGB using only LiDAR data ranked this metric among the four most important. This result suggests that, in a multivariable context, roughness helps improve model performance, probably by explaining extra intra-class variability.

Figure 5.5 - Leaf area density profile for each forest disturbance status. Blue lines are mean values from all samples, while gray areas represent the standard deviation.

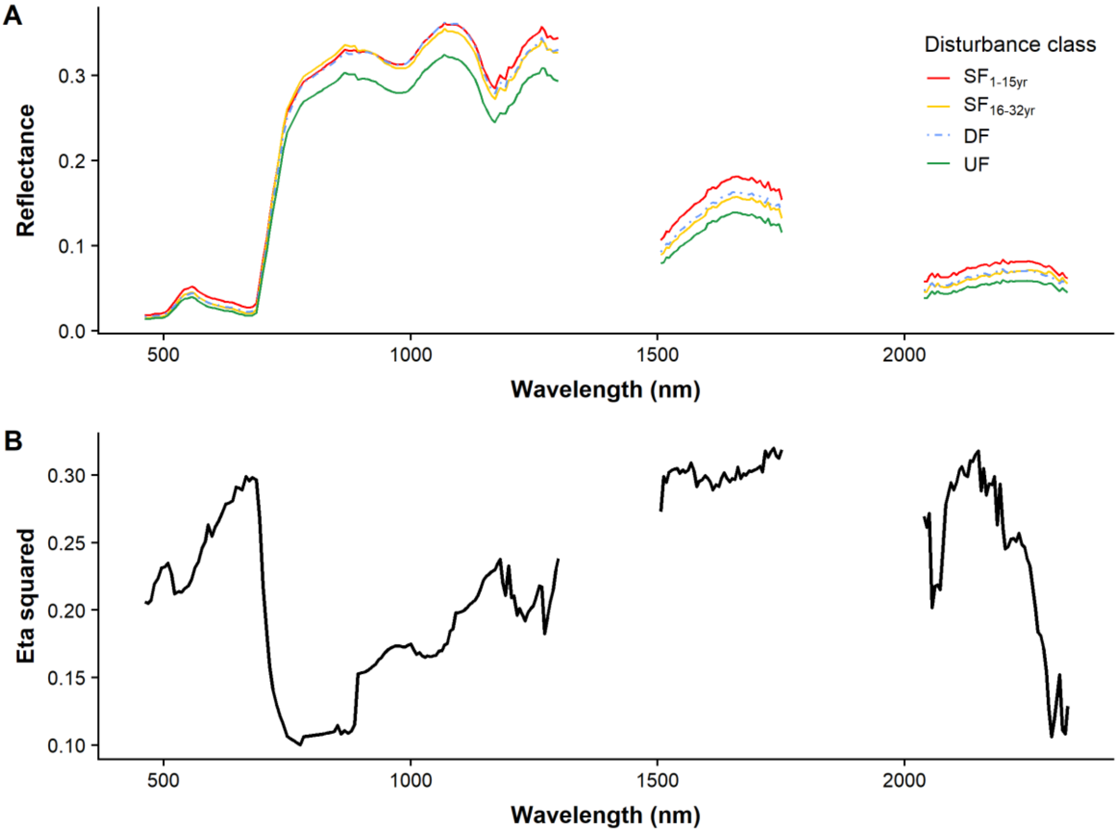


Source: Author's production.

The HSI dataset provided better discrimination between disturbed and undisturbed mature forests, as disturbed forests were generally more spectrally similar to the secondary successions (Figure 5.6A). From the HSI reflectance bands, the most relevant spectral interval for discriminating forest disturbance status (η^2[H] values > 0.3) was the SWIR (1500–2260 nm), especially bands located around 1735 nm and 2149 nm (Figure

5.6B). Red reflectance bands (~ 670 nm), related to chlorophyll absorption, were also of great importance (η^2[H] ~ 0.3) in characterizing forest disturbance. They were followed by leaf-water absorption bands in the NIR (980 nm and, especially, 1200 nm) and photosynthetic pigment features in the blue-to-green spectral region (~ 500 nm). In contrast, the spectral transition from the red-edge to the NIR (740 to 890 nm) showed the lowest η^2[H] values. The greater relevance of spectral regions associated with absorption by biochemical constituents was evidenced by the use of the continuum-removal technique. This approach enhances the vegetation absorption features of interest, while reduces the interference of other factors such as the effects of soil background, illumination or albedo. Absorption features at 1200 nm (D_{1200}) and 2100 nm (W_{2100}) were the most relevant metrics, displaying the highest η^2[H] values (0.39 and 0.37, respectively) and relative importance values for the RF and SGB models among all HSI metrics.

Figure 5.6 - (A) Average reflectance spectra of each forest disturbance class and (B) eta squared (η^2[H]) for all reflectance bands.



Source: Author's production.

In addition to the reflectance bands and absorption features, other HSI metrics were also relevant to distinguish forest classes. From the sub-pixel metrics, the proportion of pixels with shade fraction above 60% (S_{60}) had great importance, presenting a significantly different average among the four disturbance classes (Table 5.5). Deep shaded areas occur due to canopy gaps and shadowing of emergent trees, thus serving as a measure of canopy complexity. Among the vegetation indices, those with higher $\eta^2[H]$ (> 0.2) were DWSI5, PRI, PSRI, ARI1, and DWSI4. REP ($\eta^2[H] = 0.19$) was also very important for the performance of the RF and SGB models using HSI data. The advanced secondary forests were not significantly different from the undisturbed forests according to some HSI metrics, such as W_{2100} and DWSI5 (Table 5.5).

5.4 Discussion

LiDAR and HSI data contain complementary information that, when combined, improved the characterization of tropical forest disturbance status. While LiDAR performed well in classifying successional stages from differentiating them from mature forests, HSI was effective in distinguishing disturbed from undisturbed forests. Canopy structural characteristics, such as height, basal area, and biomass, have been used to characterize successional stages (LU et al., 2003). Thus, LiDAR metrics, which are directly related to canopy structure, provide an important source of information for characterizing secondary successions at different regrowth stages. For instance, the most important LiDAR metrics found here were similar to the ones used to estimate aboveground carbon density in the Borneo’s tropical forests (JUCKER et al., 2018a): canopy cover at 20 m aboveground (*Cover20*), based on the same concept that the LAD_{20_30} or LAD_{22} used here; and top of canopy height (TCH), related to the metrics H.max or H.p95 used in this study. Nonetheless, the main errors related to LiDAR-only models expressed the confusion between disturbed and undisturbed mature forests, as some disturbed areas were structurally similar to undisturbed forests. Furthermore, the recovery of secondary and disturbed forests also implicates changes in species composition displaying different functional attributes. Thus, approaches based solely on structural characteristics limit the characterization of a broad spectrum of forest disturbance status.

From the HSI data, the SWIR spectral region, especially the absorption feature around 2100 nm, was very relevant for characterizing forest disturbance status. The greater canopy complexity along with forest regeneration leads to increased canopy moisture and shadowing, decreasing SWIR reflectance. Furthermore, absorption features around 1700 nm and 2100 nm have been related to non-pigment biochemical components, such as lignin and cellulose (KOKALY et al., 2009), indicating the occurrence of dead or senescent vegetation. For instance, when including the 1660-nm SWIR band in the formulation of vegetation indices (e.g., DWSI5), Apan et al. (2004) obtained the maximum discrimination between healthy and non-healthy vegetation (sugarcane severely affected by disease). Accordingly, other studies have indicated that SWIR bands contain most of the relevant information for distinguishing forest regeneration (VIEIRA et al., 2003; WANG et al., 2019). Water absorption bands, especially at 1200 nm, were also very important. Asner et al. (2004), using EO-1 Hyperion data in the central Amazon, showed that the canopy water metrics were highly sensitive to changes in canopy leaf area and water stress. They also showed that pigment metrics related to LUE (PRI) and anthocyanin levels (ARI) were a proxy for physiological and biochemical changes from chronic water stress. Thus, the importance of those metrics to identify tropical disturbance status suggests a greater susceptibility to canopy stress in disturbed forests.

In agreement with the current results, Thenkabail et al. (2004) also found that hyperspectral bands related to absorption by biochemical constituents, such as water, chlorophyll, lignin, cellulose, and proteins, were very important to characterize tropical forest status following anthropogenic disturbance of different magnitudes. They used EO-1 Hyperion data to classify different LULC classes in African rainforests, including primary forests without evidence of anthropogenic disturbance, degraded primary forest with some evidence of anthropogenic disturbance, young secondary forest (between 9 and 15 years old), mature secondary forest (between 15 and 40 years old), mixed secondary forest with significant anthropogenic disturbance, and agricultural lands recently abandoned (between 1 and 8 years old). They reported an overall accuracy of 96%, achieved with 23 Hyperion bands. The most important spectral intervals for characterizing different vegetation types were located in the 1300–1900 nm, 1100–1300 nm, 1900–2350 nm and 600–700 nm wavelength ranges.

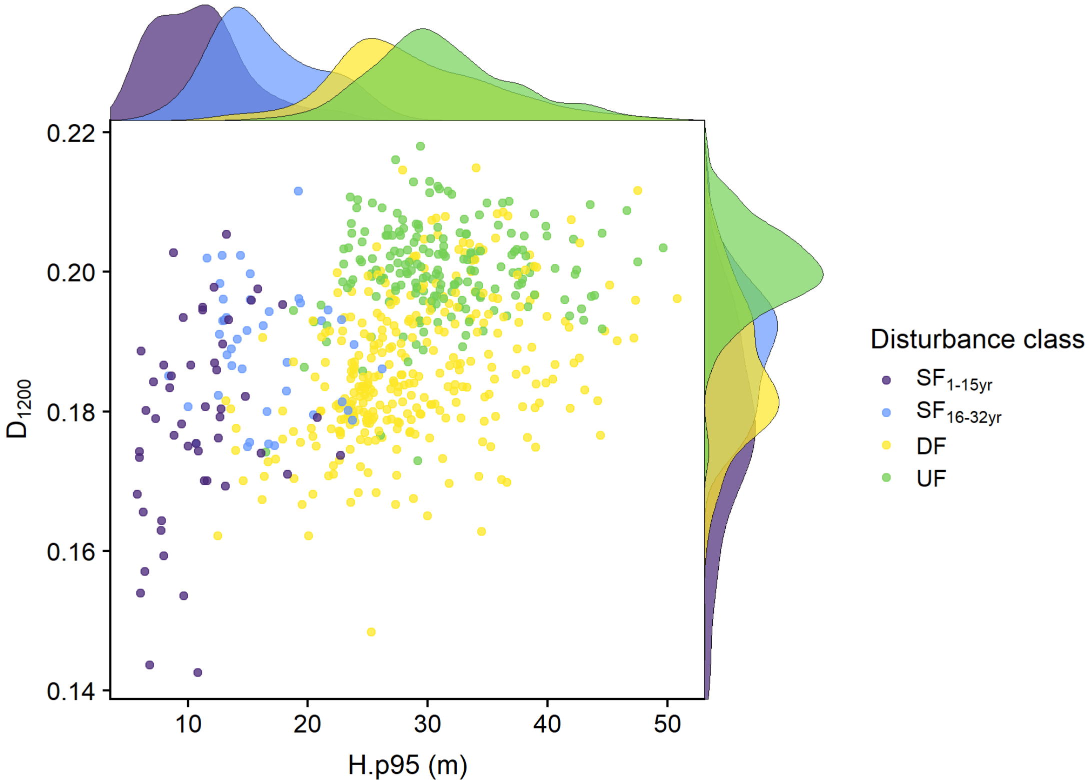
Despite the potential of combining passive and active sensors data, few studies have used this approach for the classification of land cover types and successional stages in tropical ecosystems (e.g., CARREIRAS et al., 2017; SUN et al., 2019). Carreiras et al. (2017) recognized the ability of combined single-date ALOS PALSAR dual-pol and TM/Landsat-5 reflectance data to map mature forest, non-forest and secondary forest on three sites in the Brazilian Amazon. Results presented an overall accuracy of 95-96%. For the secondary forests, the authors also retrieved the stand age, with an RMSE of 4.3-4.7 years (25.5–32.0%) for forests aged up to ~30 years. In the tropical dry forest of Costa Rica, Sun et al. (2019) used different airborne remote sensing data (waveform LiDAR, HSI, and their combination) and machine learning classifiers (Artificial Neural Network, SVM, and RF) to map secondary forest age. The best result was found with the RF classifier and the combination of LiDAR and HSI data (overall accuracy of 83%).

Secondary successions and disturbed mature forests are an integral part of tropical landscapes. However, they present different composition and structure, leading to divergent functioning patterns. Therefore, their accurate characterization and discrimination from the remaining undisturbed forests are essential for establishing conservation and management priorities. The distinct structural and functional characteristics of undisturbed forests suggest that some of their ecosystem services cannot be replaced by degraded or secondary forests (WATSON et al., 2018). Therefore, it is necessary to conserve forests that are still relatively intact and to prevent new areas from being degraded or deforested.

LiDAR metrics of disturbed mature forests generally had intermediate values between secondary and undisturbed forests. Meanwhile, for some HSI metrics, disturbed forests displayed intermediate values between initial-to-intermediate and advanced secondary successions (see example in Figure 5.7). Previous studies (BARLOW; PERES, 2008; BERENGUER et al., 2014; XAUD et al., 2013) have reported a “secondarization” of disturbed mature forests, a process that transforms closed-canopy primary forests into more open forests dominated by short-lived pioneer species due to recurrent anthropogenic disturbances. However, depending on the disturbance intensity and recurrence, disturbed mature forests can retain important structural characteristics of the former primary forests, as well as a generally heterogeneous species composition

(ITTO, 2002). Likewise, the advanced successions had distinct characteristics from the initial-to-intermediate successions and, according to some HSI metrics, were more similar to the undisturbed forests. Thus, both degraded and secondary forests have great potential to provide significant environmental benefits, as well as contribute to poverty alleviation through products and services of socio-economic importance. However, avoiding recurrent disturbance is essential to ensure the continued functioning of forests.

Figure 5.7 - Scatterplot relating an HSI metric (D_{1200} = depth of the 1200-nm absorption band) to a LiDAR metric (H.p95 = 95% percentile of height). Points are colored according to the forest disturbance class. The density plot of each metric is also shown at the respective axis.



Source: Author's production.

Sustainable management practices, such as agroforestry, can also bring benefits in highly disturbed mature forests or younger successions, producing economic value while restoring important physical attributes (e.g., soil fertility). Furthermore, management strategies for degraded and secondary forests, if well planned, can also reduce the pressure on the remaining undisturbed mature forests.

5.5 Conclusion

We concluded that the use of multisource remote sensing data, specifically the combination of LiDAR and HSI, was more effective than the use of advanced machine learning classifiers to improve discrimination between tropical forests with different disturbance status (initial-to-intermediate secondary forests, advanced secondary forests, disturbed mature forests, and undisturbed mature forests). Models based on a single remote sensing data presented a reasonable overall performance (F1 of 0.73-0.75 for LiDAR models and 0.71-0.76 for HSI models), but displayed superior accuracy in specific classes. While LiDAR produced significantly fewer errors for discriminating secondary succession classes, HSI performed significantly better than LiDAR for separating disturbed from undisturbed mature forests. This result was due to the distinct structural characteristics of secondary successions highlighted by LiDAR compared to mature forests. For disturbed mature forests, their functional characteristics derived from the HSI data, such as those related to water stress, photosynthetic efficiency, and senescent or dead vegetation, were more similar to secondary forests than the undisturbed ones.

Thus, combining the strengths of each data source significantly improved the classification performance, increasing the overall F1 in up to 8% relative to the best single-model. In general, no significant differences were observed in overall performance of the machine learning classifiers. However, the SVM had better accuracy in the advanced secondary forest when used with HSI data. It also performed slightly better than RF and SGB with multisource data.

The current study brings unprecedented insights into how advanced remote sensing technologies can be used together to improve our understanding of forest dynamics concerning anthropogenic disturbances in tropical Amazon forests. The role played by data sources, metrics and models, described in this study, represents the first step toward the production of maps to be further validated with detailed field information in the Amazon.

6 COMBINING LIDAR AND HYPERSPECTRAL DATA FOR ABOVEGROUND BIOMASS MODELING IN THE BRAZILIAN AMAZON USING DIFFERENT REGRESSION ALGORITHMS[1]

6.1 Introduction

Aboveground biomass (AGB) is a major component of the terrestrial carbon cycle and its accurate estimate is critical for supporting policies of ecosystem functioning conservation and climate change mitigation (HOUGHTON et al., 2009). Amazonian forests host Earth's most extensive areas of high plant biomass (PAN et al., 2013). However, carbon stocks and balance across the Amazon are still highly uncertain (LE QUÉRÉ et al., 2018; OMETTO et al., 2014).

Remote sensing has been recognized as an effective tool for quantifying carbon stocks over large areas, allowing accurate monitoring at the landscape scale (LU et al., 2014). Several studies have estimated AGB from different sources of remotely sensed data, such as the hyperspectral imaging (HSI) (DE JONG et al., 2003; PSOMAS et al., 2011) and Light Detection And Ranging (LiDAR) (ASNER; MASCARO, 2014; LEFSKY et al., 2002b; NELSON et al., 2017). Among the various types of sensors, LiDAR has been recognized as a consolidated technology to characterize complex forest structure due to its ability to capture three-dimensional information of the land surface (KOCH, 2010). Moreover, LiDAR is less sensitive to signal saturation than passive optical sensors. Despite its advantages, LiDAR has restricted spectral resolution, generally covering a single spectral range in the near-infrared region (LU et al., 2014). Thus, variations in biomass due to species composition and stress may not be accurately detected by this sensor.

In contrast to LiDAR, HSI (also called imaging spectrometry/spectroscopy or hyperspectral remote sensing) sensors acquire data in a large number of narrow and

[1] This chapter is an adapted version of the paper:

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, p. 111323, 2019.

The publisher authorizes the publication of the adapted version of the paper in this thesis (Appendix A).

contiguous spectral bands. HSI is capable of detecting absorption features useful for distinguishing functional and compositional traits (USTIN et al., 2004). For instance, hyperspectral sensors have been used to estimate land cover classes, plant functional types, tree species (ROTH et al., 2015), biochemical content (ASNER et al., 2015), health status (PU et al., 2008), and biophysical properties such as Leaf Area Index (LAI) (GONG et al., 2003) and biomass. On the other hand, when compared to LiDAR, the ability of the HSI instruments to detect vertical structure over dense vegetation is limited since the reflectance comes mostly from the upper canopy (FASSNACHT et al., 2014).

Integrating the complementary information provided by LiDAR and HSI sensors can therefore potentially improve the accuracy of the AGB modeling (KOCH, 2010). Several studies have investigated the potential of combining LiDAR and HSI data for classifying land cover (e.g., GEERLING et al., 2007; KOETZ et al., 2008; WANG; GLENNIE, 2015) or forest species (e.g., DALPONTE et al., 2012; GHOSH et al., 2014). However, few studies have evaluated this combination for estimating AGB (ANDERSON et al., 2008; FASSNACHT et al., 2014; LATIFI et al., 2012; LUO et al., 2017a, 2017b; SWATANTRAN et al., 2011), particularly with focus on tropical regions (CLARK et al., 2011; VAGLIO LAURIN et al., 2014). In Costa Rica, Clark et al. (2011) found that linear regression models combining a single LiDAR and hyperspectral metric were no better than the best model using two LiDAR metrics. However, they pointed out the need to analyze a wide range of LiDAR and HSI metrics, as well as other regression techniques to estimate AGB. In Sierra Leone, Vaglio Laurin et al. (2014) found improved AGB estimates using Partial Least Square Regression (PLSR) from combined LiDAR and hyperspectral data, when compared with LiDAR data alone. Thus, more research efforts are needed to explore different statistical procedures and metrics of HSI and LiDAR for AGB modeling, especially over tropical forests.

Many challenges arise from the integration of different data sources, such as the high data dimensionality, the redundancy of some metrics and the selection of the most suitable prediction model. Linear regression models (LM) have been commonly used for estimating AGB from remote sensing data, because of their simplicity and interpretability (FASSNACHT et al., 2014). However, these statistical models are less flexible than non-parametric techniques, demanding large sample sizes and being affected by multicollinearity (MANQI et al., 2014). Nonparametric machine learning techniques, such as Support Vector Regression (SVR), Stochastic Gradient Boosting (SGB), Random Forest (RF) and Cubist (CB), are more versatile than LM in identifying complex nonlinear relationships and in dealing with high data dimensionality. Such techniques may provide more accurate AGB estimates than linear regression models, especially when multisource data are used (LU et al., 2014).

Apart from identifying proper regression algorithms, an equally important challenge in multisource data integration is the selection of the most informative independent set of metrics for AGB estimation (TORABZADEH et al., 2014). In this context, feature selection methods, such as the recursive feature elimination (RFE), have the advantages of maximizing model performance (GUYON et al., 2002). RFE improves the generalization efficiency by avoiding overfitting while reducing the complexity of the model. The selection of a small subset of metrics generally facilitates the interpretation of the models and their inversion and applicability over large areas.

This chapter aims to explore optimal procedures for improving AGB modeling in the Brazilian Amazon through a comparative analysis of different data sources (airborne LiDAR and HSI, and their combination) and algorithms (linear models with (LMR) and without (LM) regularization, SVR, RF, SGB, and CB). For this purpose, we calculated a large variety of LiDAR and HSI metrics for maximizing the potential information related to vegetation biomass retrieved by each data source. By using a backward feature selection (RFE) method, we dealt with the high data dimensionality and evaluated the impact of reducing the number of input features for the models.

At the best of our knowledge, this is the first study that examines whether the use of HSI in conjunction with LiDAR data can improve AGB estimates using 12 sites regionally distributed over the Brazilian Amazon. Moreover, we addressed the synergy between airborne LiDAR and HSI data for AGB modeling from the perspectives of: (1) using both high spatial (1 m) and spectral resolution optical data; (2) detecting the metrics more related to AGB from a large set of attributes; (3) determining the optimal number of metrics required by each dataset; (4) testing the performance of different regression algorithms; and (5) examining the effect size of data source, regression algorithm and their interactions on models' performance.

6.2 Material and methods

6.2.1 Study sites and field data

This study was conducted on 12 sites in the Brazilian Amazon, representing different climate conditions (Köppen-Geiger classes Af, Am, and Aw) (KOTTEK et al., 2006), soil types (Ferralsols, Acrisols, and Gleysols) (QUESADA et al., 2011), forest structure, species composition, and disturbance history. On each site, forest inventory data were collected to obtain a reference field AGB (Table 6.1).

Table 6.1 - Field-based AGB characteristics for each site.

Site	**AGB mean±sd**	**Plots**	**Plot size**[b]	**Year**	**Source**
	(Mg.ha^{-1})	(*n*)	(m)		
MAM	232±71	8	50x50	2016	IDSM
ZF2	318±73	23	120x20	2015	LMF/INPA
DUC	277±63	11	50x50(20)	2016	SL
AUT	166±43	16	250x10	2017	FATE
TAP	142±78	5	50x50	2016	SL
SFX1	107±78	8	40x40	2012	SL
SFX2	160±86	8	40x40	2012	SL
PAR	101±54	17	125x20(2)	2013	SL
JAM	179±72	11	50x50(5)	2013	SL
ALF	174±61	8	60x40	2017	FATE
FN1	34±38	6	50x50(5)	2015	SL
FN2	170±46	11	50x50(5)	2015	SL
Total	189±101	132			

*The subplot size, when used, is given in parentheses. Abbreviations: IDSM, Instituto de Desenvolvimento Sustentável Mamirauá; LMF/INPA, Laboratório de Manejo Florestal do Instituto Nacional de Pesquisas da Amazônia; SL, Sustainable Landscapes project; FATE, Fire-Associated Transient Emissions in Amazonia.

