

ABSTRACT

Title of Dissertation: CHARACTERIZING TREE SPECIES DIVERSITY
IN THE TROPICS USING FULL-WAVEFORM
LIDAR DATA

Suzanne Mariëlle Marselis, Doctor of Philosophy,
2019

Dissertation Directed by: Professor Ralph Dubayah, Department of
Geographical Sciences

Tree species diversity is of paramount value to maintain forest health and to ensure that forests are able to provide all vital functions, such as creating oxygen, that are needed for mankind to survive. Most of the world's tree species grow in the tropical region, but many of them are threatened with extinction due to increasing natural and human-induced pressures on the environment. Mapping tree species diversity in the tropics is of high importance to enable effective conservation management of these highly diverse forests. This dissertation explores a new approach to mapping tree species diversity by using information on the vertical canopy structure derived from full-waveform lidar data. This approach is of particular interest in light of the recently launched Global Ecosystem Dynamics Investigation (GEDI), a full-waveform spaceborne lidar. First, successful derivation of vertical canopy structure metrics is

ensured by comparing canopy profiles from airborne lidar data to those from terrestrial lidar. Then, the airborne canopy profiles were used to map five successional vegetation types in Lopé National Park in Gabon, Africa. Second, the relationship between vertical canopy structure and tree species richness was evaluated across four study sites in Gabon, which enabled mapping of tree species richness using canopy structure information from full-waveform lidar. Third, the relationship between canopy structure and tree species richness across the tropics was established using field and lidar data collected in 16 study sites across the tropics. Finally, it was evaluated how the methods and applications developed here could be adapted and used for mapping pan-tropical tree species diversity using future GEDI lidar data products.

CHARACTERIZING TREE SPECIES DIVERSITY IN THE TROPICS USING FULL-WAVEFORM LIDAR DATA

By

Suzanne Mariëlle Marselis

Dissertation submitted to the faculty of the graduate school of the University of Maryland, College Park, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

2019

Advisory Committee:

Professor Ralph Dubayah, Chair
Professor Matthew Hansen
Dr. John Armston
Professor Matthew Fitzpatrick
Professor Joseph Sullivan

© Copyright by

Suzanne Mariëlle Marselis

2019

DEDICATION

To my grandmothers:

Tini Jonges & Leida Linssen

ACKNOWLEDGEMENTS

I would like to sincerely thank my adviser, Dr. Ralph Dubayah, who believed in me every step along the way. I thank you for inviting me to be part of the GEDI team and for allowing me to perform my dissertation study at the absolute forefront of spaceborne waveform lidar research. Thank you for teaching me about lidar and life.

To my committee members: Matt, Matt, John and Joe, thank you for your advice, guidance and critiques. Thank you for believing in me and for supporting me. Thank you for challenging me to always take one extra step.

I would like to express my sincerest gratitude to my beloved husband Daniel Teodoro, my supporting parents Rik and Liesbeth Marselis, and my caring sisters Nicoline and Ilonka Marselis. Without your love, patience and support I would have never been able to make it through this journey.

I would like to thank my colleagues from the University of Maryland Global Ecology Lab (GEL), the GEDI Science team members, and my friends in College Park and abroad for your support in good and bad times, your patience reading my work and your wonderful suggestions that helped me finish my dissertation.

This dissertation is supported by NASA Headquarters under the NASA Earth and Space Science Fellowship 414 Program – Grant 80NSSC17K0321.

CONTENTS

Dedication	ii
Acknowledgements	iii
Contents	iv
List of tables.....	viii
List of figures.....	ix
List of acronyms.....	xiii
List of publications.....	xiv
1. Introduction.....	1
1.1 Motivation.....	1
1.1.1 Global change and biodiversity	1
1.1.2 Conservation management.....	2
1.2 Background.....	3
1.2.1 Tropical forests	3
1.2.2 Function and diversity	4
1.2.3 Diversity mapping	5
1.2.4 Lidar for mapping diversity	7
1.3 Research objectives	9
1.4 Dissertation outline	10
2. Distinguishing vegetation types with airborne waveform lidar data in a tropical forest-savanna mosaic: a case study in Lopé National Park, Gabon.....	11
Abstract.....	11
2.1 Introduction	12
2.2 Methods.....	15
2.2.1 Study area & available datasets	15
2.2.2 Lidar data processing.....	21
2.2.3 Vegetation type classification.....	22
2.3 Results	24
2.3.1 PAI profile validation	24
2.3.2 Vegetation structure validation.....	25
2.3.3 Vegetation classification and spatial analysis.....	27

2.4 Discussion	31
2.4.1 Vegetation structural variability	31
2.4.2 Vegetation classification.....	33
2.5 Conclusion	37
Acknowledgements	38
3. Exploring the relation between remotely sensed vertical canopy structure and tree species diversity in Gabon	39
Abstract	39
3.1 Introduction	40
3.2 Method	43
3.2.1 Field data processing	43
3.2.2 Lidar data processing.....	49
3.2.3 Data analysis	50
3.3 Results	53
3.3.1 Modeling results using canopy height	53
3.3.2 Modeling results using canopy structure	54
3.3.3 Predictive modeling results.....	57
3.4 Discussion	59
3.4.1 Model performance.....	59
3.4.2 Limitations.....	61
3.4.3 Applications & future research.....	63
3.5 Conclusion	64
Acknowledgements	66
4. Evaluating the potential of full-waveform lidar for mapping pan-tropical tree species richness	68
Abstract	68
4.1 Introduction	69
4.2 Background	72
4.3 Methods	76
4.3.1 Field datasets	76
4.3.2 Lidar datasets	80
4.3.3 Canopy structure across the tropics	81
4.3.4 Species richness across the tropics	82
4.3.5 Structure-richness analysis	82
4.4 Results	86
4.4.1 Vertical forest structure across the tropics.....	86
4.4.2 Species-area relationships.....	88

4.4.3 Structure-richness models.....	88
4.5 Discussion.....	92
4.5.1 Structure-richness relationship across scales.....	92
4.5.2 Limitations.....	95
4.5.3 Future research & applications.....	97
4.6 Conclusion.....	99
Acknowledgements.....	100
5. Discussion.....	102
5.1 Synthesis.....	102
5.2 Limitations of the current study.....	105
5.3 Future directions: looking forward to data from GEDI.....	108
5.3.1 Mapping successional vegetation types.....	109
5.3.2 Estimating tree species diversity in Gabon.....	110
5.3.3 Estimating pan-tropical tree species richness.....	111
5.3.4 The future is fusion.....	112
5.3.5 Implications for ecosystem management.....	114
5.4 Final remarks.....	115
I. Supplementary material for chapter 2.....	116
I.1 TLS-LVIS comparison multiple scales.....	116
I.2 Vegetation structure metrics.....	119
I.3 Alternative classifications.....	122
II. Supplementary material for chapter 3.....	124
II.1 Dataset details.....	124
II.1.1 Study site description.....	124
II.1.2 Field data processing details.....	126
II.2 Model performance.....	129
II.3 Coefficient of variation.....	132
III. Supplementary material for chapter 4.....	133
III.1 Field data characteristics.....	133
III.2 Local model performance details.....	145
III.3 Ecological and structural distance.....	146
III.4 Detailed global modeling results.....	154
References.....	155

LIST OF TABLES

Table 1: Confusion matrix of pixel-based random forest classification using a total of 193 training pixels. Out-of-Bag estimate of the error rate is 18.7%.	29
Table 2: Information for each of the study sites regarding the total sampled area, temperature and precipitation, the tree density, species richness and Shannon diversity calculated including only trees with DBH ≥ 10 cm, at the 1 ha resolution.	48
Table 3: Sets of metrics used in linear models predicting H' and S.	52
Table 4: Information on the original plot size, the amount of total area sampled in the field and the source of the data which is either a website where the data is published and/or a publication in which the data is described.	78
Table 5: The total number of species identified for each study site and the average (\bar{x}) and standard deviation (s) of the species richness for each of the three plot sizes expressed in last three columns as $\bar{x} \pm s$ (including only live trees with DBH ≥ 10 cm).	79
Table 6: Datasets used for regional and pan-tropical analysis of the structure-richness relationships. Note that one region may not contain the same number of plots across all resolutions due to limitations in the availability of subplot and stem map information, limiting the use of data from some study sites to only one or two resolutions.	85
Table 7: Results for models predicting Shannon diversity.	130
Table 8: Results for models predicting species richness	131
Table 9: Number of plots included at each spatial resolution for each dataset with a percentage of trees identified up to genus level $> 80\%$	143