Forest inventory data comprised 132 sample plots collected between 2012 and 2017. Most plots (116) have approximately 0.25 ha and 16 plots have 0.16 ha. For the oldest plots, we assumed that potential changes in AGB due to temporal differences between forest inventories and remote sensing data acquisitions (2016-2017) had limited influence on the predictive modeling. For instance, changes in yearly AGB across the Amazon biome are, on average, 1.0 Mg.ha^{-1}.yr^{-1} in old-growth forests (BAKER et al., 2004); 2.7 Mg.ha^{-1}.yr^{-1} in selectively logged forests (RUTISHAUSER et al., 2015); and 6.1 Mg.ha^{-1}.yr^{-1} in secondary forests (POORTER et al., 2016). This variation is within the uncertainty in field AGB estimates observed here, which will be further considered in our modeling framework.

Inventory data included species identification and measurements of DBH (Diameter at Breast Height) and total tree height. Due to differences in inventory protocols among the sites, especially with respect to the sampling of palms, lianas, and standing dead trees, we only considered the living trees in the AGB calculation. DBH measurements were obtained with metric tapes for living trees with a minimum of 10 cm DBH. For most sites, all trees that met this DBH threshold were measured along the entire plot area. A subsampling strategy for smaller trees (10-35 cm) was used at sites DUC, PAR, JAM, FN1, and FN2. For these sites, we accounted for the size-dependent sampling area when aggregating individual AGB to plot-level AGB.

The total height of trees was measured using clinometers, whenever possible. When the height was not measured for every tree, a stand-specific height-DBH (H-D) relationship based on a Weibull function (FELDPAUSCH et al., 2012) was used. When no height data were available (ZF2, DUC, JAM, and ALF sites), the regional-specific H-D model proposed by Feldpausch et al. (2012) was used. For further uncertainty propagation of the AGB, each tree in the database was associated with a height error. When a measurement was present, we assumed an error of 12% of the total height, based on the median error found by Hunter et al. (2013). When the height was estimated by Weibull functions, we considered the residual standard error for the local or regional H–D model.

The identification of plant species was used to obtain the values of wood density (WD) with the *getWoodDensity* function from the R package BIOMASS (RÉJOU-MÉCHAIN et al., 2017). The global tree wood density database (CHAVE et al., 2009; ZANNE et al., 2009) was used as a reference. Each tree received a wood density value based on its species- or genus-level average if at least one value in the same genus was available in the reference database. For unidentified trees, or if the genus was not determined in the reference database, the stand-level mean wood density was assigned to the tree, based on trees for which a value was attributed. The standard deviation of wood density for each tree was stored to account for uncertainty in this variable.

Based on the DBH (in cm), height (H, in m) and WD (in g.cm^{-3}), the AGB (in Mg) of individual trees was estimated using the pantropical allometric equation of Chave et al. (2014):

$$AGB_{tree}\ (Mg) = 6.73 * 10^{-5} * (DBH^2 * H * WD)^{0.976} \qquad (6.1)$$

To account for the uncertainty introduced by the measurements and the allometric equation, we propagated the errors using the *AGBmonteCarlo* function (BIOMASS package). This Monte Carlo approach simulated 1000 AGB_{tree} by adding random errors to the measurements and the allometric model parameters (RÉJOU-MÉCHAIN et al., 2017). The individual tree biomass was divided by its associated sampling area to convert to Mg.ha^{-1}. Then, the AGB of all trees of each plot was summed to calculate the plot-level AGB. Thus, each plot had 1000 AGB values and its respective average value (AGB_{mean}). The plots covered a wide range of AGB_{mean}, varying from 2.7 Mg.ha^{-1} to 493.7 Mg.ha^{-1}, with a mean of 188.5 Mg.ha^{-1} and a standard deviation of 101.1 Mg.ha^{-1}.

6.2.2 LiDAR and HSI metrics

Airborne LiDAR and HSI data were collected to derive metrics (45 from LiDAR and 288 from HSI) that could be used as predictors for the statistical AGB models. Prior to the RFE routine, the total number of predictors was reduced by eliminating highly correlated metrics and by checking for linear dependencies (using the functions *findCorrelation* and *findLinearCombos*, respectively, from the R package *caret*). A high correlation threshold (absolute Pearson’s correlation greater than 0.98) was adopted to remove only the metrics with nearly perfect correlation since the RFE algorithm later selects the most important variables to estimate AGB. The remaining metrics after filtering by correlation and linear dependence are summarized in Table 6.2. The removed LiDAR metrics consisted of six height percentiles (H.p25, H.p30, H.p50, H.p60, H.p70, and H.p75) and five canopy cover metrics (PD_{10}, PD_{26}, PD_{30}, LAD_{10}, and LAD_{30}). The removed HSI metrics consisted of 216 reflectance bands, four vegetation indices (CRI2, DWSI1, DWSI4, and VOG1), five absorption features (depth at 670 nm and area of 495, 670, 980, and 2100 nm bands), and the mean shade fraction. Thus, three datasets were tested for AGB modeling: (1) 34 LiDAR metrics; (2) 60 HSI metrics; and (3) their combination (94 predictors). All remote sensing metrics were normalized (centered by mean and scaled by the standard deviation). The metrics have been described in detail in Chapter 3.

Table 6.2 - LiDAR and HSI metrics after filtering by correlation and linear dependency.

Data source	**Metric Type**	**Metrics**
LiDAR	Height statistics	H.max, H.mean, H.p05, H.p10, H.p20, H.p40, H.p80, H.p90, H.p95, H.sd, H.cv, H.skew, H.kurt
	Canopy cover	PD_{1st}, PD_{2_10}, PD_{10_20}, PD_{20_30}, PD_{2}, PD_{6}, PD_{14}, PD_{18}, PD_{22}, LAD_{2_10}, LAD_{10_20}, LAD_{20_30}, LAD_{2}, LAD_{6}, LAD_{14}, LAD_{18}, LAD_{22}, LAD_{26}
	Structural complexity indices	DSCI, HSCI
	Topography	Roughness
HSI	Reflectance bands	R_{461}, R_{522}, R_{604}, R_{659}, R_{694}, R_{701}, R_{852}, R_{1091},R_{1220}, R_{1506}, R_{1646}, $R_{2056,}$ R_{2155}, R_{2309}
	Vegetation indices	ARI1, ARI2, CAI, CRI1, D_{LAI}, DWSI2, DWSI3, DWSI5, EVI, GNDVI, LWVI1, LWVI2, ND_{Bleaf}, ND_{chl}, NDLI, NDNI, NDVI, NDWI, PRI, PSRI, PWI, REP, RVSI, SR, VI_{green}, VOG_{2}
	Continuum-removal absorption features	D_{495}, D_{980}, D_{1200}, D_{2100}, W_{495}, W_{670}, W_{980}, W_{1200}, W_{2100}, A_{1200}, As_{495}, As_{670}, As_{980}, As_{1200}, As_{2100}
	Sub-pixel fractions	GV, NP, S_{0_30}, S_{30_60}, S_{60}

6.2.3 Modeling framework: feature selection and model validation

Six regression algorithms were used, encompassing three main approaches: (i) linear models (LM and LMR), (ii) kernel-based models (SVR), and (iii) tree-based models (RF, SGB, and CB). For the SVR method, we tested three kernels: linear, polynomial, and RBF. We further reported the results of the RBF, which generally performed better than linear and polynomial kernels (Figure B.1). All algorithms were implemented in the R package *caret* (KUHN, 2008), which required other packages listed in Table 6.3. The six regression algorithms were applied to the three datasets in a modeling framework composed by two main steps: (1) selection of the most relevant metrics; and (2) validation of the selected models considering the field AGB uncertainty.

For feature selection, we applied the RFE algorithm (*rfe* routine of the *caret* package), using the $\mathrm{AGB}_{\mathrm{mean}}$ of each plot as the response variable. The RFE was used to assess the effect of the number of input features over the model performance. The performance was evaluated by the Root Mean Squared Error (referred to $\mathrm{RMSE}_{\mathrm{rfe}}$), quantified in a 5-fold cross-validation scheme, repeated 10 times. Model parameters were optimized by using an internal 4-fold cross-validation and selecting the parameters with the lowest RMSE. The RFE procedure started with all available predictors of each dataset. The predictors were ranked according to a criterion of importance, specific for each regression method (Table 6.3). Less important features were sequentially removed prior to modeling until the two most important variables remained. At the end of the process, the optimal feature subset size was selected, defined as the lowest number of predictors whose mean $\mathrm{RMSE}_{\mathrm{rfe}}$ was within the 95% confidence interval of the lowest $\mathrm{RMSE}_{\mathrm{rfe}}$. This approach selects the most parsimonious yet informative model.

Table 6.3 - Description of the regression models used in this study, including the parameters considered and the criteria used to rank the feature importance for AGB estimation.

Type	**Abbr.**	**Model**	**Parameters**	**Feature rank criteria**	**R package**
Linear	LM	Linear Model	-	Absolute value of t-statistic	stats
	LMR	Linear Model with Ridge Regularization	*alpha* = 0 *lambda* = 0.01, 0.5, 1	Absolute value of coefficients	glmnet, Matrix
Kernel-based	SVR	Support Vector Regression with Radial Basis Function Kernel	*cost* = 0.5, 1, 2, 4 *sigma* = e^{i} (i= -5,...,1)	Squared weights*	kernlab
Tree-based	RF	Random Forest	*ntree* = 1000 *mtry* = k/3	Increase in mean squared error by permuting a variable	randomForest
	SGB	Stochastic Gradient Boosting	*n.trees* = 50, 100, 150, 200, 250, 300 *interaction.depth* = 2 *shrinkage* = 0.1 *n.minobsinnode* = 5	Sum of the empirical improvement in squared error over all trees	gbm, plyr
	CB	Cubist	*committees* = 10, 20, 30 *neighbors* = 9	Usage (Linear combination of the rule conditions and terminal model)	Cubist

k is the number of predictors. *GUYON et al., 2002

The best set of metrics and parameters selected for each dataset and regression method (Tables B.1, B.2, and B.3) was used to train 1000 models, each with a Monte Carlo field-AGB simulation as the response variable. This yields a probability distribution of model performance, which accounts for variations due to uncertainties in the field data. We applied a 5-fold cross-validation scheme with 10 repetitions to quantify the performance of each model, in terms of coefficient of determination and RMSE (hereafter termed as CV-R^2 and CV-RMSE, respectively). The CV-RMSE was expressed both in AGB units (Mg.ha^{-1}) and as a percentage relative to the mean AGB of all sample plots (CV-RMSE%).

A two-way analysis of variance (ANOVA) was applied to examine the influence of the data source, regression method, and their corresponding interactions on model performance (CV-R^2 and CV-RMSE). Subsequently, a Tukey's test was considered for pairwise comparison of mean CV-R^2 and CV-RMSE calculated from the 1000 model runs of each 18 combinations of data sources and regression methods. Since the statistical significance (p-value) is affected by large samples, we also calculated the effect size as a measure of practical significance. For the ANOVA, we calculated the eta squared (η^2), the ratio of the sum of the squares of the factor by the total sum of squares. For multiple comparisons between models, we calculated the Cohen's d (COHEN, 1988), the absolute difference between groups, standardized by the residual standard error from the ANOVA. We considered that a difference in mean CV-R^2 or CV-RMSE between models is practically significant when $d \geq 1$, that is, two groups differ by 1 standard error or more.

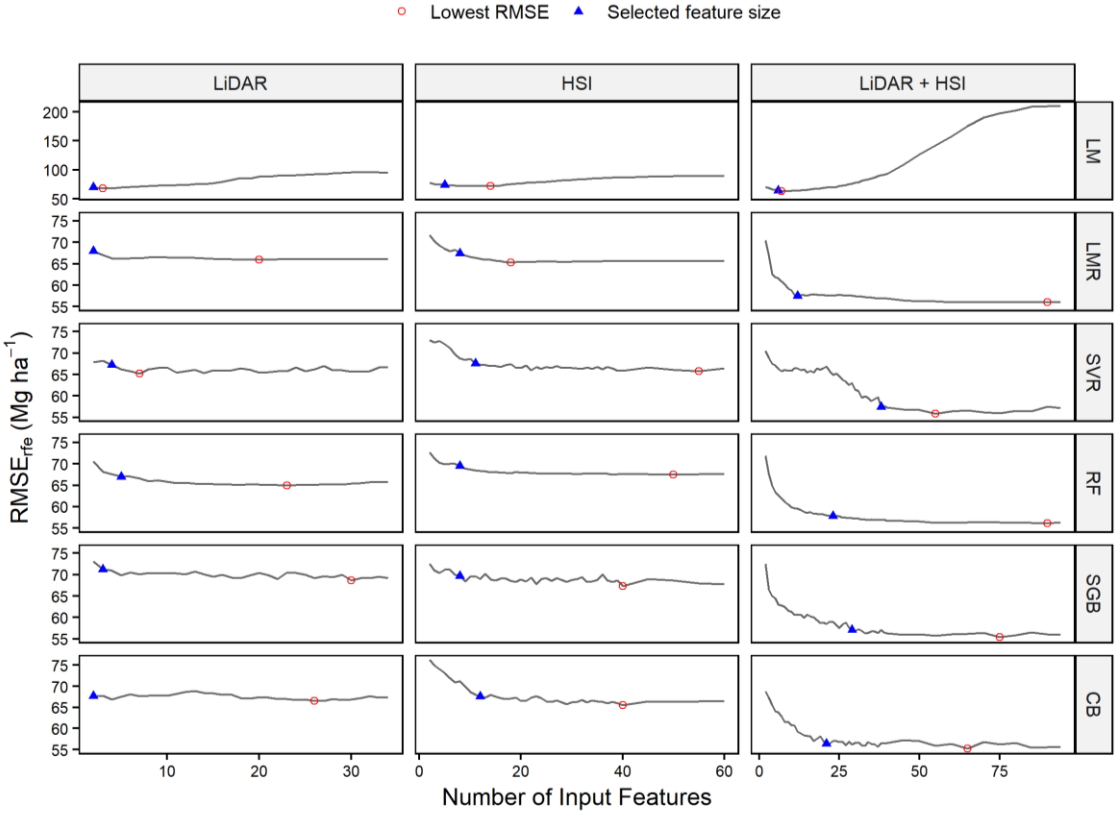
Finally, for analyzing the spatial variability of biomass, the dataset and regression method that produced the highest CV-R^2 and the lowest CV-RMSE were used to predict AGB on a regular 50×50 m grid (corresponding to the field plots area). It resulted in 1000 AGB estimations per pixel from which we calculated the AGB mean and standard deviation.

6.3 Results

6.3.1 Selection of LiDAR and HSI metrics to estimate AGB

LM was the method whose accuracy was mostly affected by the number of input variables, showing an increase in RMSE_{rfe} after reaching the best accuracy (Figure 6.1). The more variables were used in the LM, the greater the increase in RMSE_{rfe}, particularly when the two data sources were combined. This pattern was expected given the limitations of this parametric method in relation to the high dimensionality. The use of regularization (LMR models) solved well this problem. The methods LMR, SVR, and CB, when used only with LiDAR data, were less affected by the number of input metrics. Thus, adding variables into these models did not greatly improve their performance (reduction of the RMSE_{rfe} in up to 4.6%).

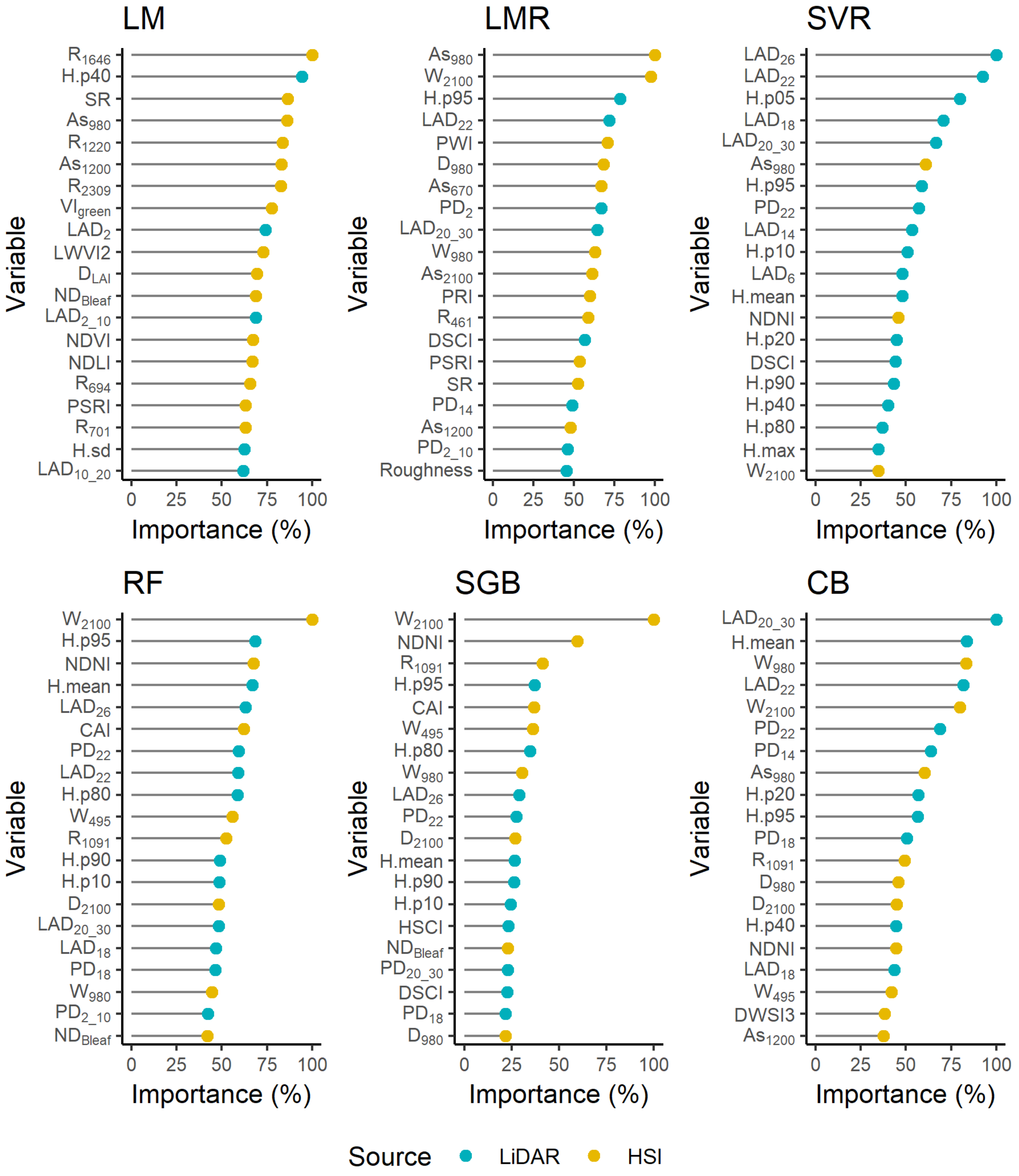
Figure 6.1 - Effect of the subset feature size on the cross-validated RMSE_{rfe} for the regression methods and data sources used. The selected feature size was the smallest possible whose RMSE_{rfe} was within the 95% confidence interval of the lowest RMSE_{rfe}. Note that the RMSE_{rfe} scale for the LM method is different from the others.



Source: Author's production.

LiDAR-only models required fewer metrics (from 2 for LM, LMR, and CB to 5 for RF) than HSI-only models (from 5 for LM to 12 for CB) to achieve optimal performance. The number of metrics selected for the multisensor models varied between 6 for LM to 38 for SVR. The contribution of each data source to the combined models, both in number of selected metrics and in their importance for the model performance, depended on the regression method considered (Figure 6.2). The linear models (LM and LMR) selected more HSI than LiDAR variables when using the combined dataset. SVR prioritized the selection of LiDAR metrics (24 against 14 HSI metrics), which had greater relative importance for the model performance. RF and SGB also selected more LiDAR than HSI variables (14 vs. 9 in RF and 17 vs. 12 in SGB), but the most influential variable was derived from the HSI data (W_{2100}). The CB method had a more even contribution from LiDAR and HSI variables (10 LiDAR metrics vs. 11 HSI metrics). However, the ranking of the variables showed that the CB had a greater influence of a LiDAR metric (LAD_{20_30}).

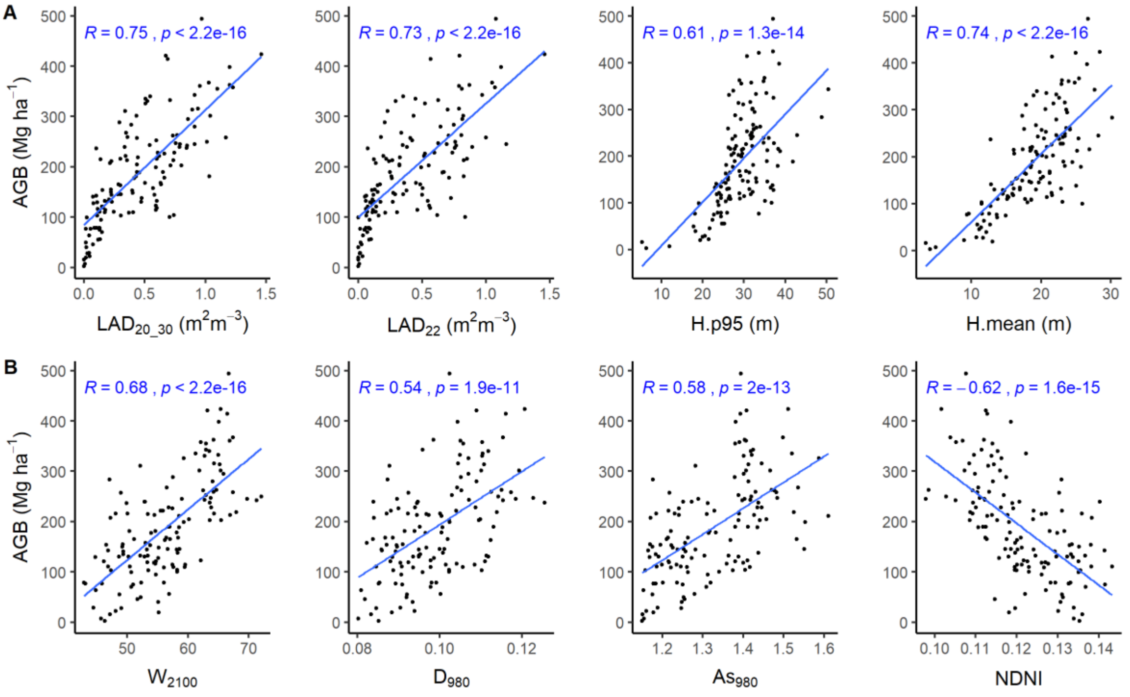
Figure 6.2 - Relative importance of the 20 highest ranked variables for each regression method with the combined dataset.



Source: Author's production.

The most important LiDAR and HSI metrics for estimating AGB were generally consistent among the different models. The most informative LiDAR metrics were related to canopy cover of the upper layers (LAD_{20_30}, LAD_{22}, PD_{22}, LAD_{26}, LAD_{18}, and PD_{18}), height percentiles (e.g., H.p95, H.p40, and H.p05), and mean height (H.mean), showing a positive association with AGB (Figure 6.3A). Structural complexity metrics (DSCI or HSCI) were selected for the methods SVR and SGB based on combined data. Some LiDAR metrics related to topography (roughness), height distribution variability (H.sd, H.cv, H.skew, and H.kurt) and canopy cover (LAD_{2_10} and LAD_{10_20}) were not selected by any model.