LIST OF FIGURES

Figure 1: Inset on left: location of data collection in Gabon. Left: canopy height collected by the LVIS instrument and location of study area. Magenta outline shows location of right panel. Right: Location of subset of field plots in the Lopé forest-savanna mosaic. 19

Figure 2: Workflow of study, grey boxes show main processes to reach study objectives. 20

Figure 3: Cumulative Plant Area Index (PAI) profiles from Terrestrial Laser Scanning (y-axis) and airborne lidar data (x-axis) in vertical intervals of 5m for eight plots covering four vegetation types. Each point represents the average cumulative PAI up to indicated height (m), calculated as the mean of all TLS scans around the plot edge (TLS) or all waveforms within the plot boundary (LVIS). Whiskers indicate PAI standard deviation of the group of scans (TLS) or waveforms (LVIS) at indicated height. 1:1 line shown in grey..... 24

Figure 4: Seven metrics (y-axis) describing characteristics of the vegetation structure of five vegetation types (colored), ordered by successional stage. Each boxplot is composed of the pixel values within all field plots covered by one specific vegetation type..... 26

Figure 5: PAI values of different vegetation height layers (left) combined into a Red-Green-Blue image (right), showing the variation in vertical vegetation structure across the study area. Zoom-in on right (location indicated with red box) reveals high structural variation over small scale. 27

Figure 6: Classification of the five successional vegetation types in the study area, using pixel-based Random Forest classification..... 29

Figure 7: Digital Terrain Model (DTM) from LVIS data (left) showing zoom-in area of other figures (red box). Stream network (6th order stream is shown as the Ogooué river) overlain with inner- and edge-Colonizing forest (middle). DTM, stream order and forest island classification led to the discrimination between bosquets and gallery forest within savannas (right)..... 30

Figure 8: Field and lidar datasets from four regions (Mondah, Mabounié, Rabi and Lopé) in central and west Gabon are used in this study. LVIS lidar acquisitions are displayed as gridded canopy height (m). Insets show distribution of field plots across each study site, markers are not to scale of field plot size..... 46

Figure 9: (a) Thirteen 1 ha plots (dark grey) were selected from the 25 ha Rabi plot and included in this study. (b) Subdivision of 1 ha plots to allow for analyzing the structure-diversity relationship at 5 spatial resolutions led to plots of sizes I. (100x100 m), II. (100x50 m), III. (50x50 m), IV. (25x25 m) and V. (20x20 m). 46

Figure 10: Canopy height explains up to 44% of the variation in tree species richness and 43% of the variation in Shannon diversity across the study sites in Gabon (cross-validated results). 54

Figure 11: Linear models using height alone as a predictive variable generally show the lowest R^2 and highest error (RMSD%). Adding variation in canopy height (RH100_sd) to the model leads to slightly better, but not statistically different, model performance (see overlapping error bars on the R^2 plots). Incorporating information on canopy structure significantly improves the S and H models at the 0.04 to 0.25 ha resolutions, when compared to using height alone. Error bars are larger at larger plot size because of smaller sample sizes. 55

Figure 12: Predicted vs. observed species richness and Shannon diversity resulting from the models using metric set 3 (PAVD profiles and height), points are colored by study site. 1.0 ha resolution models are significantly biased, 0.25 ha models show highest accuracy and statistically unbiased predictions for both species richness and Shannon diversity..... 56

Figure 13: Shannon diversity and species richness predicted in Mondah using canopy height (RH100) and standard deviation of canopy height (RH100_sd). Predictions created from gridded LVIS data products and a simulated GEDI-TanDEM-X fusion product. Points in center panel are colored by density. 57

Figure 14: ICESat data enabled predictions of H' and S within and between the study sites. ICESat richness predictions were biased slightly high compared to LVIS predictions. Points in center panel are colored by density..... 58

Figure 15: Location of field sites across the three continents, colors of each study site are consistent throughout paper. Gridlines indicate 10° intervals in longitudinal and latitudinal directions. The size of the place markers represents the relative size of the total sampled area. 77

Figure 16: Illustration of simulated GEDI waveform layout. The GEDI waveforms (red circles) have a Gaussian energy distributed with $\sigma=5.5$ m, resulting in a roughly 22 m diameter footprint. Example of simulated footprint distribution locations in a 1.0 (solid outline), 0.25 and 0.0625 ha field plot (dotted outline). 81