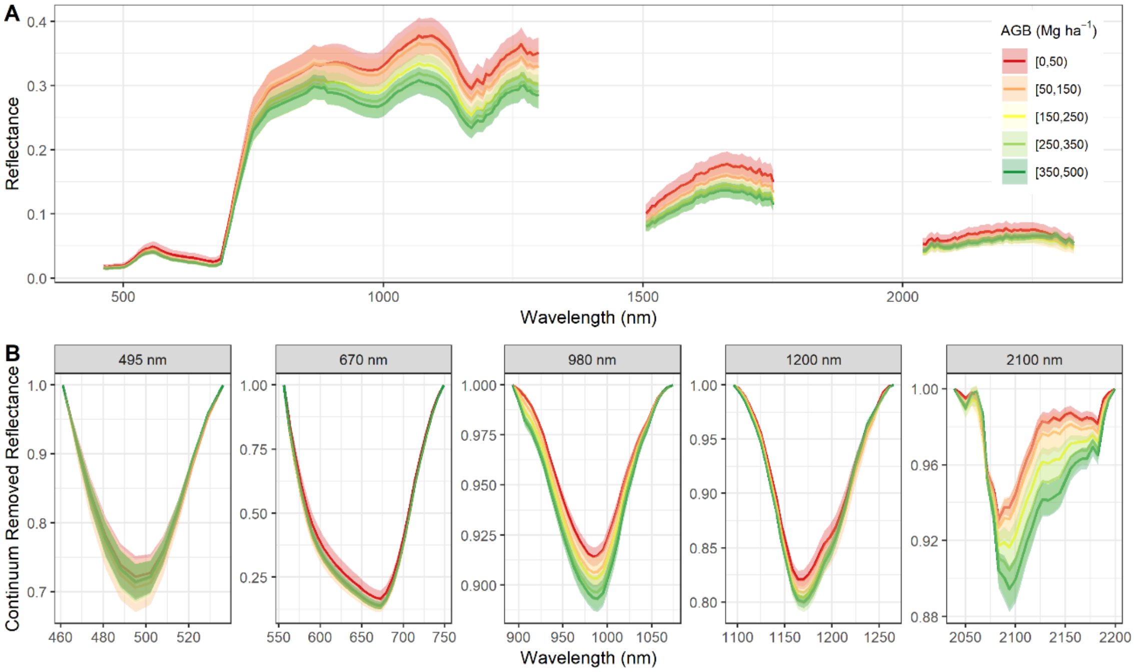
Figure 6.3 - Scatterplots of the four most important LiDAR (A) and HSI (B) metrics for aboveground biomass (AGB) estimation. The blue line represents a linear fit. The correlation coefficient (R) with p-value is showed upward in blue.



Source: Author's production.

For the HSI data, the NIR and SWIR spectral regions were the most sensitive to AGB variations, including the absorption bands at 980 nm (leaf water) and 2100 nm (lignin-cellulose) (Figure 6.3B and 6.4). Thus, metrics from these absorption bands (D_{980}, W_{980}, As_{980}, W_{2100}, and D_{2100}) were ranked as very informative for AGB estimation. Some vegetation indices and reflectance bands from the NIR (LWVI1, PWI, and R_{1091}) and SWIR (NDNI, CAI, ND_{Bleaf}, and R_{1646}) regions were also highly ranked. The proportion of shaded pixels (S_{0_30} and S_{30_60}) was also important for the AGB estimation, either directly, being selected by the methods SVR, SGB, and CB, or indirectly, being associated with the reflectance. Few metrics from the visible region (W_{495}, PRI, As_{670}, and R_{461}) were selected by the RFE. From that, the most informative was the width of the 495 nm chlorophyll absorption band (W_{495}), selected by five models (all except LM) with only HSI data and three models (RF, SGB, and CB) with the combined data. Vegetation indices resulting from a combination of visible and NIR reflectance were selected a few times (SR, PSRI, and VI_{green}) or not selected (e.g., NDVI, EVI, and GNDVI).

Figure 6.4 - Reflectance spectra (A) and continuum-removed reflectance spectra (B) across five aboveground biomass (AGB) ranges, indicated by the different colors. Spectral values are shown as mean ± standard deviation.



Source: Author's production.

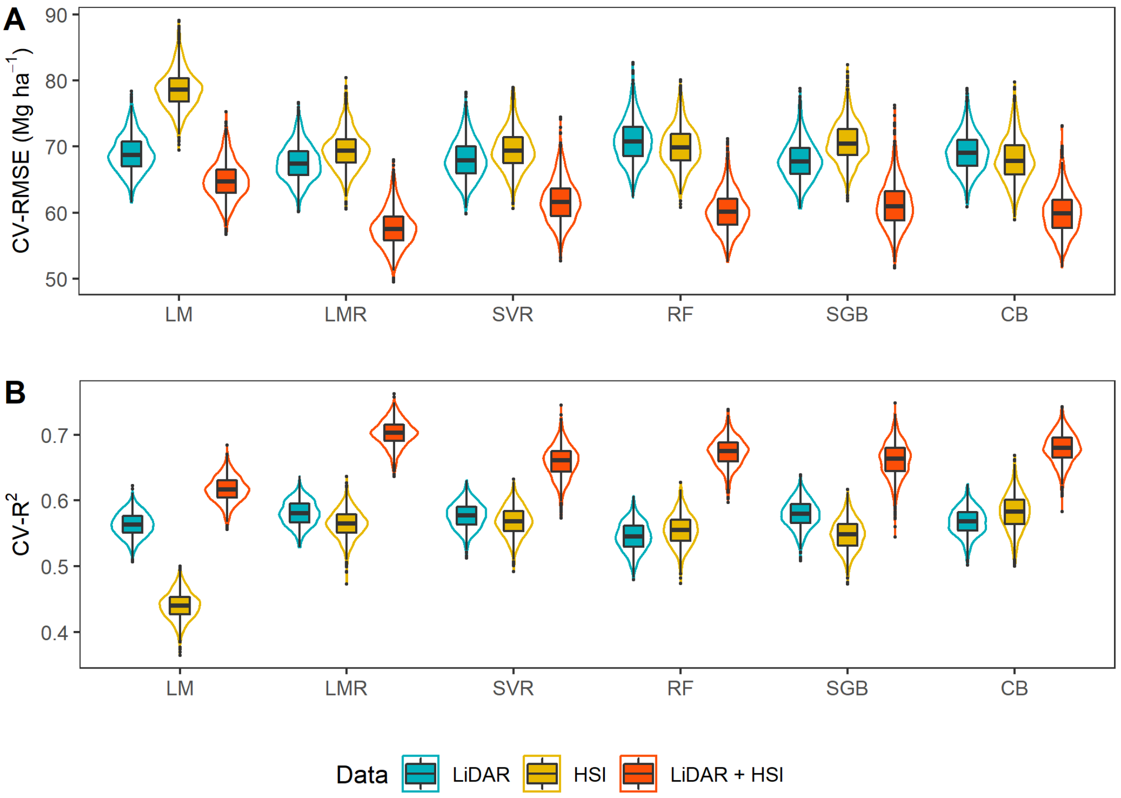
6.3.2 Performance of the data sources and regression methods for AGB modeling

The ANOVA results (Table 6.4) showed that the data source had the greatest effect on models' performance, explaining 65% of the variation in CV-R^2 and 55% of the variation in CV-RMSE. The regression method and its interaction with data source had a smaller contribution to the CV-R^2 (η^2 of 0.14 and 0.09, respectively) and CV-RMSE (η^2 of 0.10 and 0.07, respectively) variation. Therefore, there was no single best regression method. However, the LM method was less suitable for HSI and hybrid data, while the LMR presented high performance for all analyzed data sources (Figure 6.5).

Table 6.4 - Analysis of variance of the cross-validated R^2 and RMSE respective the data source, regression method, and their interaction.

Response variable: CV-R^2						
Factor	**Degree of Freedom**	**Sum of Squares**	**Mean Square**	**F value**	**p-value**	**η^2**
Data	2	50.0	25.0	51,079.6	<2e-16	0.65
Method	5	11.1	2.2	4,515.9	<2e-16	0.14
Data:Method	10	7.3	0.7	1,495.5	<2e-16	0.09
Residuals	17,982	8.8	0.0			
Response variable: CV-RMSE (Mg.ha^{-1})						
Factor	**Degree of Freedom**	**Sum of Squares**	**Mean Square**	**F value**	**p-value**	**η^2**
Data	2	336,205.1	168,102.6	18,120.1	<2e-16	0.55
Method	5	62,476.6	12,495.3	1,346.9	<2e-16	0.10
Data:Method	10	45,849.6	4,585.0	494.2	<2e-16	0.07
Residuals	17,982	166,821.0	9.3			

Figure 6.5 - Distribution of the 1000 cross-validated RMSE (A) and R^2 (B) for each regression method and data source (LiDAR, HSI, and their combination).



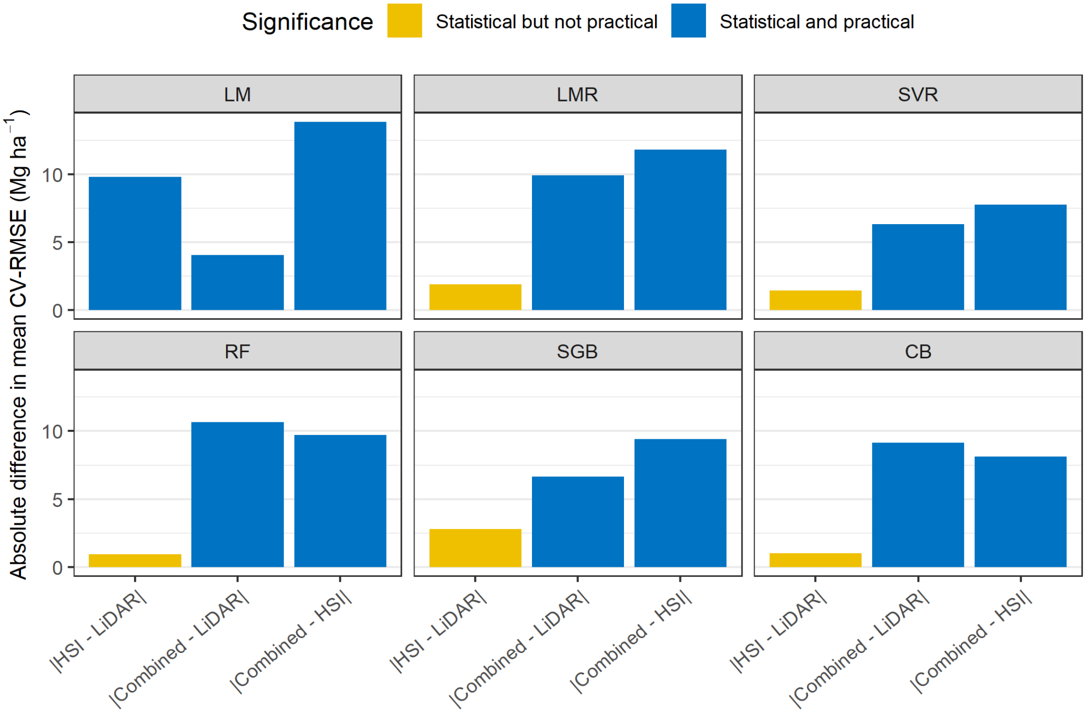
Source: Author's production.

The combination of LiDAR and HSI data improved the performance of the models for all regression methods by reducing the CV-RMSE and increasing the CV-R^2 (Figure 6.5 and Table 6.5). The improvements in CV-RMSE (reduction of 4.05-13.83 Mg.ha^{-1}) and CV-R^2 (increase of 0.05-0.18), achieved by the multisource models relative to models with single data, were both statistically and practically significant (Cohen’s d ≥ 1) for all regression algorithms (Figure 6.6). Relative to the best single-model of each method, the improvements in the combined models reached up to 15% reduction in CV-RMSE and 21% increase in CV-R^2.

Table 6.5 - Average cross-validated performance (for the 1000 model runs) for each regression method and data source.

Model	**Data**	**#Features**	**Mean CV-RMSE**		**Mean CV-R^2**
			Mg.ha^{-1}	%	
LM	LiDAR	2	68.90	36.54	0.56
	HSI	5	78.69	41.73	0.44
	Combined	6	64.85	34.39	0.62
LMR	LiDAR	2	67.60	35.85	0.58
	HSI	8	69.50	36.86	0.56
	Combined	12	57.69	30.59	0.70
SVR	LiDAR	4	68.10	36.11	0.58
	HSI	11	69.54	36.88	0.57
	Combined	38	61.78	32.77	0.66
RF	LiDAR	5	70.91	37.61	0.54
	HSI	8	69.96	37.10	0.55
	Combined	23	60.26	31.96	0.67
SGB	LiDAR	3	67.90	36.01	0.58
	HSI	8	70.68	37.48	0.55
	Combined	29	61.26	32.49	0.66
CB	LiDAR	2	69.12	36.66	0.57
	HSI	12	68.11	36.12	0.58
	Combined	21	59.98	31.81	0.68

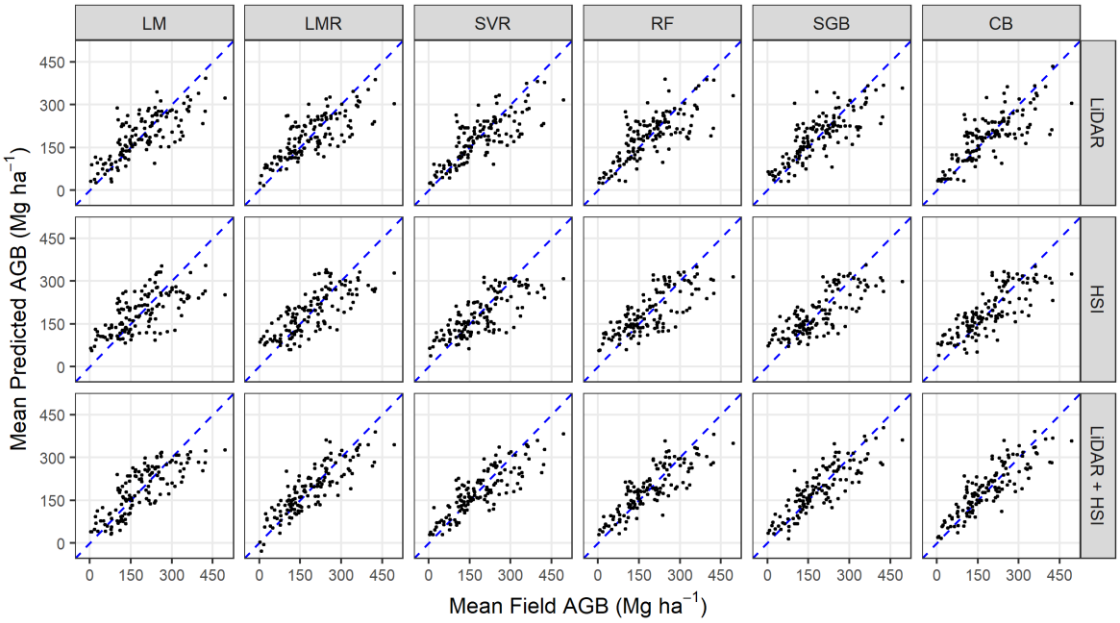
Figure 6.6 - Difference in mean cross-validated RMSE between models based on different data sources for each regression method.



Source: Author's production.

Overall, models based on single-LiDAR data performed similarly to models based on single-HSI data, with no practical difference (Figure 6.6). Only the LM method presented significantly superior performance with LiDAR when compared to the HSI data. SGB performed slightly better with LiDAR data than HSI data with a significant practical difference for CV-R^2 (increase in 0.03), but with no practical difference for CV-RMSE. All models underestimated the AGB for values greater than 300 Mg.ha^{-1}. However, models with multisensor data showed slightly lower underestimation (Figure 6.7). We also found an overestimation for low AGB values (< 50 Mg.ha^{-1}), mainly with the HSI data.

Figure 6.7 - Field AGB_{mean} versus predicted AGB (mean of cross-validated predictions from the 1000 model runs) from the different methods and data sources. The blue dashed 1:1 line is provided for reference.

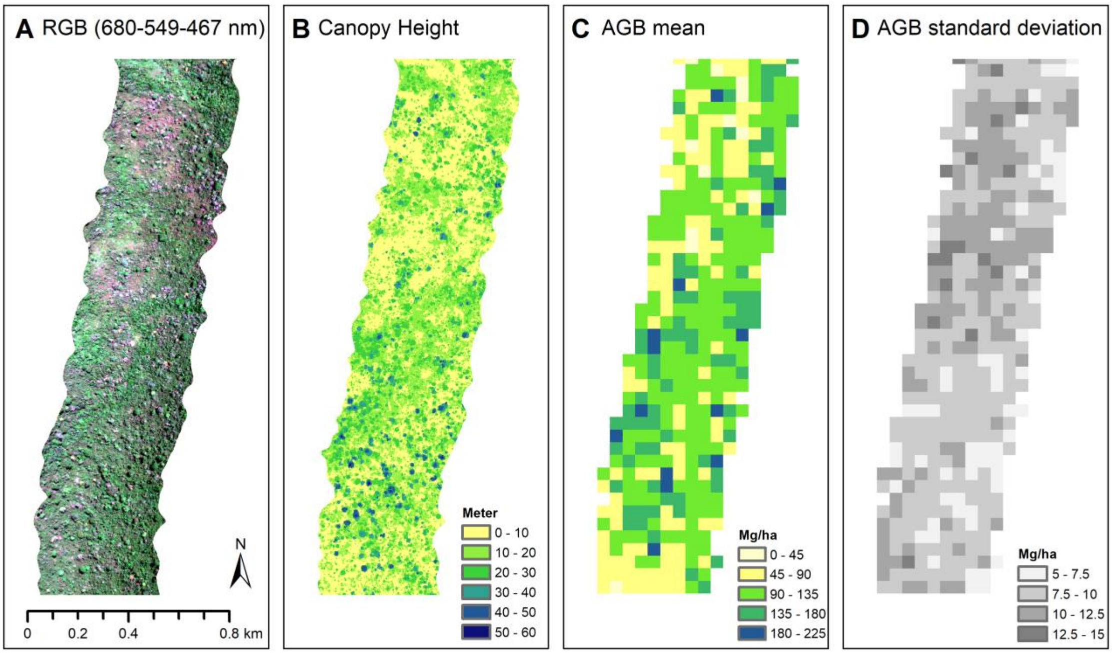


Source: Author's production.

6.3.3 Spatial variability of the predicted AGB

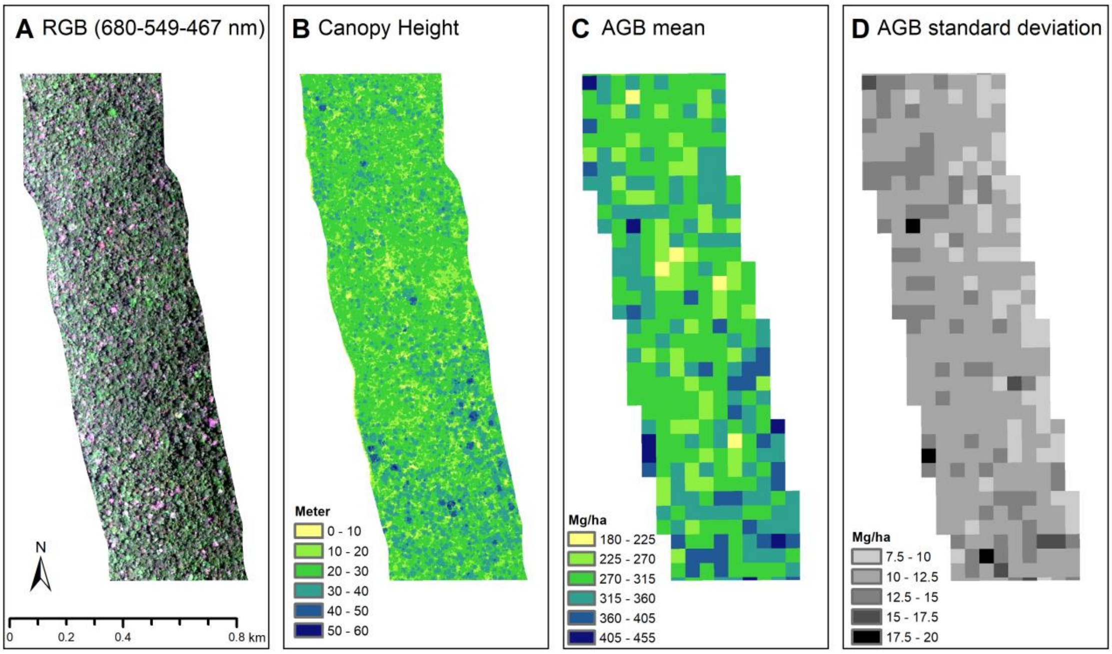
The spatial distribution of the HSI and LiDAR data and the AGB map (mean and standard deviation) derived from their combination are exemplified for the SFX1 (Figure 6.8) and DUC (Figure 6.9) sites. The AGB predictions covered both the variability within and between sites. In the SFX1 site, the predicted AGB ranged from zero Mg.ha^{-1}, due to intensively degraded areas, to 223 Mg.ha^{-1}, due to the presence of tall trees. The DUC site is an old-growth forest accounting for greater AGB. In this site, the predicted AGB varied from 193 Mg.ha^{-1} to 454 Mg.ha^{-1}, due to variations in canopy density and height. The AGB uncertainty (standard deviation) was greater (~ 15 Mg.ha^{-1}) at the tails of the predicted interval, i.e., at the locations with very low or very high predicted AGB.

Figure 6.8. (A) Spatial variability of HSI data (AISAFenix true color composite). (B) LiDAR data (Canopy Height Model). (C) Mean and (D) standard deviation of AGB predictions from the LMR method with multisensor data. Figures A and B are in 1 m resolution, while Figures C and D are in 50 m resolution. Results refer to the SFX1 site.



Source: Author's production.

Figure 6.9. (A) Spatial variability of HSI data (AISAFenix true color composite). (B) LiDAR data (Canopy Height Model). (C) Mean and (D) standard deviation of AGB predictions from the LMR method with multisensor data. Figures A and B are in 1 m resolution, while Figures C and D are in 50 m resolution. Results refer to the DUC site.



Source: Author's production.

6.4 Discussion

6.4.1 Single-LiDAR versus single-HSI AGB predictions

Our study confirmed the reliability of LiDAR-based AGB predictions in tropical ecosystems, consistent with previous tropical studies using small-footprint airborne LiDAR (ASNER; MASCARO, 2014; D’OLIVEIRA et al., 2012; HANSEN et al., 2015; KRONSEDER et al., 2012; LONGO et al., 2016; MAUYA et al., 2015; RÉJOU-MÉCHAIN et al., 2015; VAGLIO LAURIN et al., 2014; ZOLKOS et al., 2013). The LiDAR metrics selected here as important for estimating AGB were also comparable with metrics identified in other studies, such as the mean height (LATIFI et al., 2012; LONGO et al., 2016), height percentiles (LONGO et al., 2016; MANQI et al., 2014; VAGLIO LAURIN et al., 2014), and canopy-cover attributes (LATIFI et al., 2012).

Previous studies that compared LiDAR with hyperspectral sensors have shown that LiDAR was more powerful for biomass prediction (CLARK et al., 2011; FASSNACHT et al., 2014; VAGLIO LAURIN et al., 2014). Koch (2010) states that a direct AGB estimation based only on HSI data is not likely, especially in high biomass stands. However, our results suggest that single-HSI models can provide good AGB predictions even over dense tropical forests with an accuracy equivalent to LiDAR models. The wide range of HSI metrics calculated in this study, exploring the information of different vegetation properties (e.g., canopy structure, biochemistry, leaf/canopy water content, and plant physiology or stress), contributed to the good performance of the HSI models. On the other hand, while few LiDAR variables generally contained most of the information needed to estimate AGB, a larger set of HSI metrics was necessary to achieve a similar performance of the models.

The absorption bands from the SWIR and NIR regions, as well some vegetation indices from the same spectral regions, were the most influential metrics for estimating AGB with HSI data. For instance, the leaf/canopy water absorption bands, centered at 980 nm and 1200 nm, were indicated as important in the analysis. The same was verified for the 2100 nm SWIR absorption band, related to nitrogen, lignin, and cellulose (KOKALY et al., 2009). Previous studies suggested that such biochemical traits co-varied with canopy structure (KOKALY et al., 2009; SERRANO et al., 2002). Therefore, optical metrics from the SWIR and NIR spectral regions have been recommended to estimate canopy structural attributes such as LAI (GONG et al., 2003; LE MAIRE et al., 2008) and AGB (PSOMAS et al., 2011; SWATANTRAN et al., 2011). They have been used also to estimate aboveground forest productivity (SMITH et al., 2002).