Figure 17: Canopy structure expressed as the Plant Area Volume Density profile (PAVD), expressing the Plant Area Index for each 1 m vertical bin, displayed as the median of all plots within each study site (solid line), the 30th-70th percentile (darker shaded area) and 10th-90th percentile (lighter shaded area)..... 87

Figure 18: Relationships between tree species richness and area for each study site (note the change in y-axis across panels from left to right). 88

Figure 19: Relation between canopy height (left) and total plant area index across three spatial scales for all study sites across the tropics. Each point represents one plot at the specific resolution. 89

Figure 20: Cross-validated results of local structure-richness models. Open circles indicate less than 95% of the cross-validated models was significant. The models in Panama are significant across all plot sizes, whereas models in Gabon are insignificant at smaller plot size. The RMSE% is low for predictions at each study site, but little relationship exists between the predicted and observed data. All models are unbiased. 90

Figure 21: Cross-validated model performance of regional structure-richness models. Error bars indicate the 95% range of values for each metric. 91

Figure 22: Cross-validated model performance at the global scale in terms of R², RMSE%, and bias. Error bars indicate the range between which 95% of the performance values of the cross-validated models fall. All models are statistically significant and unbiased. 92

Figure 23: Cumulative PAI profiles from TLS (y-axis) and LVIS (x-axis) at 3 m vertical interval. Overall correlation results of all 8 plots: r² = 0.95, RMSE = 0.65, bias = 0.41, CCC=0.94. 117

Figure 24: Cumulative PAI profiles from TLS (y-axis) and LVIS (x-axis) at 1 m vertical interval. At the 1 m interval profiles for the individual plots can be distinguished from the figure. Overall correlation results of all eight plots: r² = 0.95, RMSE = 0.61, bias = 0.38, CCC=0.94. 118

Figure 25: Eleven metrics (y-axis) describing characteristics of vegetation structure of the five vegetation types (colored), ordered by successional stage. Each boxplot is composed of the pixel values within all field plots covered by one specific vegetation type. PAI (m²/m²) profile is provided at a 5 m vertical interval. 120

Figure 26: Standard deviation calculated from a 3x3 window around each pixel, providing information on the variation of each vegetation metric for each vegetation type (ordered by successional stage). 120

Figure 27: Mean of each structural metric calculated from 3x3 window around each pixel. Variation in metrics for each class is lower than when using single-pixel values. 121

Figure 28: Coefficient of variation of each metric calculated from 3x3 window around each pixel. 121

Figure 29 Left: (a) Classification of vegetation types with random forest model build on all vegetation metrics. Right: (b) Classification of vegetation types using two-stage classification. Savanna, Colonizing forest and ‘other forest’ were classified using pixel-based vegetation metrics, after which a random forest model with all metrics was used to re-classify ‘other forest’ to Monodominant Okoumé, Marantaceae forest and Mixed forest.	123
Figure 30: Coefficient of Variation for Species richness, Canopy height and total PAI across all sampled plot sizes.	132
Figure 31: Predicted vs. Measured species richness from local scale predictions in <i>rab</i> (top row), <i>rob</i> (middle row) and <i>bci</i> (bottom row).	145
Figure 32: Ecological distance vs. spatial distance in <i>lsv</i> (Costa Rica)	146
Figure 33: Ecological distance vs. spatial distance in <i>cha</i> (Costa Rica).....	147
Figure 34: Ecological distance vs. spatial distance in <i>bci</i> (Panama)	147
Figure 35: Ecological distance vs. spatial distance in <i>rab</i> (Gabon).....	148
Figure 36: Ecological distance vs. spatial distance in <i>rob</i> (Australia).....	148
Figure 37: Ecological distance vs. spatial distance in <i>dan</i> (Malaysia)	149
Figure 38: Ecological distance vs. spatial distance in <i>sep</i> (Borneo).....	149
Figure 39: Ecological distance vs. spatial distance in <i>s11</i> (Brazil).....	150
Figure 40: Ecological distance vs. spatial distance in <i>s12</i> (Brazil).....	150
Figure 41: Ecological distance vs. spatial distance in <i>tam</i> (Peru).....	151
Figure 42: Ecological distance vs. spatial distance in <i>lop</i> (Gabon)	151
Figure 43: Ecological distance vs. spatial distance in <i>mon</i> (Gabon)	152
Figure 44: Ecological distance vs. spatial distance in <i>mab</i> (Gabon)	152
Figure 45: Ecological distance vs. spatial distance in <i>mal</i> (DRC).....	153
Figure 46: Ecological distance vs. spatial distance in <i>kea</i> (DRC)	153
Figure 47: Examples of observed vs. predicted tree species richness using cross-validated global models from one random Monte-Carlo simulation at each of the three spatial resolutions. Colors of points coincide with colors in Figure 15.....	154

LIST OF ACRONYMS

ACRONYM	DEFINITION
AGEOS	Gabonese Studies and Space Observations Agency
ALS	Airborne Laser Scanning
ANPN	Agence Nationale des Parcs Nationaux
BCI	Barro Colorado Island
BIC	Bayesian Information Criterion
CBD	Convention of Biological Diversity
CENAREST	Centre National de la Recherche Scientifique et Technologique
CP5	Cumulative Plant Area Index between 0-5 meter vertical, occurs in text with 5 m increments, from 5 to 50.
DAAC	Distributed Active Archive Center
DBH	Diameter at Breast Height
DLR	German Aerospace Center
DTM	Digital Terrain Model
EBV	Essential Biodiversity Variable
ESA	European Space Agency
GEDI	Global Ecosystem Dynamics Investigation
GEL	Global Ecology Lab
GEOBON	The Group on Earth Observations Biodiversity Observation Network
GLAS	Geoscience Laser Altimeter System
H'	Tree species diversity expressed by Shannon diversity index
ICESAT	Ice, Cloud, and land Elevation Satellite
LVIS	Land Vegetation and Ice Sensor
NASA	National Aeronautics and Space Administration
NESSF	NASA Earth and Space Science Fellowship
NP	National Park
ONERA	French Aeronautics Space and Defense Research Lab
ORNL	Oak Ridge National Laboratory
P5	Plant Area Index between 0-5 meter vertical, occurs in text with 5 m increments, from 5 to 50.
PAI	Plant Area Index
PAVD	Plant Area Volume Density
RH100	Relative Height at which 100% of the energy was returned
RH100_SD	Standard deviation of RH100
RMSD	Root Mean Squared Difference
S	Tree species richness
SAR	Synthetic Aperture Radar

SI	Supplementary Information
TANDEM-X	TerraSAR-X add-on for Digital Elevation Measurement
TLS	Terrestrial Laser Scanning

LIST OF PUBLICATIONS

The work carried out for this dissertation research has resulted in a number of peer-reviewed publications and conference posters and presentations which are all listed here, ordered by the chapter in which they are presented in this thesis.

Chapter 3

The contents of chapter 3 are published in Remote Sensing of Environment:

Marselis, S. M., Tang, H., Armston, J. D., Calders, K., Labrière, N., & Dubayah, R. (2018). Distinguishing vegetation types with airborne waveform lidar data in a tropical forest-savanna mosaic: A case study in Lopé National Park, Gabon. *Remote sensing of environment*, 216, 626-634. <https://doi.org/10.1016/j.rse.2018.07.023>

Part of the contents of Chapter 3 have been presented on the following conferences:

Marselis, S., Tang, H., Blair, J.B., Hofton, M.A., Armston, J., Dubayah, R. (2017). Characterizing Exterior and Interior Tropical Forest Structure Variability with Full-Waveform Airborne LIDAR Data in Lopé, Gabon. AGU Fall Meeting 2017, New Orleans, LA, USA. B21B-1961.

Marselis, S.M., Armston, J., Fitzpatrick, M., Dubayah, R. (2017). Characterization of vegetation structure at the tropical forest-savanna ecotone in Gabon. *Silvilaser 2017*, Blacksburg, Virginia, USA.

Chapter 4

The contents of chapter 4 are published in Environmental Research Letters:

Marselis, S.M., Tang, H., Armston, J., Abernethy, K, Alonso, A., Barbier, N., Bissiengou, P., Jeffery, K., Kenfack, D., Labrière, N., Lee, S.K., Lewis, S., Memiaghe, H., Poulsen, J.R., White, L., Dubayah, R. (2019). Exploring the relation between remotely sensed vertical canopy structure and tree species diversity in Gabon. Environmental Research Letters 14 094013 <https://doi.org/10.1088/1748-9326/ab2dcd>

Part of the content of Chapter 4 have been presented on the following conferences:

Marselis, S.M., Armston, J.D., Dubayah, R., the GEDI Science Definition Team. (2018). Predicting Tree Diversity with Full-Waveform Lidar Data in Gabon. AGU Fall Meeting 2018, Washington, DC, USA. B41L-2873.