In addition to the high spectral resolution, the high spatial resolution of the hyperspectral images used in this study contributed positively to the AGB models. The spatial resolution of 1 m provides information on the distribution of crowns and gaps. This resolution can be more directly related to the forest inventory information used to establish the models. Sub-pixel-based metrics, such as the proportion of shaded pixels, served as a measure of the canopy spatial arrangement, improving the AGB models. The proportion of shade increased with increasing amounts of AGB, reducing the overall reflectance. These results are consistent with those found by Barbier and Couteron (2015), who observed a negative linear relationship between the mean reflectance and the maximum DBH, a measure of forest structure, due to the shade proportion. Moreover, studies based on texture metrics from high spatial resolution optical data have shown good potential to provide non-saturating proxies for stand parameters, including AGB (BARBIER; COUTERON, 2015; PLOTON et al., 2017). As a result, Barbier and Couteron (2015) state that LiDAR is not the only option for monitoring canopy structure and carbon stocks in tropical forests.

Our findings showed that some HSI indices commonly used for biomass estimation (e.g., SR and NDVI) saturated for AGB above 100 $Mg.ha^{-1}$. Thus, these metrics may be more useful for estimating AGB in simpler stand structures than in dense forests. In contrast, the most relevant HSI metrics for AGB estimation found here were almost unaffected by saturation at high AGB values.

6.4.2 Single-models versus combined-models

The improvements in AGB models based on the integration of LiDAR and hyperspectral data were consistent with the studies performed in temperate mixed forests from the USA (ANDERSON et al., 2008), tropical forests from Africa (VAGLIO LAURIN et al., 2014) and in wetland vegetation from China (LUO et al., 2017b). The gain in explained variance (R^2) by the use of the hybrid approach reported in these studies, when compared with the best single-model, was within the range found in our investigation (absolute increase of 6-9%). In contrast, some studies performed in tropical (CLARK et al., 2011) and temperate forests (FASSNACHT et al., 2014; LATIFI et al., 2012; LUO et al., 2017a) have shown only slight (around 2% absolute increase of R^2) or no improvements in AGB estimation after combining LiDAR and HSI for AGB modeling.

The differences in results from the literature can be explained by several factors that influence the performance of the AGB models. Examples of these factors include the regression technique chosen for analysis; the number and type of metrics selected as potential input data; the type of vegetation under study; and the quality of the field and remote sensing data used to obtain the models. For instance, some studies suggest that LiDAR data provide a more straightforward connection with vegetation structure, being able to produce satisfactory predictions with relatively simple techniques, such as linear regression approaches (LONGO et al., 2016; MANQI et al., 2014). On the other hand, hyperspectral measurements relate indirectly with biophysical properties, and thus, may need more complex models (TORABZADEH et al., 2014). In this study, the conventional linear regression model was not very suited for high dimensional datasets (based on HSI and multisensor source). However, the linear model with regularization showed superior performance by solving the issue of multicollinearity between the metrics. Nevertheless, this good performance was only possible because several metrics used in this study were non-saturating with large amounts of AGB, showing a consistent linear relationship with it. Studies based on metrics that saturate over dense vegetation may not find the same results.

Few studies have applied machine learning techniques for estimating AGB from multisource remote sensing data. Fassnacht et al. (2014) verified that the RF method outperformed other approaches (stepwise linear regression, SVR, Gaussian processes, and k-nearest neighbor) when using combined LiDAR and HSI data. Feng et al. (2017) compared different data sources and modeling approaches (linear, nonlinear, RF, and SVR) under stratification and non-stratification conditions of vegetation types. For the combination of LiDAR and RapidEye data, RF had the best performance under stratification. RF emerged also as the best algorithm for different data sources in the study by Cao et al. (2018) when compared to SVR, neural networks, k-nearest neighbor, and generalized linear mixed model. In our study, the regression method had little effect on the models' performance. With the exception of the LM with HSI and hybrid data, all the evaluated algorithms were useful for estimating AGB from remote sensing data.

Selecting the most appropriate metrics to estimate AGB is another factor that can affect model performance. We showed that it was possible to reduce considerably the number of metrics used as input data without losing much accuracy in AGB estimates. Even regression methods not entirely affected by the high data dimensionality can benefit from the reduction in the number of features. More elaborated feature selection procedures can produce parsimonious models for practical applications. Therefore, models based on multisource datasets require strategies to overcome the trade-off between the high data dimensionality and the loss of information for achieving a proper number of features.

The characteristics of the vegetation are also relevant to AGB prediction. In our study, we considered a wide variety of vegetation types, from intact old-growth forests to secondary forests, also including areas under different levels of degradation by fire, logging, or fragmentation. The information gain in AGB modeling provided by the HSI may be related to the discrimination of different vegetation types or conditions, such as canopy stress (SWATANTRAN et al., 2011). Other studies based on multispectral data (VIEIRA et al., 2003; ZHANG W. et al., 2017) have demonstrated the potential of metrics related to the NIR and SWIR spectral regions and shade fractions for differentiating vegetation at different regrowth stages. Nevertheless, it is important to note that variations in the remote sensing data acquisitions, especially the varying view-illumination geometry, may affect some metrics (GALVÃO et al., 2013). In our study, the remote sensing data acquisition was designed to reduce such effects by orienting simultaneously most of the flight lines in the same direction (N-S). Although some variations in the average SZA remained (SZA = 30°±7° across sites), since the data were collected at different locations, we observed that such variations did not produce a systematic effect on the residuals of our best model (Figure B.2).

Modeling AGB in the extensive and highly diverse Amazon region has some obstacles such as the acquisition of high quality and standardized field data. For instance, the variable size of the field plots may be a source of uncertainty in the data analysis. Small plots are more susceptible to spatial heterogeneity, GPS location errors, and boundary effects (i.e., confusion in the inclusion/exclusion of trees at the edges of the plot). Moreover, the shape of the plots may also favor the edge effect, in cases of large perimeter-area ratio (MAUYA et al., 2015). In our study, most plots were larger than 0.24 ha, the minimum area required to achieve model errors lower than 20% of field biomass (ZOLKOS et al., 2013). The few plots smaller than this size or the plots with greater perimeter-area ratio did not influence the residuals of our best model (Figure B.3). Another issue is the scarcity of field data in some underrepresented regions and the considerable uncertainties related to field measurements and allometric equations. Terrestrial LiDAR offers a possible alternative to address this issue by improving field estimates of AGB, and therefore, the calibration and validation of models based on remote sensing data (STOVALL; SHUGART, 2018).

6.5 Conclusions

In this chapter, we explored the potential of combining LiDAR and HSI data for estimating AGB in the Brazilian Amazon, using six regression methods and a great number and type of metrics. We concluded that:

(1) Both LiDAR and HSI data used alone can effectively estimate AGB in tropical forests of the Amazon if proper metrics and regression methods are considered. However, HSI models required more input variables (5-12) than LiDAR models (2-5) for estimating AGB.

(2) The accuracy of the AGB estimates was improved in up to 15% in RMSE and 21% in R^2 after using the hybrid dataset relative to the single model of best performance.

(3) The most informative LiDAR metrics for estimating AGB were related to the upper canopy cover and tree height percentiles.

(4) The most important HSI metrics were associated with the NIR and SWIR spectral regions, mainly the water and lignin-cellulose absorption bands.

(5) From ANOVA, results showed that the source of remote sensing data (HSI, LiDAR, or their combination) had a more important effect than the regression algorithms to estimate AGB. Thus, there was no single best regression method.

This chapter contributes to the investigation of the potential of LiDAR and hyperspectral remote sensing to estimate the AGB of tropical forests. More accurate estimates of forest carbon are highly required, considering the current scenario of global environmental changes.

7 ANTHROPOGENIC AND ENVIRONMENTAL DRIVERS OF ABOVEGROUND BIOMASS IN MATURE AND SECONDARY TROPICAL FORESTS

7.1 Introduction

Amazon rainforests are crucial for the global carbon balance by storing large amounts of carbon, about 86 Pg C in above and belowground plant biomass (SAATCHI et al., 2007). However, they also represent a carbon source when subjected to natural or anthropogenic disturbances that lead to total (deforestation) or partial (forest degradation) forest cover loss. Thus, opportunities to mitigate climate change arise by avoiding anthropogenic disturbances and regenerating already disturbed areas. To meet these efforts, it is essential to monitor forest aboveground biomass (AGB) and understand the factors that control its potential recovery from different disturbance types such as deforestation, wildfires, and selective logging.

Both the AGB density of mature forests and the rate of accumulation following deforestation or degradation vary as a function of environmental factors (GUARIGUATA; OSTERTAG, 2001; VILANOVA et al., 2018). The water availability, determined by annual rainfall and drought intensity, has been mentioned as a key factor for biomass variability (POORTER et al., 2016; SAATCHI et al., 2007). Other factors, such as topographic features, may also affect the biomass potential of an area (ASNER et al., 2009b). Furthermore, local forest disturbances summed to the global climate change may cause irreversible loss of structural and functional characteristics of tropical forests (GIBSON et al., 2011). Thereby, it is highly relevant to understand the AGB loss caused by different types of anthropogenic disturbances compared to undisturbed forests in different environmental gradients.

Field surveys have provided important insights into the factors influencing AGB variability in the Amazon (BERENGUER et al., 2014; POORTER et al., 2016; QUESADA et al., 2011). However, field data are often limited in sample coverage and may be biased to easily accessible locations. Remote sensing technology adds new possibilities for analysis, enabling large-scale data collection, including areas of difficult access. In addition, the combination of advanced remote sensing sources can help to reduce uncertainties in regional scale AGB estimates. In previous chapters of this thesis, we verified that combining LiDAR and HSI data better characterized different forest status according to anthropogenic disturbances (chapter 5) and reduced the uncertainty of AGB models (chapter 6) in the Brazilian Amazon.

Here, multisource remote sensing data (airborne LiDAR and HSI) were used to estimate AGB in 600 samples over the Brazilian Amazon for addressing the following questions:

(i) How different types of anthropogenic forest disturbance (deforestation and degradation by fire and/or selective logging) can affect the AGB of the Brazilian Amazon forests?

(ii) To what extent can variability in AGB of mature forests and secondary successions be explained by anthropogenic (disturbance type and post-disturbance time) and environmental factors (climate and topography)?

(iii) What is the rate of AGB recovery over time and how can climate and topography affect this recovery?

7.2 Material and methods

7.2.1 Study sites and sample plots

This study was conducted at 12 sites in the Brazilian Amazon biome, spanning a variety of anthropogenic and environmental conditions. The sites were distributed over four Brazilian states: the sites MAM, ZF2, DUC, and AUT in the Amazonas state; the sites TAP, SFX1, SFX2, and PAR in the Pará state; JAM site in the Rondônia state; and the sites ALF, FN1, and FN2 in the Mato Grosso state. Each site included a transect of approximately 12.5 km x 0.3 km where airborne LiDAR and HSI data were collected. In the sites AUT, DUC, and TAP, two transects were available.

In order to sample the variability in forest structure and environmental conditions over the sites, 50 square plots of 0.25 ha (50x50 m) were distributed on each site, separated by at least 100 m from each other and spatially balanced within the transects. Therefore, a total of 600 samples (50 samples for each of the 12 sites) were allocated in forested areas over the Brazilian Amazon.

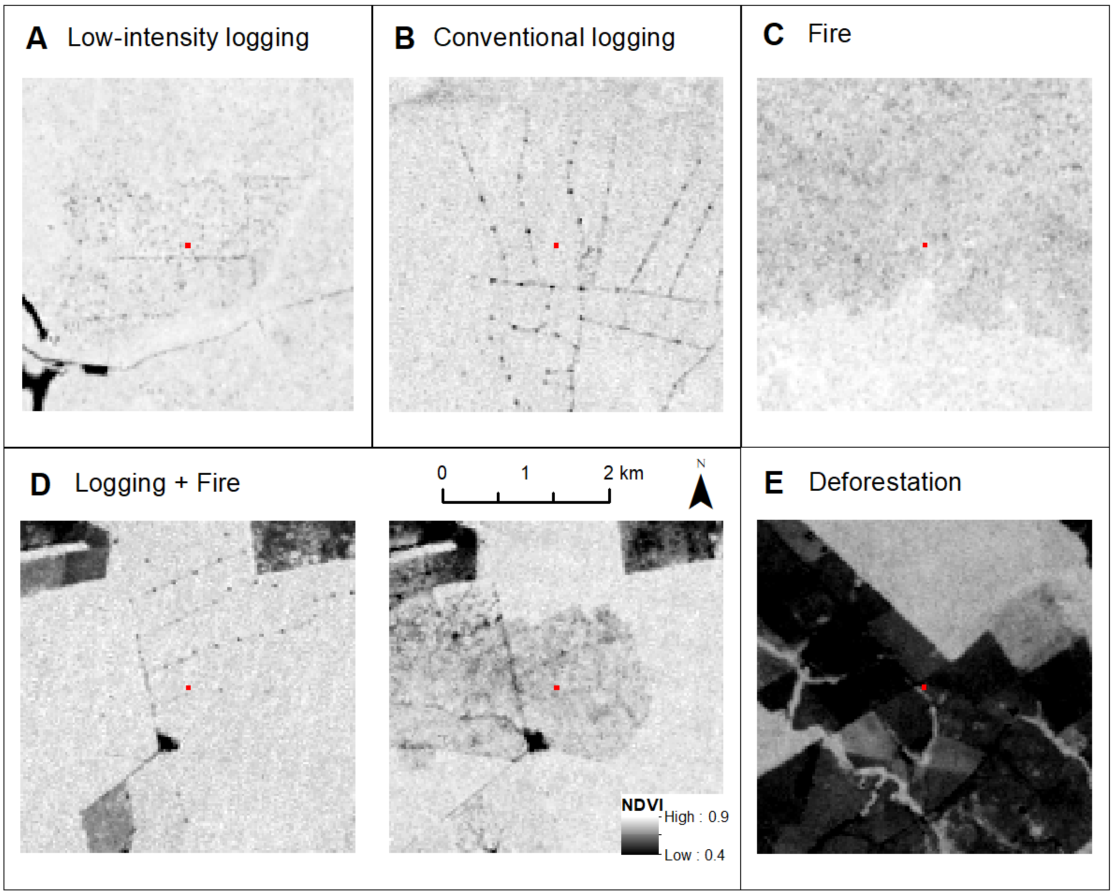
7.2.2 Anthropogenic variables

To obtain the history of anthropogenic forest disturbances over the sites, Landsat images from 1984 to 2017 were visually inspected on the Google Earth Engine platform. Based on the Landsat time series, we obtained two anthropogenic variables: the type of forest disturbance and the time since the last disturbance from 2017 (reference year).

The detected types of disturbance (Figure 7.1) included deforestation (*n* = 92 samples), degradation by fire only (*n* = 146), degradation by conventional logging only (*n* = 54), degradation by both logging and fire (*n* = 68), and low-intensity logging (*n* = 49). For the JAM site, where low-intensity logging was authorized by forest concession, the boundaries of the annual production areas are available online (SFB, 2020). Areas with no evidence of fires but located close to roads (< 300 m) were also considered as disturbed by low-intensity logging since small-scale degradation events are common in such fragmented areas. 191 samples were identified as undisturbed forests, that is, forests with no signs of anthropogenic disturbance since the first year of satellite observation in our study (1984).

Disturbance timing ranged from 1 to 32 years. For areas where multiple degradation events (fire, logging, or both) were detected, we considered the time from the last degradation event. For secondary forests, i.e. forests that regenerated following deforestation, we also referred to the time since deforestation as the age of the stand. Some secondary forests had evidence of understory fire following regrowth. Since not all trees are affected in these cases, we did not consider these fire events when accounting for the age of these forests. For the undisturbed forests, there was no data to confirm whether disturbance events occurred before 1984. In order to include these forests in the analysis, we filled in the time since last disturbance by attributing an arbitrary value of 35 years.

Figure 7.1 - Types of forest disturbance. Images show a post-disturbance Landsat NDVI (Normalized Difference Vegetation Index) for a subset area (4 km^2) around a sample plot (small red squares).



Source: Author's production.

7.2.3 Environmental variables

To evaluate the AGB variability as a function of environmental factors, we obtained climatic and topographic data over the 600 samples of the study area. Two climatic variables related to water availability were considered: Mean Annual Precipitation (MAP) and Climatic Water Deficit (CWD). MAP was obtained from the WorldClim version 2 (FICK; HIJMANS, 2017), in a spatial resolution of 30 s (~1 km^2). The WorldClim data provide a set of bioclimatic variables (MAP is the BIO12 variable) based on the average for the period 1970-2000. CWD was obtained from http://chave.ups-tlse.fr/pantropical_allometry.htm, based on Chave et al. (2014), in a spatial resolution of 2.5 arc-minute (~4.5 km^2). CWD (in mm per year) is the amount of water lost during dry months (defined as months where evapotranspiration exceeds rainfall). CWD is by definition negative, and sites with CWD of 0 are not seasonally water-stressed.

Regarding the topographic variable, a LiDAR-based terrain roughness was considered for characterizing the local topographic variability. Roughness was defined as the difference between the highest and lowest altitude in a 3 × 3 moving window (WILSON et al., 2007). The altitude used to compute the roughness was obtained from a LiDAR-based DTM with a spatial resolution of 10 m. Although the LiDAR data was also used to estimate AGB in this study, no AGB model has considered the terrain roughness as a predictor.

7.2.4 AGB estimation from multisensor data

Airborne LiDAR and HSI data were available for all samples, from which we calculated metrics related to structural and functional characteristics of the vegetation. These metrics were used to estimate AGB from three models that performed best in Chapter 6: the LMR, CB and RF methods, based on a subset of combined LiDAR and HSI metrics. These models were chosen because they displayed the lowest RMSE values, with no practical difference between them. Since there was no best single model, the average predicted AGB was calculated from the estimates of these three models. Other studies have suggested that model averaging generally performs better than single-model predictions (EXBRAYAT et al., 2013; HU et al., 2015). Some values predicted by the LMR method were negative, which were corrected to zero before averaging, as negative AGB values are not possible.

7.2.5 Statistical analysis

To examine whether there were significant differences in predicted AGB among forests submitted to diverse anthropogenic disturbances (deforestation; degradation by fire, conventional logging, low-intensity logging, or both conventional logging and fire; and undisturbed forests), a Kruskal-Wallis test, followed by a pairwise Wilcoxon test (with Holm correction for multiple testing), was applied.

To analyze the effect of anthropogenic and environmental factors on the AGB, we divided the 600 samples into two groups: mature (*n*= 508) and secondary (*n*= 92) forests. Secondary forests are areas affected by deforestation followed by vegetation regrowth, while mature forests included both undisturbed and degraded forests in our analysis. This division is necessary due to the different patterns of regrowth of these forests. The natural regeneration of secondary forests depends on seeds that are dispersed from the nearby remaining forests or that germinate from the dormant seed bank. In contrast, regeneration after fire or logging is typically dominated by plants that remained on the site (PUTZ; REDFORD, 2010).

Thus, a multivariable linear regression model was used to assess the extent to which AGB variability is affected by different anthropogenic and environmental factors in secondary and mature forests. To select the explanatory variables, we first fitted a model with all possible variables and then used a stepwise selection by AIC (function *stepAIC* of the R package MASS). Using this strategy, we selected the optimal set of variables that produced the lowest AIC. The full model for mature forests included, as possible explanatory variables, the type of disturbance (fire, conventional logging, low-intensity logging, and logging + fire), time since last disturbance, CWD, MAP, and natural logarithm (ln) of roughness. An ln transformation for roughness produced a more linear relationship of this variable with AGB. On the other hand, the linearity was not observed for the other variables after the ln transformation, which was not implemented. For secondary forests, the same explanatory variables used for mature forests were considered, except the disturbance type. To further investigate patterns of AGB recovery over time in secondary forests, we also built simple linear regressions considering AGB as a function of forest age in different climatic conditions: low to moderate water-stressed areas (CWD > -400) and highly water-stressed areas (CWD < -400). The intercept for all regression models with secondary forests was set to zero since we expect no AGB in the stand age zero.

To determine the effect of each selected variable on the AGB, we calculated the magnitude of the t-value, i.e. the coefficient estimate divided by its standard error. To validate the models, we examined the distribution of residuals against fitted values and the normality of residuals, using QQ (quantile-quantile) plots and the Shapiro-Wilk test.

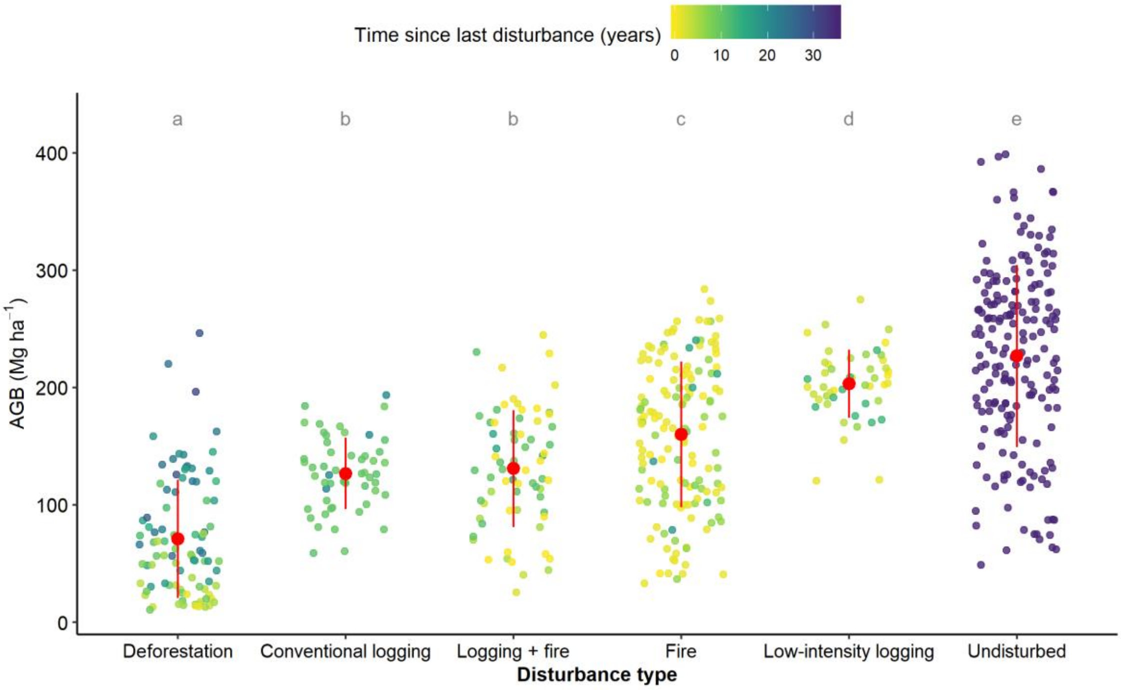
All statistical analysis considered a significance level of 0.05 and was performed in the R version 3.4.0.

7.3 Results

7.3.1 AGB variation as a function of the type of disturbance

The type of disturbance had a significant contribution to the AGB variability across the Brazilian Amazon. Vegetation regenerating after deforestation and/or degradation by fire/logging had a significantly lower average AGB than undisturbed forests (Figure 7.2). Undisturbed forests exhibited an average AGB of 226.73 Mg.ha^{-1} and high inter-site variability, especially due to the lower values in the seasonally flooded MAM site (Table 7.1). Secondary forests accounted for the lowest mean AGB (70.93 Mg.ha^{-1}), ranging from values around 15 Mg.ha^{-1} in younger successions (up to 10 years old) to values generally greater than 100 Mg.ha^{-1} in older successions (> 15 years). Disturbed mature forests displayed, on average, intermediate AGB values between secondary and undisturbed forests. However, some areas degraded by fire or logging + fire showed equivalent AGB to younger secondary successions. AGB average of forests disturbed by conventional logging (126.74 Mg.ha^{-1}) had no significant difference from the forests disturbed by both logging and fire (130.84 Mg.ha^{-1}), being 42-44% lower than undisturbed forests. Forests degraded only by fire showed high AGB variability, with a mean (159.99 Mg.ha^{-1}) 29% lower than the undisturbed ones. Finally, forests disturbed by low-intensity logging had the smallest AGB difference compared to undisturbed forests, with an average AGB of 203.13 Mg ha^{-1} (10% lower than AGB average of undisturbed forests).

Figure 7.2 - AGB variability according to the disturbance type. The time since last disturbance is displayed in different colors. The average AGB ± standard deviation is shown as red points and lines. Distinct letters indicate significant differences between average AGB from a pairwise Wilcoxon test with a Holm correction.



Source: Author's production.

Table 7.1 - Summary of the aboveground biomass (AGB) estimated by the remote sensing models for each site and disturbance type.

Site	**Disturbance type**	***n***	**Estimated AGB (Mg.ha^{-1})**			
			Min.	Max.	Mean	Sd.
MAM	Undisturbed	50	48.6	228.5	135.5	49.5
ZF2	Undisturbed	50	227.3	392.0	283.0	36.7
DUC	Deforestation	21	28.4	246.2	132.4	47.3
	Undisturbed	29	236.7	398.4	306.7	44.2
AUT	Deforestation	22	13.0	86.7	44.4	24.5
	Fire	21	41.7	229.1	156.3	47.8
	Undisturbed	7	125.0	202.4	165.4	29.5
TAP	Deforestation	14	50.5	158.6	92.4	34.3
	Logging + fire	6	171.4	244.8	201.9	26.0
	Fire	25	120.8	283.8	218.4	38.7
	Undisturbed	5	254.3	332.3	289.9	38.1
SFX1	Deforestation	3	15.5	52.2	28.2	20.7
	Fire	47	33.0	199.4	111.0	49.1
SFX2	Deforestation	2	68.9	75.4	72.2	4.6
	Fire	48	63.7	256.3	175.2	54.4
PAR	Deforestation	18	10.6	66.0	35.2	16.6
	Logging + fire	32	51.4	228.8	128.7	45.4
JAM	Low-intensity logging	35	155.0	275.0	210.9	24.8
	Undisturbed	15	179.9	334.5	229.2	40.1
ALF	Deforestation	2	49.5	89.0	69.2	28.0
	Fire	4	200.1	240.2	221.4	18.8
	Low-intensity logging	14	120.7	231.8	183.8	31.3
	Undisturbed	30	172.7	297.5	222.1	31.4
FN1	Deforestation	9	13.0	103.5	37.5	29.4
	Conventional logging	17	60.4	169.0	110.1	31.5
	Logging + fire	23	25.4	230.2	111.4	50.5
	Fire	1	102.0	102.0	102.0	-
FN2	Deforestation	1	138.9	138.9	138.9	-
	Conventional logging	37	58.8	193.3	134.4	26.8
	Logging + fire	7	118.6	178.4	143.7	21.1
	Undisturbed	5	121.5	187.5	155.8	29.3

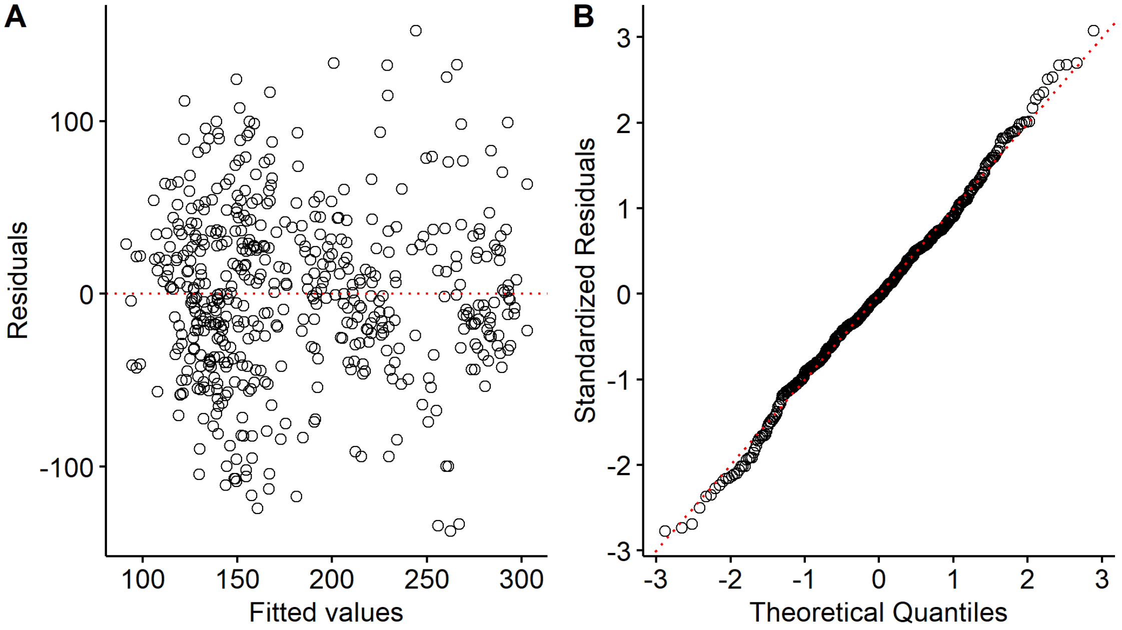
7.3.2 AGB drivers in mature forests

As expected, the results of the multivariable linear regression model for mature forests (Table 7.2, Figure 7.3) confirmed that the occurrence of anthropogenic disturbances had the greatest effect on AGB variability, especially disturbances caused by fires, conventional logging, or both logging and fire. Considering the mature forests across all sites, the variables selected by the stepwise procedure indicate that disturbance by fire, conventional logging, or logging + fire, MAP, roughness, CWD, and time since last disturbance, explained 55% of the AGB variability. Time since last disturbance showed no significant effect on AGB (p-value < 0.05), as well as the disturbance due to low-intensity logging, not selected by the stepwise procedure. MAP had an unexpected negative effect on AGB, which can be explained by the low AGB values shown at the MAM site. This site has the highest MAP (> 3000 mm) over the study area, which combined with its low terrain roughness, characterizes a poor drainage condition that leads to seasonal flooding, negatively affecting biomass (Figure 7.4). By excluding this site from the analysis, MAP no longer had a significant negative effect on AGB and the disturbance due to low-intensity logging became significant. The disturbance timing, however, continued to have no significant effect on AGB and was not selected by the stepwise procedure. Topography still affected AGB, especially over undisturbed forests (R^2 of 0.43 for all undisturbed sites and 0.12 when excluding MAM). In short, the AGB in mature forests was affected mainly by the disturbance type, followed by terrain roughness and CWD.

Table 7.2 - Multivariable linear regression results for mature forests considering all sites and excluding the MAM site.

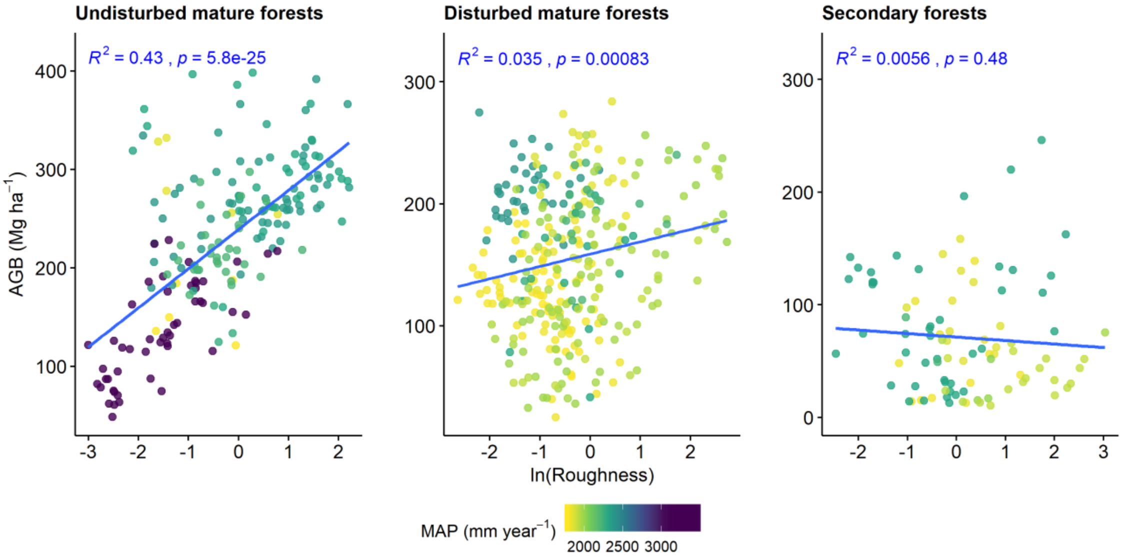
Variable	**Coeff.**	**CI 95%**	**Std. Error**	**t-value**	**p-value**	**R^2**
Mature forests (n = 508)						
Intercept	513.16	[450.96; 575.36]	31.66	16.21	8.5E-48	0.55
Disturbed by fire	-99.83	[-117.80; -81.86]	9.15	-10.91	5.1E-25	
MAP	-0.11	[-0.13; -0.09]	0.01	-10.26	1.5E-22	
Disturbed by conv. logging	-99.88	[-120.44; -79.32]	10.47	-9.54	6.1E-20	
Disturbed by logging + fire	-95.67	[-116.61; -74.74]	10.65	-8.98	5.5E-18	
ln(Roughness)	16.29	[11.88; 20.71]	2.25	7.25	1.6E-12	
CWD	0.18	[0.12; 0.25]	0.03	5.28	1.9E-07	
Time	0.49	[-0.05; 1.02]	0.27	1.79	7.4E-02	
Mature forests with MAP < 3000 mm (n = 458)						
Intercept	276.18	[262.69; 289.66]	6.86	40.24	2.4E-151	0.55
Disturbed by fire	-91.73	[-103.67; -79.78]	6.08	-15.09	5.9E-42	
Disturbed by logging + fire	-91.43	[-113.99; -68.86]	11.48	-7.96	1.4E-14	
Disturbed by conv. logging	-87.66	[-110.27; -65.05]	11.50	-7.62	1.5E-13	
ln(Roughness)	13.25	[8.54; 17.97]	2.40	5.53	5.5E-08	
Disturbed by low-int. logging	-32.82	[-50.39; -15.24]	8.94	-3.67	2.7E-04	
CWD	0.11	[0.05; 0.17]	0.03	3.54	4.4E-04	

Figure 7.3 - Distribution of residuals versus fitted values (A) and QQ plot (B) for the multivariable regression with all mature forests. The regression model without the MAM site presented the same patterns in residuals distribution (Figure B.4).



Source: Author's production.

Figure 7.4 - Relationship between estimated AGB and log-transformed roughness for undisturbed, disturbed, and secondary forests. Blue lines represent a linear fit. Points are colored by the MAP (Mean Annual Precipitation).

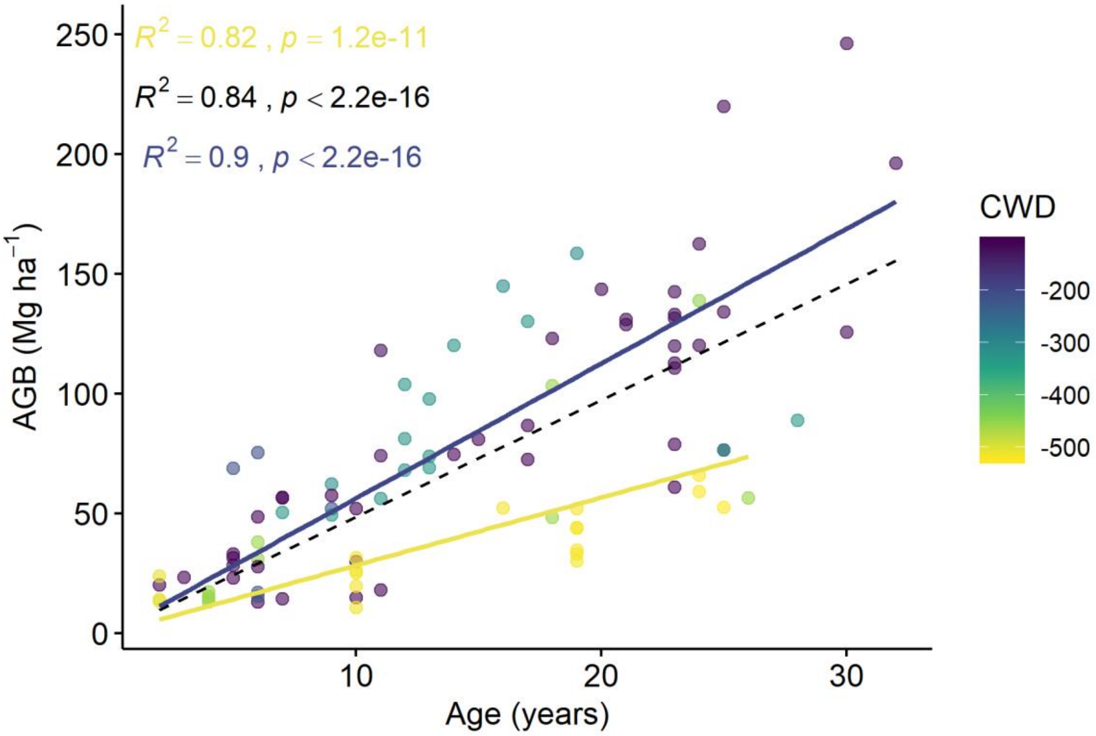


Source: Author's production.

7.3.3 AGB drivers in secondary forests

For secondary forests, the stand age was the most important variable, explaining 84% of the AGB variability across all sites (Figure 7.5). However, CWD and MAP also had a significant effect on second-growth AGB, explaining, along with age, 86% of their variability. Keeping climate conditions constant, the recovery rate of biomass in secondary forests was 4.42 $Mg.ha^{-1}.yr^{-1}$. In assessing the relationship between age and AGB of secondary forests subjected to different water stress gradients, we found that the recovery rate of AGB was slower (2.83 $Mg.ha^{-1}.yr^{-1}$) where climate stress was high (CWD < -400 $mm.yr^{-1}$) and faster (5.63 $Mg.ha^{-1}.yr^{-1}$) in areas with less water stress (CWD > -400 $mm.yr^{-1}$). The results of simple and multivariable linear regressions for secondary forests are summarized in Table 7.3. Figure 7.6 shows the distribution of residuals for the multivariable regression with secondary forests.

Figure 7.5 - Estimated AGB as a function of the age of vegetation regrowth. The yellow line represents a linear fit for samples with Climatic Water Deficit (CWD) < -400. The blue line represents a linear fit for samples with CWD > -400. The black dashed line represents a linear fit for all samples. Points are colored by the CWD.

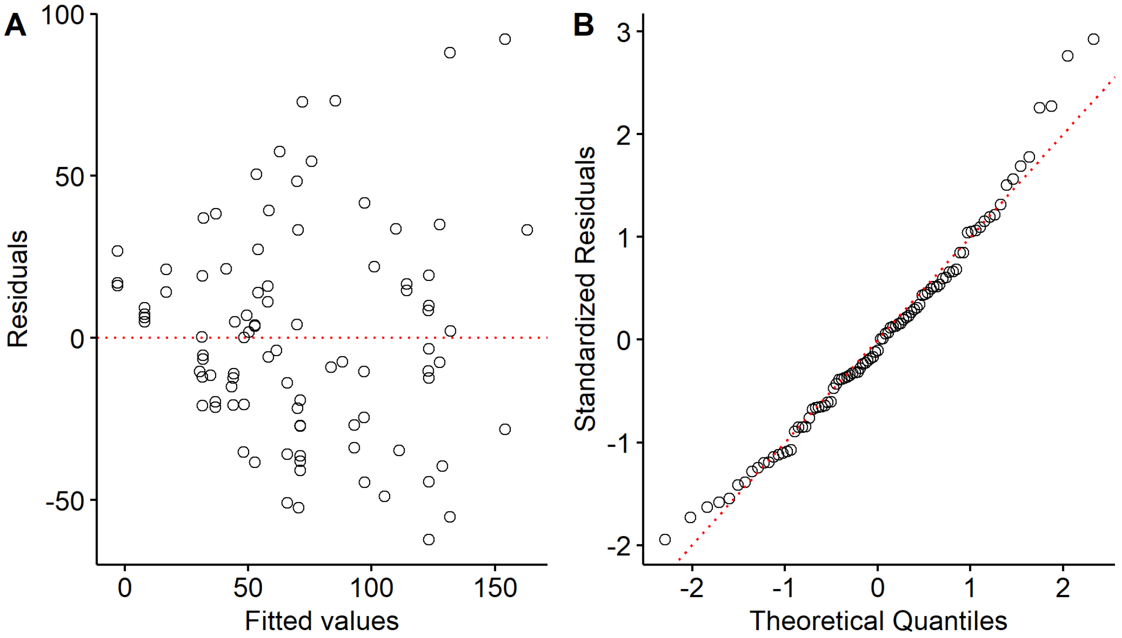


Source: Author's production.

Table 7.3 - Results of linear regressions for secondary forests considering only the stand age as the explanatory variable and also other environmental variables subjected to a stepwise selection.

Variable	Coeff.	CI 95%	Std. Error	t-value	p-value	R^2
Simple regression for secondary forests (n = 92)						
Age	4.86	[4.41, 5.31]	0.23	21.55	<2E-16	0.84
Multivariable regression for secondary forests (n = 92)						
Age	4.42	[3.56, 5.29]	0.43	10.18	1.4E-16	0.86
CWD	0.07	[0.04; 0.11]	0.02	4.06	1.0E-04	
MAP	0.01	[0.01; 0.02]	0.00	3.29	1.4E-03	
Simple regression for secondary forests with CWD < -400 (n = 28)						
Age	2.83	[2.31, 3.35]	0.25	11.19	1.2E-11	0.82
Simple regression for secondary forests with CWD > -400 (n = 64)						
Age	5.63	[5.15, 6.10]	0.24	23.62	<2E-16	0.90

Figure 7.6 - Distribution of residuals versus fitted values (A) and QQ plot (B) for the multivariable regression with secondary forests.



Source: Author's production.

7.4 Discussion

We found that anthropogenic disturbance had a fundamental impact on AGB variability in the Brazilian Amazon. Deforestation is the major source of carbon emissions in tropical forests (ARAGÃO et al., 2014). However, once the land use is discontinued, secondary forests show a high resilience to regenerate biomass over time (POORTER et al., 2016). Thus, the age of vegetation regrowth was the main factor controlling the amount of biomass following deforestation. In contrast, for mature forests, the time since last disturbance did not show a significant effect on AGB. Instead, the type of disturbance was mainly responsible for biomass variations. This is due to the great heterogeneity of forest damage produced by different types of degradation, also being related to the severity and recurrence of degradation events. Longo et al. (2016) analyzed the variability of AGB derived from inventory data and LiDAR-based models in the Brazilian Amazon and observed that disturbances of different intensities and recurrences, such as reduced-impact logging and multiple fires, often had similar disturbance age. Therefore, the type of degradation was more important than the time since last disturbance for explaining AGB variability.

Moreover, the patterns of AGB recovery can also depend on the disturbance type. A meta-analysis on aboveground carbon trajectories through time after logging and fire in tropical forests showed that logging emissions were concentrated at the beginning of the disturbance event (ANDRADE et al., 2017). However, fire emissions tend to peak several years after the disturbance event, as trees take longer to die due to fire damage and subsequent synergistic impacts such as drought and diseases. Additionally, we observed that some burned forests, previously logged or not, had very low AGB stocks that were more similar to AGB from earlier second-growth than undisturbed forests. Other studies in the Brazilian Amazon have also indicated that anthropogenic disturbances caused by fires engender a secondarization process that reduces carbon stocks even after a long period of the degradation event (BARLOW; PERES, 2008; BERENGUER et al., 2014; RAPPAPORT et al., 2018; SILVA et al., 2018; XAUD et al., 2013). These findings reveal the high impact of fire on carbon loss and persistence, highlighting the urgent need to avoid fires in tropical forests.

Nonetheless, the effect of anthropogenic disturbances on the AGB depletion in mature forests remains largely unaccounted for in carbon emission estimates. We provided an overview effect of anthropogenic disturbances on mature Amazonian forests that, on average, stored up to 44% less AGB than undisturbed forests. This result is consistent with those reported by Berenguer et al. (2014) based on ground data, in which forests disturbed by both selective logging and understory fires had, on average, 40% less aboveground carbon than undisturbed forests. Similarly, Longo et al. (2016) verified that forests disturbed by conventional logging showed an AGB depletion of about 42% compared to intact forests.

Besides the anthropogenic factors, environmental conditions related to topography and climate act as pivotal drivers of tropical forests AGB. Topography constrains soil nutrient and hydraulic conditions, playing an important control over the AGB of mature tropical forests (ASNER et al., 2009b; BERENGUER et al., 2014; JUCKER et al., 2018b; QUESADA et al., 2011). Here, terrain roughness was an important variable for explaining AGB variability in mature forests, especially in the undisturbed ones. Low biomass values associated with low terrain roughness suggest that this topographic variable may act as a proxy for two important factors: variation of soil properties and occurrence of flooded areas. However, the effect of topography on AGB decreased in areas with a higher predominance of anthropogenic disturbances, playing no significant effect in secondary forests. As secondary forests recover, becoming more similar to undisturbed forests, the topography is expected to play an important role in AGB variability.

Regarding climatic conditions, CWD affected the variability of AGB in both mature and secondary forests, suggesting a greater recovery potential in less water-stressed areas. In mature forests of Venezuela, Vilanova et al. (2018) verified that water deficit and turnover rates are key drivers of forest biomass. In areas where the turnover rates are low, mostly because of shorter dry seasons, the forests tend to have higher AGB with stands dominated by medium to high wood density species. The synergistic effect of extreme droughts and fire incidence in Amazonian mature forests has been inducing increased tree mortality and a shift in forest composition to pioneer species with lower wood density, thus sequestering less carbon (ARAGÃO et al., 2014; BARLOW; PERES, 2008; SILVA et al., 2018).

In secondary forests, the AGB recovery rate in areas with lower water deficit (5.6 $Mg.ha^{-1}.yr^{-1}$) was twice as high as in water-stressed areas (2.8 $Mg.ha^{-1}.yr^{-1}$). The results found by Poorter et al. (2016) corroborate our findings in which both higher rainfall and lower seasonal water deficit increase the potential AGB of secondary forests. However, the average AGB recovery rate of 6.1 $Mg.ha^{-1}.yr^{-1}$ found by Poorter et al. (2016) was more similar to the rate of the less water-stressed areas found here. The results of Poorter et al. (2016) were based on compiled data from chronosequence studies in Neotropical secondary forests, of which three were carried out in the Brazilian Amazon (JUNQUEIRA et al., 2010; VIEIRA et al., 2003; WILLIAMSON et al., 2012). It is important to note that the age range of our study (up to 32 years, with 75% of the samples with up to 22 years old) is more limited than that considered by Poorter et al. (2016) (up to 100 years). Moreover, the recovery rates calculated here cannot be extrapolated to older ages, as AGB growth tends to decrease with advanced ages. Nevertheless, our estimate of AGB recovery rate considering all secondary forest samples (4.4 $Mg.ha^{-1}.yr^{-1}$) was consistent with that found by Lennox et al. (2018) across 20 years of succession, of 4.5 $Mg.ha^{-1}.yr^{-1}$.

The results found here have important implications in the context of climate change, given the expected increase in drought conditions in the Amazon forests (MALHI et al., 2009), already observed in a few locations over the Brazilian Amazon (ALMEIDA et al., 2017). More intense and frequent droughts may affect the functioning of Amazonian forests, by increasing tree mortality (PHILLIPS et al., 2010), vulnerability to forest fires, and consequently, exacerbating carbon emissions (ARAGÃO et al., 2014). Moreover, those drier conditions potentially threaten biomass recovery following natural or anthropogenic disturbances.

Other factors not considered in this study may also influence the AGB variability, potentially interacting in complex ways. For example, soil fertility (BECKNELL; POWERS, 2014), intensity of prior land use (JAKOVAC et al., 2015), and surrounding landscape forest cover (BERENGUER et al., 2014) has been reported as important factors affecting tropical forest structure and resilience. Therefore, future studies that consider these factors may expand the understanding of the variability of AGB in the Brazilian Amazon.