Marselis, S.M., Tang, H., Armston, J., Dubayah, R. (2018). Predicting tree diversity with full-waveform lidar data in Gabon. ForestSAT, College Park, MD, USA. Poster-ID: 588.

Chapter 5

Part of the content of Chapter 5 have been presented on the following conference:

Marselis, S., Tang, H., Armston, J., Hancock, S., Minor, D., Duncanson, L.,
Fitzpatrick, M., Hansen, M., Dubayah, R. (2019) Exploring the potential of GEDI-
derived canopy structure for mapping tree species diversity in the tropics. Silvilaser
2019, Iguazu falls, Brazil. Presentation ID 26.

1. INTRODUCTION

Over the last decades, airborne and spaceborne remote sensing systems have revolutionized the way we can observe, study and understand the Earth's biosphere. Remote sensing allows us to gather information about remote regions at an unprecedented rate and to monitor the changes that are occurring rapidly across the face of the Earth. The continuous development of new methods to extract information from these new data is therefore relevant in the advancement of this field. In this dissertation, I investigate the use of full-waveform lidar data, a specific type of remote sensing data, for mapping tree species diversity in the tropics.

1.1 MOTIVATION

1.1.1 GLOBAL CHANGE AND BIODIVERSITY

Change is a fundamental characteristic of the Earth and it allows for the provision of essential ecosystem services such as climate regulation, disturbance prevention, fresh water supply, soil formation, pollination and waste recycling (de Groot et al., 2002). The Earth has been in a relatively stable state over the last 10,000 years, but global pressures have increased tremendously since the industrial revolution and have induced environmental change at unprecedented and alarming rate (Rockstrom et al., 2009) increasing glacier melting (Marzeion et al., 2014), environmental pollution (Vorosmarty et al., 2010), and deforestation (Hansen et al., 2013). Pressures on forests, both long-term stressors and short-term disturbances, are caused by human and natural drivers of change. Natural stressors, for example, can be changes in

precipitation or seasonality and the occurrence of extreme wind events affecting vegetation mortality and growth (Bonan, 2008). Humans can have a direct impact on vegetation, e.g. through deforestation for agriculture or urbanization, or afforestation through plantations, or an indirect impact, e.g. through dam construction which changes water supply downstream (Hansen et al., 2013; Ligon et al., 1995; Vitousek et al., 1997). In the last 300 years, 35% of pre-agricultural forest cover has been lost and 82% of the remaining forest is degraded (Watson et al., 2018). In the last century, species extinctions have increased to an estimated 1,000 – 10,000 times the background extinction rate, and habitat destruction occurs disproportionately in high-diversity tropical regions where small-ranged species live (Pimm et al., 2014). The research in this dissertation focusses specifically on tropical forests because of their extremely high diversity and vulnerability.

1.1.2 CONSERVATION MANAGEMENT

Even though we are currently able to map deforestation (Hansen et al., 2013) and estimate species extinction rates (Pimm et al., 2014) at unprecedented accuracy, there are still many improvements needed in research and applications, to allow for effective conservation management of our planet and to protect it to the degree that all future generations of humans will be able to survive. Recently, initiatives have been proposed in order to streamline the closing of gaps on this knowledge front and move towards operational data products for conservation management as quickly as possible. The Convention of Biological Diversity (CBD) has proposed the Essential Biodiversity Variables (EBV's); a list of relevant biodiversity indicators to achieve the Aichi Biodiversity Targets, a set of goals for the conservation of global

biodiversity (Pereira et al., 2013). To reach these goals, the CBD has proposed a list of biodiversity metrics to be derived from spaceborne remote sensing data; a challenge to be addressed over the next decade (Skidmore et al., 2015). In this dissertation I address this challenge through exploring a method for comprehensive mapping of tree species diversity across the tropics to enable better conservation management. These maps will strengthen our understanding of the value of tree species diversity in intact forests, and elucidate interactions between climate and ecosystems across spatial scales.

1.2 BACKGROUND

1.2.1 TROPICAL FORESTS

Tropical forests are spread over five main biogeographical regions. The Neotropics in Central and South America, the African mainland, the island of Madagascar, the Sundaland tropics in South-East Asia and the region of Papua-New Guinea and Australia (Corlett and Primack, 2011). The climate across the tropical zones varies greatly and the Inter Tropical Convergence Zone controls the amount and timing of rainfall. In Africa, the rainless months are often misty and overcast, providing quite different conditions than in tropical America and continental Asia where dry periods are sunny (Corlett and Primack, 2011). Historically, biogeographical processes are known to be the most important driver of tree species diversity patterns at a large continental scale (Keil and Chase, 2019).