7.5 Conclusion

In this chapter, we used airborne LiDAR and HSI data to estimate AGB of forests under different anthropogenic and environmental conditions over the Brazilian Amazon. Based on this multisensor-derived AGB, we concluded that:

(1) Anthropogenic disturbances significantly reduced the AGB compared to undisturbed forests. Forest degradation accounted for an AGB depletion of up to 44%.

(2) In mature forests, the occurrence of anthropogenic disturbances had the greatest effect on AGB variability, especially those caused by fires, conventional logging, or both logging and fire. Topography and climatic water deficit also displayed a significant effect. However, time since last disturbance had no significant effect on AGB variation of mature forests.

(3) Secondary forest age was the major factor explaining AGB recovery following deforestation, with a rate of 4.4 Mg.ha^{-1}.yr^{-1}. However, the AGB resilience of secondary forests depended on climatic water deficit and rainfall. Therefore, it is expected that the AGB recovery potential will vary according to the gradient of water availability in the Amazon.

These results were consistent with other field-based studies, highlighting the potential of multisensor data to capture AGB variations due to anthropogenic and environmental factors over the complex forests of the Brazilian Amazon.

8 CONCLUDING REMARKS

The combined analysis of results from Chapters 5, 6 and 7 showed some common and interesting aspects. For instance, the HSI and LiDAR metrics that had the greatest influence on the classification of forest disturbance status were also the most important for estimating AGB. This is because AGB integrates important forest structural and functional information associated with forest disturbance and regrowth, such as tree height, basal area, number of trees per area, and wood density. Moreover, climatic conditions related to water availability also showed a great relevance for the AGB variation, which was evidenced by the importance of HSI metrics related to canopy moisture and water stress. This suggests that HSI metrics are sensitive to physiological and compositional characteristics due to adaptation to drier climate as well as changes in forest functioning such as water regulation given microclimatic changes generated by anthropogenic disturbances. Those results highlight the relevance of AGB as a measure of forest ecosystem functioning and the usefulness of integrating different remote sensing data for its characterization.

The moist Amazonian forests showed high AGB resilience overtime after deforestation. Thus, the recovery of secondary forests from initial to advanced successions has great potential to increase the carbon reservoir of the Amazon. However, the recovery of highly degraded forests, mainly by fire, is uncertain and may lead to an opposite pathway, from mature forests to earlier successional stages.

Most studies on forest ecosystems based on LiDAR and HSI data integration have been performed locally, on single study sites (e.g., ANDERSON et al., 2008; DALPONTE et al., 2012; JONES et al, 2010; THOMAS et al., 2008). Although it is important to consider different spatial scales in the study of multisensor data integration, the full realization of its potential as a source of forest information requires an ability to generalize in different environmental conditions and human-induced disturbance dynamics. Thus, the integrated use of LiDAR and HSI data can also help to understand the dynamics of complex Amazonian forests from a regional perspective.

As an overall conclusion of this thesis, the combined use of LiDAR and hyperspectral remote sensing data significantly improved the discrimination of forests subjected to different types of anthropogenic disturbances (Chapter 5) and the estimation of aboveground biomass in the Brazilian Amazon (Chapter 6). The gain of information provided by the multisensor data is highly required to support strategies of conservation of priority areas, sustainable management, and reduction of greenhouse gas emissions as well as improving our understanding of forest dynamics in the face of increasing human pressure and climate change risks. For instance, integrating LiDAR and HSI data provided a unique opportunity to assess the relative importance of environmental and anthropogenic factors affecting forest aboveground biomass across a broad set of sites (Chapter 7). Therefore, the use of information derived from multiple remote sensing sources has proved useful for the study of the highly biodiverse and complex Amazon forests.

REFERENCES

ALMEIDA, C. T. et al. Spatiotemporal rainfall and temperature trends throughout the Brazilian Legal Amazon, 1973-2013. **International Journal of Climatology**, v.37, p.2013-2026, 2017.

ANDERSON, J. E. et al. Integrating waveform lidar with hyperspectral imagery for inventory of a northern temperate forest. **Remote Sensing of Environment**, v. 112, n. 4, p. 1856–1870, 2008.

ANDRADE, R. B. et al. Scenarios in tropical forest degradation: carbon stock trajectories for REDD +. **Carbon Balance and Management**, v. 12, n. 6, p. 1–7, 2017.

APAN, A. et al. Detecting sugarcane “orange rust” disease using EO-1 Hyperion hyperspectral imagery. **International Journal of Remote Sensing**, v. 25, n. 2, p. 489–498, 2004.

ARAGÃO, L. E. O. C. et al. Environmental change and the carbon balance of Amazonian forests. **Biological Reviews**, v. 89, n. 4, p. 913–931, 2014.

ASNER, G. P. Biophysical and biochemical sources of variability in canopy reflectance. **Remote Sensing of Environment**, v.64, p.134−153, 1998.

ASNER, G. P. et al. Drought stress and carbon uptake in an Amazon forest measured with spaceborne imaging spectroscopy. **PNAS**, v. 101, n. 16, p. 6039–6044, 2004.

ASNER, G. P. et al. A contemporary assessment of change in humid tropical forests. **Conservation Biology**, v. 23, n. 6, p. 1386–1395, 2009a.

ASNER, G. P. et al. Environmental and biotic controls over aboveground biomass throughout a tropical rain forest. **Ecosystems**, v. 12, p. 261–278, 2009b.

ASNER, G. P. et al. Quantifying forest canopy traits: Imaging spectroscopy versus field survey. **Remote Sensing of Environment**, v. 158, p. 15–27, 2015.

ASNER, G. P.; MARTIN, R. E. Spectral and chemical analysis of tropical forests: scaling from leaf to canopy levels. **Remote Sensing of Environment**, v. 112, n. 10, p. 3958–3970, 2008.

ASNER, G. P.; MASCARO, J. Mapping tropical forest carbon: calibrating plot estimates to a simple LiDAR metric. **Remote Sensing of Environment**, v. 140, p. 614–624, 2014.

BAJWA, S. G.; KULKARNI, S. S. Hyperspectral data mining. In: THENKABAIL, P. S.; LYON, J. G.; HUETE, A. R. (Ed.). **Hyperspectral remote sensing of vegetation**. Boca Raton, FL: CRC Press, 2011. p. 93–120.

BAKER, T. R. et al. Increasing biomass in Amazonian forest plots. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 359, n. 1443, p. 353–365, 2004.

BALDECK, C. A. et al. Operational tree species mapping in a diverse tropical forest with airborne imaging spectroscopy. **PLoS ONE**, v. 10, n. 7, p. 21, 2015.

BARBIER, N.; COUTERON, P. Attenuating the bidirectional texture variation of satellite images of tropical forest canopies. **Remote Sensing of Environment**, v. 171, p. 245–260, 2015.

BARLOW, J.; PERES, C. A. Fire-mediated dieback and compositional cascade in an Amazonian forest. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 363, n. 1498, p. 1–8, 2008.

BASAK, D.; PAL, S.; PATRANABIS, D. C. Support vector regression. **Neural Information Processing: Letters and Reviews**, v. 11, n. 10, p. 203–224, 2007.

BECKNELL, J. M.; POWERS, J. S. Stand age and soils as drivers of plant functional traits and aboveground biomass in secondary tropical dry forest. **Canadian Journal of Forest Research**, v. 44, p. 604–613, 2014.

BELGIU, M.; DRAGUT, L. Random forest in remote sensing: a review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 114, p. 24–31, 2016.

BERENGUER, E. et al. A large-scale field assessment of carbon stocks in human-modified tropical forests. **Global Change Biology**, v. 20, p. 3713–3726, 2014.

BIGDELI, B.; SAMADZADEGAN, F.; REINARTZ, P. Fusion of hyperspectral and LIDAR data using decision template-based fuzzy multiple classifier system. **International Journal of Applied Earth Observation and Geoinformation**, v. 38, p. 309–320, 2015.

BISPO, P. D. C. et al. Mapping forest successional stages in the Brazilian Amazon using forest heights derived from TanDEM-X SAR interferometry. **Remote Sensing of Environment**, v. 232, p. 111194, 2019.

BLACKARD, J. A. et al. Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information. **Remote Sensing of Environment**, v. 112, n. 4, p. 1658–1677, 2008.

BLACKBURN, G. A. Remote sensing of forest pigment using airborne imaging spectrometer and LIDAR imagery. **Remote Sensing of Environment**, v. 82, p. 311–321, 2002.

BONAN, G. B. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. **Science**, v. 320, n. 5882, p. 1444–1449, 2008.

BOSCHETTI, M.; BOSCHETTI, L.; OLIVERI, S.; CASATI, L.; CANOVA, I. Tree species mapping with airborne hyper-spectral MIVIS data: the Ticino Park study case. **International Journal of Remote Sensing**, v.28, p.1251-1261, 2007.

BOUVIER, M. et al. Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data. **Remote Sensing of Environment**, v. 156, p. 322–334, 2015.

BOYD, D. S.; DANSON, F. M. Satellite remote sensing of forest resources: three decades of research development. **Progress in Physical Geography**, v. 29, n. 1, p. 1–26, 2005.

BREIMAN, L. E. O. Random forest. **Machine Learning**, v. 45, p. 5–32, 2001.

BROADBENT, E. N. et al. Linking rainforest ecophysiology and microclimate through fusion of airborne LiDAR and hyperspectral imagery. **Ecosphere**, v. 5, n. 5, art57, 2014.

BROGE, N. H.; LEBLANC, E. Comparing prediction power and stability of broadband and hyperspectral indices for estimation of green leaf area index and canopy chlorophyll density. **Remote Sensing of Environment**, v.76, p. 156–172, 2000.

CAO, L. et al. Integrating airborne LiDAR and optical data to estimate forest aboveground biomass in arid and semi-arid regions of China. **Remote Sensing**, v. 10, 2018.

CARREIRAS, J. M. B. et al. Mapping major land cover types and retrieving the age of secondary forests in the Brazilian Amazon by combining single-date optical and radar remote sensing data. **Remote Sensing of Environment**, v. 194, p. 16–32, 2017.

CARREIRAS, J. M. B.; VASCONCELOS, M. J.; LUCAS, R. M. Understanding the relationship between aboveground biomass and ALOS PALSAR data in the forests of Guinea-Bissau (West Africa). **Remote Sensing of Environment**, v. 121, p. 426–442, 2012.

CASTILLO, M. et al. LIDAR remote sensing for secondary tropical dry forest identification. **Remote Sensing of Environment**, v. 121, p. 132–143, 2012.

CHAVE, J. et al. Towards a worldwide wood economics spectrum. **Ecology Letters**, v. 12, n. 4, p. 351–366, 2009.

CHAVE, J. et al. Improved allometric models to estimate the aboveground biomass of tropical trees. **Global Change Biology**, v. 20, n. 10, p. 3177–3190, 2014.

CHEN, Q. et al. Integration of airborne lidar and vegetation types derived from aerial photography for mapping aboveground live biomass. **Remote Sensing of Environment**, v. 121, p. 108–117, 2012.

CHEN, Q. LiDAR remote sensing of vegetation biomass. In: WANG, G.; WENG, Q. (Ed.). **Remote sensing of natural resources.** Boca Raton: CRC Press, 2014. cap.21, p399-420.

CHIRICI, G. et al. Stochastic gradient boosting classification trees for forest fuel types mapping through airborne laser scanning and IRS LISS-III imagery. **International Journal of Applied Earth Observation and Geoinformation**, v. 25, n. 1, p. 87–97, 2013.

CLARK, M. L. et al. Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. **Remote Sensing of Environment**, v. 115, n. 11, p. 2931–2942, 2011.

CLARK, M. L.; KILHAM, N. E. Mapping of land cover in northern California with simulated hyperspectral satellite imagery. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 119, p. 228–245, 2016.

CLARK, M. L.; ROBERTS, D. A.; CLARK, D. B. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. **Remote Sensing of Environment**, v. 96, n. 3–4, p. 375–398, 2005.

CLARK, R. N.; ROUSH, T. L. Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications. **Journal of Geophysical Research**, v. 89, n. B7, p. 6329–6340, 1984.

COHEN, J. **Statistical power analysis for behavioural sciences**. 2.ed. Hillsdale, NJ: Erlbaum, 1988.

CURRAN, P. J.; KUPIEC, J. A.; SMITH, G. M. Remote sensing the biochemical composition of a slash pine canopy. **IEEE Transactions on Geoscience and Remote Sensing**, v.35, p.415−420, 1997.

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.

DA SILVA, R. D. et al. Spectral/textural attributes from ALI/EO-1 for mapping primary and secondary tropical forests and studying the relationships with biophysical parameters. **GIScience & Remote Sensing**, v. 51, n. 6, p. 677–694, 2014.

DALPONTE, M. et al. Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. **IEEE Transactions on Geoscience and Remote Sensing**, v. 46, n. 5, p. 1416–1427, 2008.

DALPONTE, M.; BRUZZONE, L.; GIANELLE, D. Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. **Remote Sensing of Environment**, v. 123, p. 258–270, 2012.

DAMODARAN, B. B.; COURTY, N.; LEFÈVRE, S. Sparse Hilbert Schmidt independence criterion and Surrogate-Kernel-based feature selection for hyperspectral image classification. **IEEE Transactions on Geoscience and Remote Sensing**, v. 55, n. 4, p. 2385–2398, 2017.

DE JONG, S. M.; PEBESMA, E. J.; LACAZE, B. Above-ground biomass assessment of Mediterranean forests using airborne imaging spectrometry: the DAIS Peyne experiment. **International Journal of Remote Sensing**, v. 24, n. 7, p. 1505–1520, 2003.

DE MOURA, Y. M. et al. Spectral analysis of amazon canopy phenology during the dry season using a tower hyperspectral camera and modis observations. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 131, p. 52–64, 2017.

DIRZO, R.; RAVEN, P. H. Global state of biodiversity and loss. **Annual Review of Environment and Resources**, v. 28, p. 137–167, 2003.

DUZAN, H.; SHARIFF, N. S. B. M. Ridge regression for solving the multicollinearity problem: review of methods and models. **Journal of Applied Sciences**, v. 15, n. 3, p. 393–404, 2015.

ELITH, J.; LEATHWICK, J. R.; HASTIE, T. A working guide to boosted regression trees. **Journal of Animal Ecology**, v. 77, n. 4, p. 802–813, 2008.

EXBRAYAT, J.-F. et al. Using multi-model averaging to improve the reliability of catchment scale nitrogen predictions. **Geoscientific Model Developmen**, v. 6, p. 117–125, 2013.

FANG, H. et al. Leaf area index. In: LIANG, S.; LI, X.; WANG, J. (Ed.). **Advanced remote sensing:** terrestrial information extraction and applications. [S.l.]: Academic Press, 2012. cap.11, p.347-381.

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.

FELDPAUSCH, T. R. et al. Tree height integrated into pantropical forest biomass estimates. **Biogeosciences**, v. 9, n. 8, p. 3381–3403, 2012.

FENG, Y. et al. Examining effective use of data sources and modeling algorithms for improving biomass estimation in a moist tropical forest of the Brazilian Amazon. **International Journal of Digital Earth**, v. 10, n. 10, p. 996–1016, 2017.

FÉRET, J. B.; ASNER, G. P. Tree species discrimination in tropical forests using airborne imaging spectroscopy. **IEEE Transactions on Geoscience and Remote Sensing**, v. 51, n. 1, p. 73–84, 2013.

FERRAZ, A. A. et al. Lidar detection of individual tree size in tropical forests. **Remote Sensing of Environment**, v. 183, p. 318–333, 2016.

FICK, S. E.; HIJMANS, R. J. WorldClim 2 : new 1-km spatial resolution climate surfaces for global land areas. **International Journal of Climatology**, v. 37, n. 12, p. 4302–4315, 2017.

FILIPPI, A. M.; GÜNERALP, I.; RANDALL, J. Hyperspectral remote sensing of aboveground biomass on a river meander bend using multivariate adaptive regression splines and stochastic gradient boosting. **Remote Sensing Letters**, v. 5, n. 5, p. 432–441, 2014.

FRIEDMAN, J. H. Stochastic gradient boosting. **Computational Statistics & Data Analysis**, v. 38, p. 367–378, 2002.

GALVÃO, L. S. et al. Possibilities of discriminating tropical secondary succession in Amazônia using hyperspectral and multiangular CHRIS / PROBA data. **International Journal of Applied Earth Observation and Geoinformation**, v. 11, p. 8–14, 2009.

GALVÃO, L. S. et al. On intra-annual EVI variability in the dry season of tropical forest: a case study with MODIS and hyperspectral data. **Remote Sensing of Environment**, v. 115, n. 9, p. 2350–2359, 2011.

GALVÃO, L. S. et al. Crop type discrimination using hyperspectral data. In: THENKABAIL, P. S.; LYON, J. G.; HUETE, A. R. (Ed.). **Hyperspectral remote sensing of vegetation**. Boca Raton, FL: CRC Press, 2012. cap. 17, p. 397-421.

GALVÃO, L. S. et al. View-illumination effects on hyperspectral vegetation indices in the Amazonian tropical forest. **International Journal of Applied Earth Observations and Geoinformation**, v. 21, p. 291–300, 2013.

GALVÃO, L. S.; FORMAGGIO, A. R.; TISOT, D. A. Discrimination of sugarcane varieties in southeastern Brazil with EO-1 Hyperion data. **Remote Sensing of Environment**, v. 94, n. 4, p. 523–534, 2005.

GAMON, J. A.; PEÑUELAS, J.; FIELD, C. B. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. **Remote Sensing of Environment**, v. 41, p. 35–44, 1992.

GAO, B. A normalized difference water index for remote sensing of vegetation liquid water from space. **Remote Sensing of Environment**, v. 58, n. 3, p. 257-266, 1996.

GEERLING, G. W. et al. Classification of floodplain vegetation by data fusion of spectral (CASI) and LiDAR data. **International Journal of Remote Sensing**, v. 28, n. 19, p. 4263–4284, 2007.

GHOSH, A. et al. A framework for mapping tree species combining hyperspectral and LiDAR data: role of selected classifiers and sensor across three spatial scales. **International Journal of Applied Earth Observation and Geoinformation**, v. 26, n. 1, p. 49–63, 2014.

GIBSON, L. et al. Primary forests are irreplaceable for sustaining tropical biodiversity. **Nature**, v. 478, n. 7369, p. 378–381, 2011.

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

GITELSON, A. A.; KAUFMAN, Y. J.; MERZLYAK, M. N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. **Remote Sensing of Environment**, v.58, p.289–298, 1996.

GITELSON, A. A.; KAUFMAN, Y. J.; STARK, R.; RUNDQUIST, D. Novel algorithms for remote estimation of vegetation fraction. **Remote Sensing of Environment**, v. 80, n. 1, p. 76–87, 2002.

GITELSON, A. A.; KEYDAN, G. P.; MERZLYAK, M. N. Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. **Geophysical Research Letters**, v. 33, n. 11, p. 2–6, 2006.

GODINHO, S.; GUIOMAR, N.; GIL, A. Using a stochastic gradient boosting algorithm to analyse the effectiveness of Landsat 8 data for montado land cover mapping: application in southern Portugal. **International Journal of Applied Earth Observation and Geoinformation**, v. 49, p. 151–162, 2016.

GOETZ, A. F. H. et al. Imaging spectrometry for earth remote sensing. **Science**, v. 228, n. 4704, p. 1147–1153, 1985.

GOLDBERGS, G. et al. Hierarchical integration of individual tree and area-based approaches for savanna biomass uncertainty estimation from airborne LiDAR. **Remote Sensing of Environment**, v. 205, p. 141–150, 2018.

GONG, P.; PU, R.; BIGING, G. S.; LARRIEU, M. R. Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 41, p. 1355–1362, 2003.

GRUSSU, G. et al. Optimum plot and sample sizes for carbon stock and biodiversity estimation in the lowland tropical forests of Papua New Guinea. **Forestry**, v. 89, p. 150–158, 2016.

GUARIGUATA, M. R.; OSTERTAG, R. Neotropical secondary forest succession: changes in structural and functional characteristics. **Forest Ecology and Management**, v. 148, p. 185–206, 2001.

GUYON, I.; WESTON, J.; BARNHILL, S.; VAPNIK, V. Gene selection for cancer classification using support vector machines. **Machine Learning,** v. 46, p. 389–422, 2002.

GUYOT, G.; BARET, F. Utilisation de la haute resolution spectrale pour suivre l'etat des couverts vegetaux. In: GUYENNE, T. D.; HUNT, J. J. (Ed.). **Spectral signatures of objects in remote sensing**. Paris: ESA, 1988. p. 279-286.

HAKALA, T.; SUOMALAINEN, J.; KAASALAINEN, S.; CHEN, Y. Full waveform hyperspectral LiDAR for terrestrial laser scanning. **Optics Express**, v. 20, p. 7119-7127, 2012.

HANSEN, E. H. et al. Modeling aboveground biomass in dense tropical submontane rainforest using airborne laser scanner data. **Remote Sensing**, v. 7, n. 1, p. 788–807, 2015.

HEINZEL, J.; KOCH, B. Investigating multiple data sources for tree species classification in temperate forest and use for single tree delineation. **International Journal of Applied Earth Observation and Geoinformation**, v. 18, n. 1, p. 101–110, 2012.

HOUGHTON, R. A.; HALL, F.; GOETZ, S. J. Importance of biomass in the global carbon cycle. **Journal of Geophysical Research: Biogeosciences,** v. 114, p. 1–13, 2009.

HU, X. et al. Combining models is more likely to give better predictions than single models. **Phytopathology**, v. 105, p. 1174–1182, 2015.

HUETE, A. R. et al. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote Sensing of Environment**, v. 83 n. 1–2, p. 195–213, 2002.

HUNTER, M. O.; KELLER, M.; VICTORIA, D.; MORTON, D. C. Tree height and tropical forest biomass estimation. **Biogeosciences**, v. 10, p. 8385–8399, 2013.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - INPE. **Desmatamento – Amazônia Legal.** Disponível em: http://terrabrasilis.dpi.inpe.br/downloads/. Acesso em: 05 jan. 2019.

INTERNATIONAL PANEL ON CLIMATE CHANGE - IPCC. **Climate change and land**: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. Genebra: IPCC, 2019.

INTERNATIONAL TROPICAL TIMBER ORGANIZATION - ITTO. **ITTO guidelines for the restoration, management and rehabilitation of degraded and secondary tropical forests**. Yokohama, Japan: ITTO, 2002.

ISENBURG, M. **LAStools** - efficient LiDAR processing software (version 171030, unlicensed). 2018. Available from: http://rapidlasso.com/LAStools.

JAKOVAC, C. C. et al. Loss of secondary-forest resilience by land-use intensification in the Amazon. **Journal of Ecology**, v. 103, p. 67–77, 2015.

JONES, T. G.; COOPS, N. C.; SHARMA, T. Assessing the utility of airborne hyperspectral and LiDAR data for species distribution mapping in the coastal Pacific Northwest, Canada. **Remote Sensing of Environment**, v. 114, n. 12, p. 2841–2852, 2010.

JOHN, R. et al. Grassland canopy cover and aboveground biomass in Mongolia and Inner Mongolia: Spatiotemporal estimates and controlling factors. **Remote Sensing of Environment**, v. 213, p. 34–48, 2018.

JORDAN, C. F. Derivation of leaf-area index from quality of light on the forest floor. **Ecology**, v.50, p.663-666, 1969.

JUCKER, T. et al. Estimating aboveground carbon density and its uncertainty in Borneo's structurally complex tropical forests using airborne laser scanning. **Biogeosciences**, v. 15, n. 12, p. 3811–3830, 2018a.

JUCKER, T. et al. Topography shapes the structure, composition and function of tropical forest landscapes. **Ecology Letters**, v. 21, n. 7, p. 989–1000, 2018b.

JUNQUEIRA, A. B.; SHEPARD-JR., G. H.; CLEMENT, C. R. Secondary forests on anthropogenic soils in Brazilian Amazonia conserve agrobiodiversity. **Biodiversity and Conservation**, v. 19, p. 1933–1961, 2010.

KARATZOGLOU, A. et al. kernlab – an S4 package for Kernel methods in R. **Journal of Statistical Software**, v. 11, n. 9, p. 1–20, 2004.

KOCH, B. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 65, n. 6, p. 581–590, 2010.

KOETZ, B. et al. Multi-source land cover classification for forest fire management based on imaging spectrometry and LiDAR data. **Forest Ecology and Management**, v. 256, n. 3, p. 263–271, 2008.

KOKALY, R. F. et al. Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. **Remote Sensing of Environment**, v. 113, n. SUPPL. 1, p. S78–S91, 2009.

KOTTEK, M. et al. World map of the Köppen-Geiger climate classification updated. **Meteorologische Zeitschrift**, v. 15, n. 3, p. 259–263, 2006.

KRONSEDER, K. et al. Above ground biomass estimation across forest types at different degradation levels in central kalimantan using lidar data. **International Journal of Applied Earth Observation and Geoinformation**, v. 18, n. 1, p. 37–48, 2012.

KUHN, M. Building predictive models in R using the caret package. **Journal of Statistical Software**. v. 28, n. 5, p. 1-26, 2008.

LARGE, A. R. G.; HERITAGE, G. L. Laser scanning: evolution of the discipline. In: HERITAGE, G. L.; LARGE, A. R. G. (Ed.). **Laser scanning for the environmental sciences**. [S.l.]: Wiley-Blackwell, 2009. cap.1, p.1-20.

LATIFI, H.; FASSNACHT, F.; KOCH, B. Forest structure modeling with combined airborne hyperspectral and LiDAR data. **Remote Sensing of Environment**, v. 121, p. 10–25, 2012.

LAWRENCE, R. et al. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. **Remote Sensing of Environment**, v. 90, n. 3, p. 331–336, 2004.

LE MAIRE, G. et al. Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. **Remote Sensing of Environment**, v. 112, n. 10, p. 3846–3864, 2008.

LE QUÉRÉ, C. et al. Global carbon budget 2018**. Earth System Science Data Discussions,** v. 10, p. 2141–2194, 2018.

LEFSKY, M. A. et al. Lidar remote sensing for ecosystem studies. **BioScience**, v. 52, n. 1, p. 19–30, 2002a.

LEFSKY, M. A. et al. Lidar remote sensing of above-ground biomass in three biomes. **Global Ecology and Biogeography**, v. 11, n. 5, p. 393–399, 2002b.

LEHNERT, L. W.; MEYER, H.; BENDIX, J. hsdar: manage, analyse and simulate hyperspectral data in R. R package version 0.7.1. **Journal of Statistical Software**, v.89, n.12, p.1-23, 2018.

LENNOX, G. D. et al. Second rate or a second chance? assessing biomass and biodiversity recovery in regenerating Amazonian forests. **Global Change Biology**, v. 24, n. 12, p. 5680–5694, 2018.

LIANG, S.; WANG, J.; JIANG, B. A systematic view of remote sensing. In: LIANG, S.; LI, X.; WANG, J. (Ed.). **Advanced remote sensing:** terrestrial information extraction and applications. [S.l.]: Academic Press, 2012. cap.1, p.1-31.

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

LU, D. et al. Classification of successional forest stages in the Brazilian Amazon basin. **Forest Ecology and Management**, v. 181, n. 3, p. 301–312, 2003.

LU, D. et al. Aboveground forest biomass estimation with Landsat and LiDAR data and uncertainty analysis of the estimates. **International Journal of Forestry Research**, v.2012, 16 p., 2012.

LU, D. et al. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. **International Journal of Digital Earth**, v. 9, n. 1, p. 63–105, 2014.

LUO, S. et al. Fusion of airborne LiDAR data and hyperspectral imagery for aboveground and belowground forest biomass estimation. **Ecological Indicators**, v. 73, p. 378–387, 2017a.

LUO, S. et al. Retrieving aboveground biomass of wetland Phragmites australis (common reed) using a combination of airborne discrete-return LiDAR and hyperspectral data. **International Journal of Applied Earth Observation and Geoinformation**, v. 58, p. 107–117, 2017b.

MAGURRAN, A. E. **Measuring biological diversity**. [S.l.]: Blackwell Science, 2004.

MALHI, Y. et al. Exploring the likelihood and mechanism of a climate-change-induced dieback of the Amazon rainforest. **Proceedings of the National Academy of Sciences**, v. 106, n. 49, p. 20610–20615, 2009.

MAN, Q. et al. Light detection and ranging and hyperspectral data for estimation of forest biomass: a review. **Journal of Applied Remote Sensing**, v. 8, n. 1, p. 81598, 2014.

MANQI, L. et al. Forest biomass and carbon stock quantification using airborne LiDAR data: a case study over Huntington Wildlife Forest in the Adirondack Park. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 7, n. 7, p. 3143–3156, 2014.

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 and Management**, v. 10, n. 1, p. 1–14, 2015.

MCGAUGHEY, R .J. **FUSION/LDV**: software for LIDAR data analysis and visualization, manual. Seattle: USFS Pacific Northwest Research Station, 2014.

MERTON, R. N. Monitoring community hysteresis using spectral shift analysis and the red-edge vegetation stress index. In: ANNUAL JPL AIRBORNE EARTH SCIENCE WORKSHOP, 7., 1998. **Proceedings…** NASA, 1998.

MERZLYAK, M. N. et al. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. **Physiologia Plantarum**, v. 106, n. 1, p. 135–141, 1999.

MONNET, J. M.; CHANUSSOT, J.; BERGER, F. Support vector regression for the estimation of forest stand parameters using airborne laser scanning. **IEEE Geoscience and Remote Sensing Letters**, v. 8, n. 3, p. 580–584, 2011.

MORAN, E. F. et al. Effects of soil fertility and land-use on forest succession in Amazônia. **Forest Ecology and Management**, v. 139, p. 93–108, 2000.

MOUNTRAKIS, G.; IM, J.; OGOLE, C. Support vector machines in remote sensing: a review. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 66, n. 3, p. 247–259, 2011.

MURA, M. D. et al. Challenges and opportunities of multimodality and data fusion in remote sensing. **Proceedings of the IEEE**, v. 103, n. 9, p.1585-1601, 2015.

NÆSSET, E. Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. **Remote Sensing of Environment**, v. 113, n. 1, p. 148–159, 2009.

NÆSSET, E.; GOBAKKEN, T. Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. **Remote Sensing of Environment**, v. 112, n. 6, p. 3079–3090, 2008.

NAGLER, P. L.; DAUGHTRY, C. S. T.; GOWARD, S. N. Plant litter and soil reflectance. **Remote Sensing of Environment**, v. 71, n. 2, p. 207–215, 2000.

NAIDOO, L. et al. Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a random forest data mining environment. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 69, p. 167–179, 2012.

NELSON, R. et al. Lidar-based estimates of aboveground biomass in the continental US and Mexico using ground, airborne, and satellite observations. **Remote Sensing of Environment**, v. 188, p. 127–140, 2017.

NILSSON, M. Estimation of tree heights and stand volume using an airborne lidar system. **Remote Sensing of Environment**, v. 56, n. 1, p. 1–7, 1996.

NIU, Z.; YAN, C. Canopy biochemical characteristics. In: LIANG, S.; LI, X.; WANG, J. (Ed.). **Advanced remote sensing:** terrestrial information extraction and applications. [S.l.]: Academic Press, 2012. cap.10, p.301-346.

NOBRE, C. A. et al. Land-use and climate change risks in the Amazon and the need of a novel sustainable development paradigm. **Proceedings of the National Academy of Sciences**, v. 113, n. 39, p. 10759–10768, 2016.

NUMATA, I. Characterization on pastures using field and imaging spectrometers. In: THENKABAIL, P. S.; LYON, J. G.; HUETE, A. (Ed.). **Hyperspectral remote sensing of vegetation**. Boca Raton: CRC Press, 2012. cap. 9, p. 207-225.

OMETTO, J. P. et al. Amazon forest biomass density maps: Tackling the uncertainty in carbon emission estimates. **Climatic Change**, v. 124, n. 3, p. 545–560, 2014.

OSBORNE, J.; WATERS, E. Four assumptions of multiple regression that researchers should always test. **Practical Assessment, Research and Evaluation**, v. 8, n. 2, p. 1, 2002.

PAN, Y. et al. The structure, distribution, and biomass of the world’s forests. **Annual Review of Ecology, Evolution, and Systematics**, v. 44, n. 1, p. 593–622, 2013.

PEÑUELAS, J.; PINOL, J.; OGAYA, R.; LILELLA. I., Estimation of plant water content by the reflectance water index WI (R900/R970). **International Journal of Remote Sensing**, v. 18, p. 2869–2875, 1997.

PHILLIPS, O. L. et al. Drought-mortality relationships for tropical forests. **New Phytologist**, v. 187, p. 631–646, 2010.

PLOTON, P. et al. Toward a general tropical forest biomass prediction model from very high resolution optical satellite images. **Remote Sensing of Environment**, v. 200, p. 140–153, 2017.

POHL, C.; VAN GENDEREN, J. L. Multisensor image fusion in remote sensing: concepts, methods and applications. **International Journal of Remote Sensing**, v. 19, n. 5, p. 823-854, 1998.

POORTER, L. et al. Biomass resilience of neotropical secondary forests. **Nature**, v. 530, n. 7589, p. 211–214, 2016.

PLOURDE, L. C.; OLLINGER, S. V.; SMITH, M.; MARTIN, M. E. Estimating species abundance in a northern temperate forest using spectral mixture analysis. **Photogrammetric Engineering and Remote Sensing**, v.73, p.829-840, 2007.

PSOMAS, A. et al. Hyperspectral remote sensing for estimating aboveground biomass. **International Journal of Remote Sensing**, v. 32, n. 24, p. 9007–9031, 2011.

PU, R. et al. Using CASI hyperspectral imagery to detect mortality and vegetation stress associated with a new hardwood forest disease. **Photogrammetric Engineering & Remote Sensing**, v. 74, n. 1, p. 65–75, 2008.

PU, R. Tree species classification. In: WANG, G.; WENG, Q. (Ed.). **Remote sensing of natural resources.** Boca Raton: CRC Press, 2014. cap.14, p. 239-258.

PUTZ, F. E.; REDFORD, K. H. The importance of defining ‘forest’: tropical forest degradation, deforestation, long-term phase shifts, and further transitions. **Biotropica**, v. 42, n. 1, p. 10–20, 2010.

QUESADA, C. A. et al. Soils of Amazonia with particular reference to the RAINFOR sites. **Biogeosciences**, v. 8, n. 6, p. 1415–1440, 2011.

RAPPAPORT, D. I. et al. Quantifying long-term changes in carbon stocks and forest structure from Amazon forest degradation. **Environmental Research Letters**, v. 13, n. 6, p. 065013, 2018.

RÉJOU-MÉCHAIN, M. et al. Using repeated small-footprint LiDAR acquisitions to infer spatial and temporal variations of a high-biomass neotropical forest. **Remote Sensing of Environment**, v. 169, p. 93–101, 2015.

RÉJOU-MÉCHAIN, M. et al. Biomass: an R package for estimating above-ground biomass and its uncertainty in tropical forests. **Methods in Ecology and Evolution**, v. 8, n. 9, p. 1163–1167, 2017.

ROTH, K. L. et al. The impact of spatial resolution on the classification of plant species and functional types within imaging spectrometer data. **Remote Sensing of Environment**, v. 171, p. 45–57, 2015.

ROUSE, J.W. et al. Monitoring vegetation systems in the great plains with ERTS. In: ERTS-1 SYMPOSIUM, n. 3, Washington, DC. **Proceedings**…Washington: NASA, 1973. p. 309-317.

ROUSSEL, J. R., AUTY, D. **lidR**: airborne LiDAR data manipulation and visualization for forestry applications. R package version 1.6.1. 2018. Available from: https://CRAN.R-project.org/package=lidR.

RULEQUEST.. **Data mining with cubist**. 2018. Available from: https://www.rulequest.com/cubist-info.html.

RUTISHAUSER, E. et al. Rapid tree carbon stock recovery in managed Amazonian forests. **Current Biology**, v. 25, n. 18, p. R787–R788, 2015.

SAATCHI, S. S. et al. Distribution of aboveground live biomass in the Amazon basin. **Global Change Biology**, v. 13, p. 816–837, 2007.

SANCHES, I. D. A.; SOUZA FILHO, C. R.; KOKALY, R. F. Spectroscopic remote sensing of plant stress at leaf and canopy levels using the chlorophyll 680nm absorption feature with continuum removal. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 97, p. 111–122, 2014.

SCHEPERS, L. et al. Burned area detection and burn severity assessment of a heathland fire in belgium using airborne imaging spectroscopy (APEX). **Remote Sensing**, v. 6, n. 3, p. 1803–1826, 2014.

SERRANO, L.; PEÑUELAS, J.; USTIN, S. L. Remote sensing of nitrogen and lignin in mediterranean vegetation from AVIRIS data: decomposing biochemical from structural signals. **Remote Sensing of Environment**, v. 81, p. 355, 2002.

SERVIÇO FLORESTAL BRASILEIRO - SFB. **Madeflona industrial madeireira**: execução financeira e técnica da concessão (Jamari - UMF I), 2020. Available from: http://www.florestal.gov.br/florestas-sob-concessao/92-concessoes-florestais/florestas-sob-concessao/299-umf-i-madeflona-industrial-madeireira-execucao-financeira-e-tecnica-da-concessao-jamari-umf-i.

SHIPPERT, P. Introduction to hyperspectral image analysis. **Online Journal of Space Communication**, v. 3, p. 13, 2003.

SILVA, C. V. J. et al. Drought-induced Amazonian wildfires instigate a decadal-scale disruption of forest carbon dynamics. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 373, n. 1760, p. 20180043, 2018.

SINGH, K. K. et al. When big data are too much: effects of LiDAR returns and point density on estimation of forest biomass. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 9, n. 7, p. 3210–3218, 2016.

SMITH, M. L. et al. Direct estimation of aboveground forest productivity through hyperspectral remote sensing of canopy nitrogen. **Ecological Applications**, v. 12, n. 5, p. 1286–1302, 2002.

SMITH, M. L.; MARTIN, M. E.; PLOURDE, L.; OLLINGER, S. V. Analysis of hyperspectral data for estimation of temperate forest canopy nitrogen concentration: comparison between airborne (AVIRIS) and spaceborne (Hyperion) sensor. **IEEE Transaction of Geoscience and Remote Sensing**, v.41, p.1332−1337, 2003.

SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. **Information Processing and Management**, v. 45, n. 4, p. 427–437, 2009.

SOMMER, C. et al. Feature-based tree species classification using hyperspectral and Lidar data. **EARSeL eProceedings,** v.14, n. 2, p. 49–70, 2015.

STARK, S. C. et al. Amazon forest carbon dynamics predicted by profiles of canopy leaf area and light environment. **Ecology Letters**, v. 15, n. 12, p. 1406–1414, 2012.

STOVALL, A. E. L.; SHUGART, H. H. Improved biomass calibration and validation with terrestrial lidar: Implications for future LiDAR and SAR missions. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 11, n. 10, p. 3527–3537, 2018.

STRAND, J. et al. Spatially explicit valuation of the Brazilian Amazon Forest's Ecosystem Services. **Nature Sustainability**, v. 1, p. 657–664, 2018.

STRAUB, C.; WANG, Y.; IERCAN, O. Airborne laser scanning: methods for processing and automatic feature extraction for natural artificial objects. In: In: HERITAGE, G. L.; LARGE, A. R. G. (Ed.). **Laser scanning for the environmental sciences**. [S.l.]: Wiley-Blackwell, 2009. cap.8, p. 115-132.

SUN, C.; CAO, S.; SANCHEZ-AZOFEIFA, G. A. Mapping tropical dry forest age using airborne waveform LiDAR and hyperspectral metrics. **International Journal of Applied Earth Obsservation and Geoinformation**, v. 83, e101908, 2019.

SWATANTRAN, A. et al. Mapping biomass and stress in the Sierra Nevada using lidar and hyperspectral data fusion. **Remote Sensing of Environment**, v. 115, n. 11, p. 2917–2930, 2011.

THENKABAIL, P. S. et al. Hyperion, IKONOS, ALI, and ETM+ sensors in the study of African rainforests. **Remote Sensing of Environment**, v. 90, n. 1, p. 23–43, 2004.

THOMAS, V. et al. Mapping stand-level forest biophysical variables for a mixedwood boreal forest using lidar: an examination of scanning density. **Canadian Journal of Forest Research**, v. 36, n. 1, p. 34–47, 2006.

THOMAS, V. et al. Canopy chlorophyll concentration estimation using hyperspectral and lidar data for a boreal mixedwood forest in northern Ontario, Canada. International **Journal of Remote Sensing**, v. 29, n. 4, p. 1029–1052, 2008.

THOMAS, V. et al. Leaf area and clumping indices for a boreal mixed-wood forest: Lidar, hyperspectral, and Landsat models. **International Journal of Remote Sensing**, v. 32, n. 23, p. 8271–8297, 2011.

TORABZADEH, H.; MORSDORF, F.; SCHAEPMAN, M. E. Fusion of imaging spectroscopy and airborne laser scanning data for characterization of forest ecosystems - A review. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 97, p. 25–35, 2014.

TYUKAVINA, A. et al. Pan-tropical hinterland forests: mapping minimally disturbed forests. **Global Ecology and Biogeography**, v. 25, p. 151–163, 2016.

USTIN, S. L. et al. Using imaging spectroscopy to study ecosystem processes and properties. **BioScience**, v. 54, n. 6, p. 523, 2004.

VAGLIO LAURIN, G. et al. Above ground biomass estimation in an African tropical forest with lidar and hyperspectral data. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 89, p. 49–58, 2014.

VAUHKONEN, J. et al. **Forestry applications of airborne laser scanning – concepts and case studies**. Netherlands: Springer Science, 2014.

VIEIRA, I. C. G. et al. Classifying successional forests using Landsat spectral properties and ecological characteristics in eastern Amazônia. **Remote Sensing of Environmen**t, v. 87, n. 4, p. 470–481, 2003.

VILANOVA, E. et al. Environmental drivers of forest structure and stem turnover across Venezuelan tropical forests. **PLoS ONE**, v. 13, n. 6, p. 1–27, 2018.

VOGELMANN, J. E.; ROCK, B. N.; MOSS, D. M. Red edge spectral measurements from sugar maple leaves. **International Journal of Remote Sensing**, v.14, n. 8, p. 1563−1575, 1993.

WANG, H.; GLENNIE, C. Fusion of waveform LiDAR data and hyperspectral imagery for land cover classification. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 108, p. 1–11, 2015.

WANG, Y. et al. Remote sensing of environment mapping tropical disturbed forests using multi-decadal 30 m optical satellite imagery. **Remote Sensing of Environment**, v. 221, p. 474–488, 2019.

WATSON, J. E. M. et al. The exceptional value of intact forest ecosystems. **Nature Ecology and Evolution**, v. 2, n. 4, p. 599–610, 2018.

WEI, G.; SHALEI, S.; BO, Z.; SHUO, S.; FAQUAN, L.; XUEWU, C. Multi-wavelength canopy LiDAR for remote sensing of vegetation: design and system performance. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 69, p. 1–9, 2012.

WILLIAMSON, G. B. et al. Convergence and divergence in alternative successional pathways in Central Amazonia. **Plant Ecology & Diversity**, v. 7, n. 1–2, p. 341–348, 2012.

WILSON, M. F. J. et al. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. **Marine Geodesy**, v. 30, p. 3–35, 2007.

XAUD, H. A. M.; MARTINS, F. DA S. R. V.; SANTOS, J. R. DOS. Forest ecology and management tropical forest degradation by mega-fires in the northern Brazilian Amazon. **Forest Ecology and Management**, v. 294, p. 97–106, 2013.

XIAO, J. et al. Remote sensing of the terrestrial carbon cycle : a review of advances over 50 years. **Remote Sensing of Environment**, v. 233, p. 37, 2019.

YANG, C.; EVERITT, J. H.; FLETCHER, R. S.; JENSEN, R. R.,; MAUSEL, P. W. Evaluating AISA+ hyperspectral imagery for mapping black mangrove along the south Texas Gulf Coast. **Photogrammetric Engineering and Remote Sensing**, v.75, p.425-435, 2009.

ZANNE, A. E. et al. **Data from**: Towards a worldwide wood economics spectrum, Dryad Data Repository, 2009. doi:10.5061/dryad.234.

ZARCO-TEJADA, P. J.; et al. Scaling-up and model inversion methods with narrowband optical indices for chlorophyll estimation in closed forest canopies with hyperspectral data. **IEEE Transactions on Geoscience and Remote Sensing**, v.39, p.1491–1507, 2001.

ZHANG, J. Multi-source remote sensing data fusion: status and trends. **International Journal of Image and Data Fusion**, v. 1, n. 1, p. 5-24, 2010.

ZHANG, J.; YANG, J. Data Fusion. In: LIANG, S.; LI, X.; WANG, J. **Advanced remote sensing:** terrestrial information extraction and applications. [S.l.]: Academic Press, 2012. cap.4, p.91-109.

ZHANG, C.; SELCH, D.; COOPER, H. A Framework to combine three remotely sensed data sources for vegetation mapping in the Central Florida Everglades. **Wetlands**, v. 36, n. 2, p. 201–213, 2016.

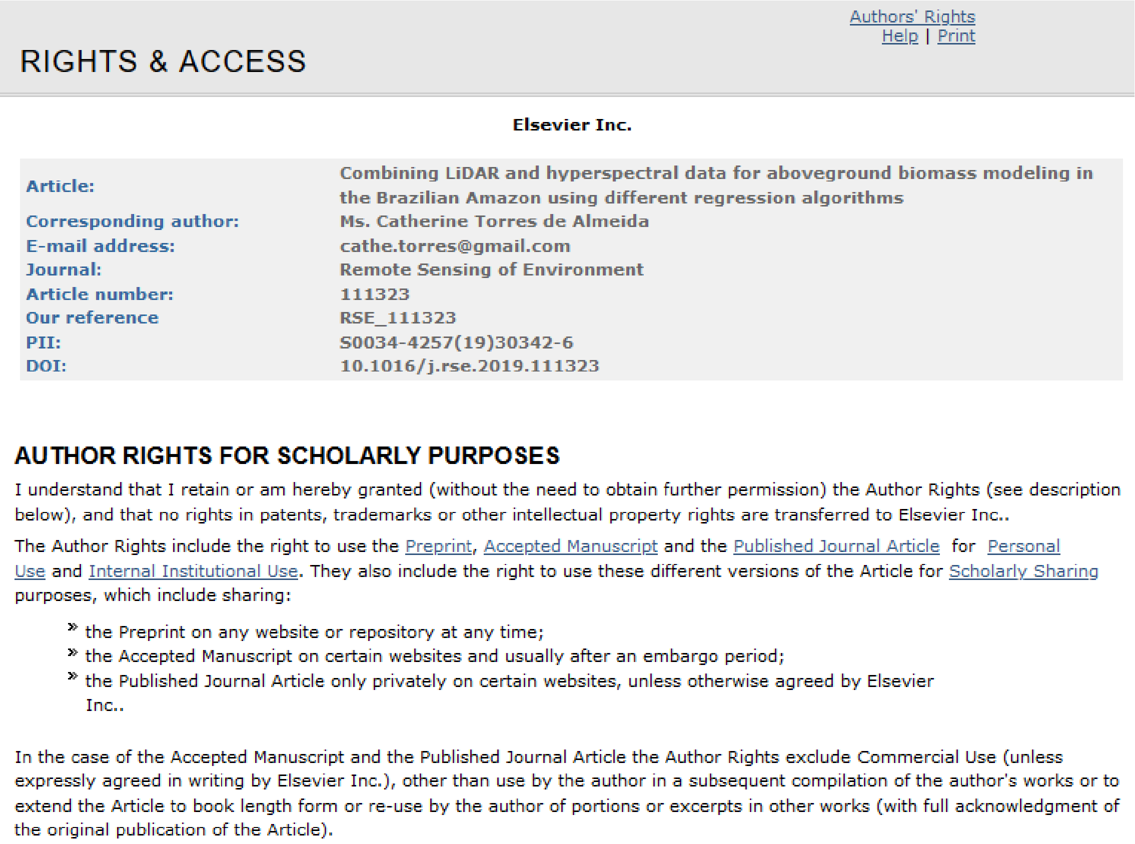
ZHANG, W. et al. Characterizing forest succession stages for wildlife habitat assessment using multispectral airborne imagery. **Forests**, v. 8, n. 7, 2017.

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

ZHAO, K. et al. Utility of multitemporal lidar for forest and carbon monitoring: tree growth, biomass dynamics, and carbon flux. **Remote Sensing of Environment**, v. 204, n. August 2017, p. 883–897, 2018.

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

APPENDIX A – AUTHOR RIGHTS FOR SCHOLARLY PURPOSES



RIGHTS & ACCESS

Elsevier Inc.

Article:	**Combining LiDAR and hyperspectral data for aboveground biomass modeling in the Brazilian Amazon using different regression algorithms**
Corresponding author:	**Ms. Catherine Torres de Almeida**
E-mail address:	**cathe.torres@gmail.com**
Journal:	**Remote Sensing of Environment**
Article number:	**111323**
Our reference	**RSE_111323**
PII:	**S0034-4257(19)30342-6**
DOI:	**10.1016/j.rse.2019.111323**

AUTHOR RIGHTS FOR SCHOLARLY PURPOSES

I understand that I retain or am hereby granted (without the need to obtain further permission) the Author Rights (see description below), and that no rights in patents, trademarks or other intellectual property rights are transferred to Elsevier Inc..

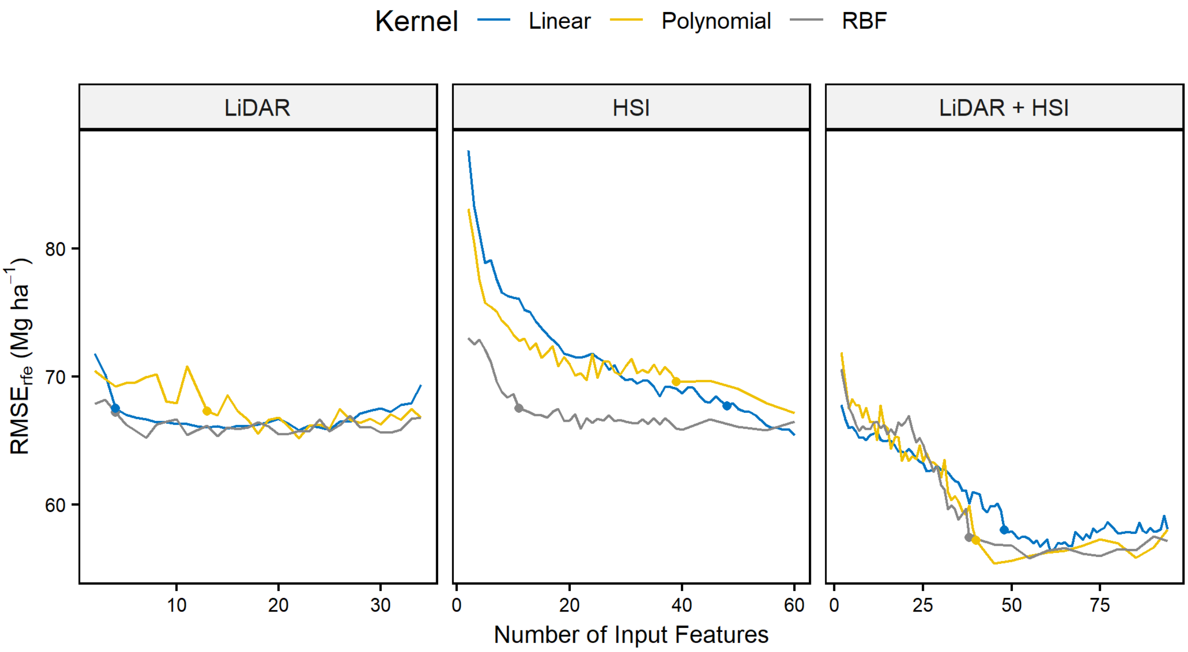
The Author Rights include the right to use the Preprint, Accepted Manuscript and the Published Journal Article for Personal Use and Internal Institutional Use. They also include the right to use these different versions of the Article for Scholarly Sharing purposes, which include sharing:

- the Preprint on any website or repository at any time;
- the Accepted Manuscript on certain websites and usually after an embargo period;
- the Published Journal Article only privately on certain websites, unless otherwise agreed by Elsevier Inc..

In the case of the Accepted Manuscript and the Published Journal Article the Author Rights exclude Commercial Use (unless expressly agreed in writing by Elsevier Inc.), other than use by the author in a subsequent compilation of the author's works or to extend the Article to book length form or re-use by the author of portions or excerpts in other works (with full acknowledgment of the original publication of the Article).

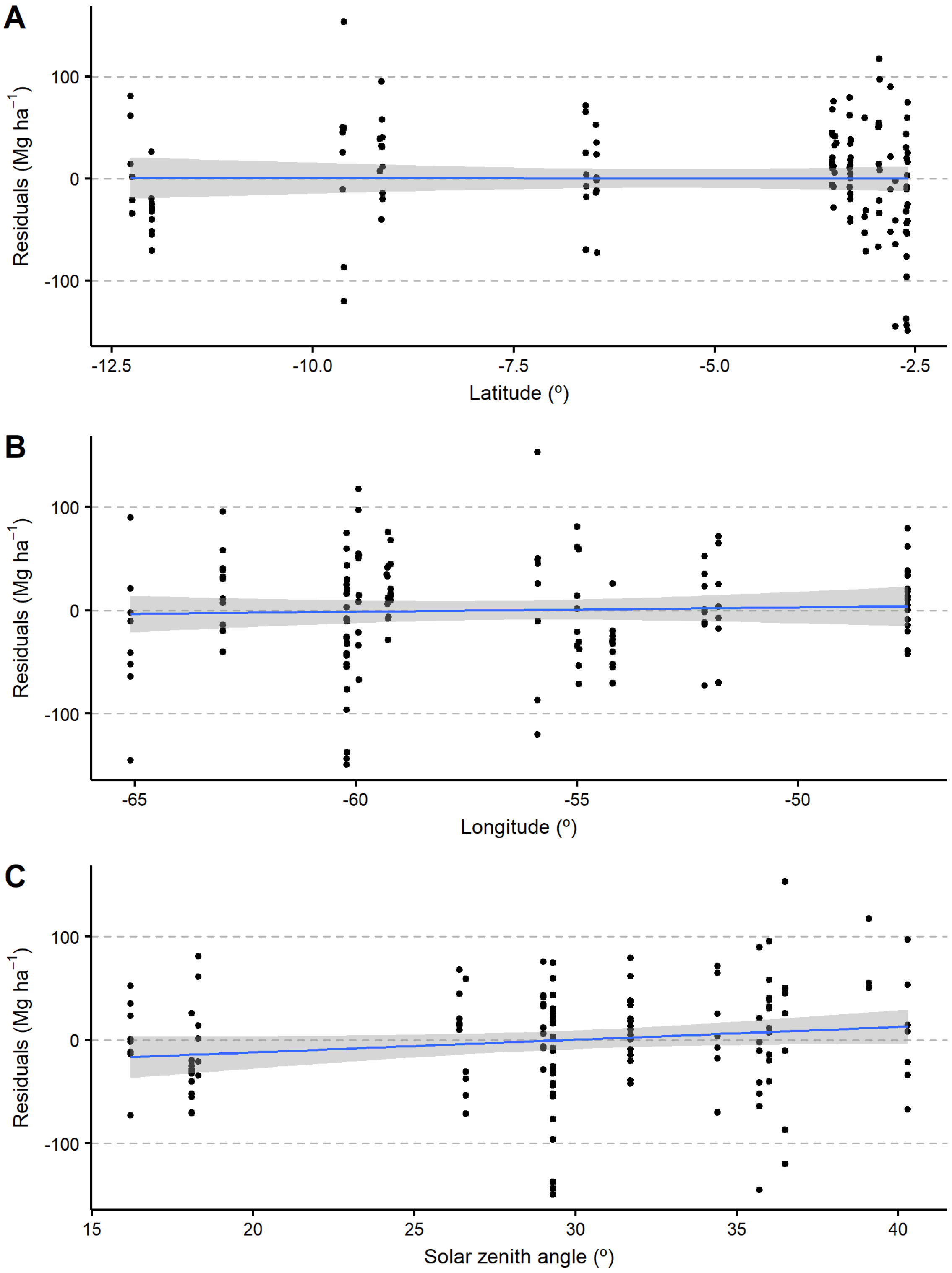
APPENDIX B - SUPPLEMENTARY FIGURES AND TABLES

Figure B.1 - Comparison of the feature selection results for the SVR (Support Vector Regression) method with three different kernels: linear, polynomial and RBF (Radial Basis Function). The points represent the selected feature size. The RBF kernel was chosen because it produced the smallest $RMSE_{rfe}$ values with the lowest number of metrics selected.



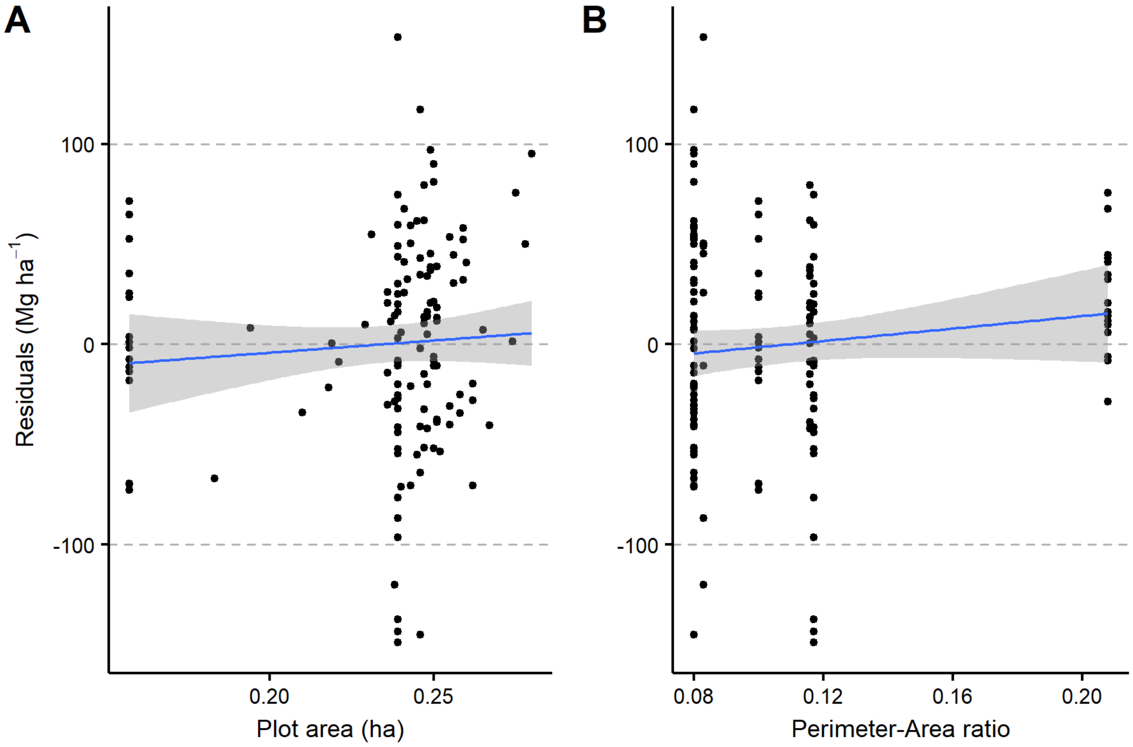
Source: Author's production.

Figure B.2 - Analysis of the residuals from the best model (LMR with LiDAR plus hyperspectral data), as a function of field plot location (latitude in A and longitude in B) and of the solar zenith angle at the time of the hyperspectral data acquisition (C). The blue line represents a linear fit and the gray area represents its 95% confidence interval.



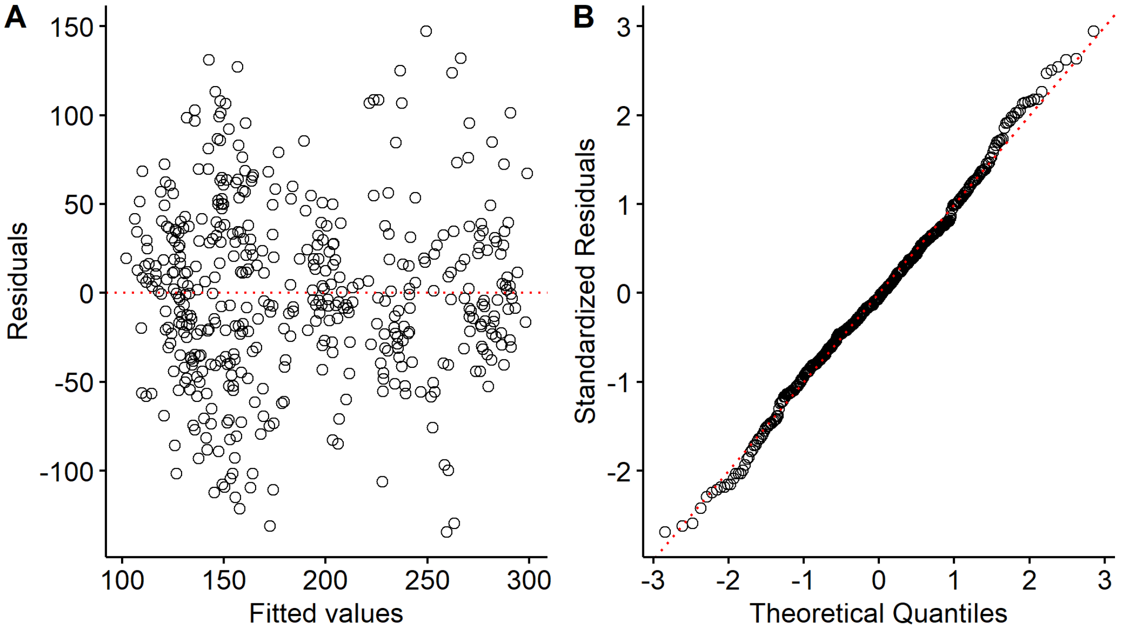
Source: Author's production.

Figure B.3. Analysis of the residuals from the best model (LMR with LiDAR plus hyperspectral data), as a function of the field plot area (A) and perimeter-area ratio (B). The blue line represents a linear fit and the gray area represents its 95% confidence interval.



Source: Author's production.

Figure B.4 - Distribution of residuals versus fitted values (A) and QQ plot (B) for the multivariable regression with the mature forests, with exception of the MAM site.



Source: Author's production.

Table B.1. LiDAR metrics selected by RFE (Recursive Feature Elimination) for each regression algorithm (LM: Linear Model, LMR: Linear Model with Regularization, SVR: Support Vector Regression, RF: Random Forest, SGB: Stochastic Gradient Boosting, and CB: Cubist). Metrics in bold have relative importance greater than 80%. Model tuning selected the following parameters: *lambda* 0.01 for LMR; *cost* 4 and *sigma* 0.05 for SVR; *n.trees* 100 for SGB; and *committees* 10 for CB.

Rank	LiDAR Model					
	LM	*LMR*	*SVR*	*RF*	*SGB*	*CB*
1	**PD_{2_10}**	**LAD_{20_30}**	**H.p05**	**LAD_{20_30}**	**LAD_{20_30}**	**LAD_{20_30}**
2	**LAD_{22}**	H.p95	**LAD_{20_30}**	**PD_{22}**	**H.p80**	PD_{22}
3			**LAD_{22}**	**LAD_{22}**	**PD_{1st}**	
4			**LAD_{18}**	**LAD_{26}**		
5				**H.p80**		

Abbreviations of the LiDAR metrics: H.pX= X^{th} (05, 80, or 95^{th}) percentile of height distribution of first returns above 2 m; PD_{2_10}= number of first returns between 2 and 10 m divided by the number of all first returns; PD_{22}= number of first returns above 22 m divided by the number of all first returns; PD_{1st}= number of first returns above 2m divided by the number of all returns above 2m; LAD_{20_30}= Leaf Area Density between 20 and 30 m; and LAD_h= Leaf Area Density above the height h (18, 22, or 26).

Table B.2. Hyperspectral (HSI) metrics selected by RFE (Recursive Feature Elimination) for each regression algorithm (LM: Linear Model, LMR: Linear Model with Regularization, SVR: Support Vector Regression, RF: Random Forest, SGB: Stochastic Gradient Boosting, and CB: Cubist). Metrics in bold have relative importance greater than 80%. Model tuning selected the following parameters: *lambda* 0.01 for LMR; *cost* 2 and *sigma* 0.018 for SVR; *n.trees* 50 for SGB; and *committees* 20 for CB.

Rank	HSI Model					
	LM	*LMR*	*SVR*	*RF*	*SGB*	*CB*
1	**LWVI1**	**W_{2100}**	**NDNI**	**W_{2100}**	**W_{2100}**	**W_{2100}**
2	R_{1646}	**NDNI**	W_{2100}	**W_{495}**	**W_{495}**	W_{495}
3	ND_{Bleaf}	LWVI1	As_{980}	NDNI	NDNI	W_{980}
4	LWVI2	As_{980}	W_{495}	D_{2100}	CAI	NDNI
5	R_{852}	W_{980}	LWVI1	R_{1091}	R_{1091}	LWVI1
6		W_{495}	S_{30_60}	CAI	D_{2100}	D_{LAI}
7		R_{461}	DWSI2	As_{980}	W_{980}	PSRI
8		ND_{Bleaf}	NDLI	PRI	D_{980}	S_{0_30}
9			D_{2100}			VI_{green}
10			As_{670}			R_{701}
11			S_{0_30}			As_{980}
12						GV

Abbreviations of the HSI metrics: R_{λ}= reflectance of a band centered at the wavelength λ (461, 701, 852, 1091, or 1646 nm); CAI= Cellulose Absorption Index; D_{LAI}= Difference for Leaf Area Index; DWSI2= Disease Water Stress Index 2; LWVI1= Leaf Water Vegetation Index 1; LWVI2= Leaf Water Vegetation Index 2; ND_{Bleaf}= Normalized Difference for Leaf Biomass; NDLI= Normalized Difference Lignin Index; NDNI= Normalized Difference Nitrogen Index; PRI= Photochemical Reflectance Index; PSRI= Plant Senescence Reflectance Index; VI_{green}= Vegetation Index green; D_c, W_c, and As_c = depth, width, and asymmetry, respectively, of the absorption band centered at 495, 670, 980, and 2100 nm; GV= mean green vegetation fraction; $S_{0\_30=}$ proportion of pixels with shade fraction below 30%; and $S_{30\_60=}$ proportion of pixels with shade fraction between 30 and 60%.

Table B.3 - Metrics selected by RFE (Recursive Feature Elimination) for each regression algorithm (LM: Linear Model, LMR: Linear Model with Regularization, SVR: Support Vector Regression, RF: Random Forest, SGB: Stochastic Gradient Boosting, and CB: Cubist) with hybrid datasets (LiDAR + hyperspectral). Metrics in bold have relative importance greater than 80%. Model tuning selected the following parameters: *lambda* 0.01 for LMR; *cost* 4 and *sigma* 0.018 for SVR; *n.trees* 150 for SGB; and *committees* 20 for CB.

Rank	Combined Model					
	LM	*LMR*	*SVR*	*RF*	*SGB*	*CB*
1	**R$_{1646}$**	**As$_{980}$**	**LAD$_{26}$**	**W$_{2100}$**	**W$_{2100}$**	**LAD$_{20\_30}$**
2	**H.p40**	**W$_{2100}$**	**LAD$_{22}$**	H.p95	NDNI	**H.mean**
3	**SR**	H.p95	H.p05	NDNI	R$_{1091}$	**W$_{980}$**
4	**As$_{980}$**	LAD$_{22}$	LAD$_{18}$	H.mean	H.p95	**LAD$_{22}$**
5	**R$_{1220}$**	PWI	LAD$_{20\_30}$	LAD$_{26}$	CAI	W$_{2100}$
6	**As$_{1200}$**	D$_{980}$	As$_{980}$	CAI	W$_{495}$	PD$_{22}$
7		As$_{670}$	H.p95	PD$_{22}$	H.p80	PD$_{14}$
8		PD$_{2}$	PD$_{22}$	LAD$_{22}$	W$_{980}$	As$_{980}$
9		LAD$_{20\_30}$	LAD$_{14}$	H.p80	LAD$_{26}$	H.p20
10		W$_{980}$	H.p10	W$_{495}$	PD$_{22}$	H.p95
11		As$_{2100}$	LAD$_{6}$	R$_{1091}$	D$_{2100}$	PD$_{18}$
12		PRI	H.mean	H.p90	H.mean	R$_{1091}$
13			NDNI	H.p10	H.p90	D$_{980}$
14			H.p20	D$_{2100}$	H.p10	D$_{2100}$
15			DSCI	LAD$_{20\_30}$	HSCI	H.p40
16			H.p90	LAD$_{18}$	ND$_{Bleaf}$	NDNI
17			H.p40	PD$_{18}$	PD$_{20\_30}$	LAD$_{18}$
18			H.p80	W$_{980}$	DSCI	W$_{495}$
19			H.max	PD$_{2\_10}$	PD$_{18}$	DWSI3
20			W$_{2100}$	ND$_{Bleaf}$	D$_{980}$	As$_{1200}$
21			PD$_{18}$	PD$_{20\_30}$	PD$_{6}$	PSRI
22			D$_{980}$	D$_{980}$	S$_{0\_30}$	
23			LAD$_{2}$	H.p05	PD$_{1st}$	
24			HSCI		PD$_{2\_10}$	
25			D$_{2100}$		H.p05	
26			R$_{1091}$		DWSI2	
27			PD$_{14}$		LAD$_{20\_30}$	

28	As_{1200}	S_{30_60}
29	S_{30_60}	H.p40
30	GV	
31	S_{0_30}	
32	D_{1200}	
33	PD_{2_10}	
34	R_{1220}	
35	PWI	
36	PD_{1st}	
37	PD_{10_20}	
38	R_{1646}	

Abbreviations of the metrics: H.mean= mean height of first returns above 2 m; H.max= maximum height; H.pX= X^{th} (05, 10, 20, 40, 80, 90, or 95^{th}) percentile of height distribution of first returns above 2 m; PD_h= number of first returns above a height h (2, 6, 14, 18, or 22) divided by the number of all first returns; PD_{a_b}= number of first returns between a height interval a_b (2_10, 10_20, or 20_30) divided by the number of all first returns; PD_{1st}= number of first returns above 2m divided by the number of all returns above 2m; LAD_h= Leaf Area Density above the height h (2, 6, 14, 18, 22, or 26); LAD_{20_30}= Leaf Area Density between 20 and 30 m; DSCI= Simpson Structural Complexity Index; HSCI= Shannon Structural Complexity Index; R_λ= reflectance of a band centered at the wavelength λ (1091, 1220, or 1646 nm); CAI= Cellulose Absorption Index; DWSI2= Disease Water Stress Index 2; DWSI3= Disease Water Stress Index 3; ND_{Bleaf}= Normalized Difference for Leaf Biomass; NDNI= Normalized Difference Nitrogen Index; PRI= Photochemical Reflectance Index; PSRI= Plant Senescence Reflectance Index; PWI= Plant Water Index; SR= Simple Ratio; D_c, W_c, and As_c = depth, width, and asymmetry, respectively, of the absorption band centered at 495, 670, 980, 1200, and 2100 nm; GV= mean green vegetation fraction; $S_{0\_30=}$ proportion of pixels with shade fraction below 30%; and $S_{30\_60=}$ proportion of pixels with shade fraction between 30 and 60%.