The global estimate of the number of tropical tree species lays between 40,000 and 53,000 species (Slik et al., 2015). The highest tree species diversity is found, in

decreasing order, in the Neotropics, then on Sundaland and last in African tropical forest (Corlett and Primack, 2011; Mutke and Barthlott, 2005). Tree species composition in the Neotropics evolved to be the most distinct of the three, as Central and South America were separated from Africa and South-East Asia about 70 million years ago, when they broke off from the Gondwana continent. The floristics in Africa and South-East Asia are more similar to each other, although Africa only has 10-20% of the species in families shared between the two regions. This disparity in species number was likely caused by drought in Africa during the last ice age which is also thought to have caused the characteristic African savanna-forest ecosystem (Corlett and Primack, 2011). At large spatial scales, historical biogeographical processes are most important for the distribution of tree species richness, but environmental variables have a stronger influence on spatial patterns of species richness at regional or local scales (Keil and Chase, 2019). For example, topographic heterogeneity, including the uplift of mountains in the Himalaya and the Andes, is an important factor that has affected speciation over large scales during past periods of climate change (Kreft and Jetz, 2007). On the other hand, at a local scale, soil nutrients explained 36-51% of tree species richness (within 25 – 50 ha plots) in study sites in Colombia, Ecuador and Panama (John et al., 2007).

1.2.2 FUNCTION AND DIVERSITY

Forests fulfill many key functions and are indispensable for a healthy earth in which humans and other organisms can thrive. For example, tropical forests play a key role in the regulation of local, regional and global climate, the regulation of watersheds, habitat provision for other organisms and indigenous cultures and they are an

important source of medicinal components (Watson et al., 2018). More intact and diverse forests are better able to provide these ecosystem services (Tuck et al., 2016). For example, intact forests are more efficient at storing CO₂ from the atmosphere (Longo et al., 2016). Thus, understanding the patterns and drivers of tropical tree species diversity is essential.

Ecosystems and their inherent diversity can be generally characterized along three principal dimensions: composition, structure and function. These three dimensions are key to biodiversity, which can be measured at the genetic, species and ecosystem level (Noss, 1990). Structural diversity of forests is commonly described by a combination of structural attributes related to variation in vertical foliage arrangement, canopy height, canopy density and deadwood (McElhinny et al., 2005). Following niche theory, a diverse combination of structural attributes is expected to promote the abundance and diversity of species by creating a large variety of ecological niches and habitat elements (MacArthur and MacArthur, 1961).

1.2.3 DIVERSITY MAPPING

Inventory and mapping of biodiversity has been a challenge for many decades. Identifying and monitoring species existence and extinction in the field is difficult. In 2011, it was estimated that 86% of species on Earth are still unknown (Mora et al., 2011). Despite this challenge, there are a lot of things we do know about the world's biodiversity on a large scale. For example, biodiversity is higher near the equator than near the poles (Whittaker et al., 2001) and 44% of all vascular plant species are confined in endemic hotspots covering just 1.4% of the land surface (Myers et al.,

2000). The hotspots of vascular plant species richness are in Central America, in the eastern Amazon, along the south-east coast of Brazil, along the east coast of Madagascar, in the northern part of the island of Borneo and along the northern slopes of the Himalaya's, stretching east into South-East Asia (Mutke and Barthlott, 2005). Three maps of vascular plant diversity derived by (Kreft and Jetz, 2007), using different methods, show alternative distributions with lower diversity along the east coast of Brazil and in south-east Asia's mainland. All of these analyses were based on field data, in which the spatial distribution of the available field data is a limiting factor in extrapolating the species richness between field sites and across data-scarce areas. Combining the field-inventory species richness data with environmental information has enhanced the understanding of the drivers of such global patterns. Tropical tree species diversity is known to increase with increasing precipitation, forest stature (stage of growth), soil fertility, time since catastrophic disturbance and rate of canopy turnover, and to decrease with seasonality, latitude and altitude (Givnish, 1999).

The list of variables that can be correlated with tree species richness at different scales is long and changes by observation scale (Keil and Chase, 2019). Moreover, the strength of the relationship between a variable and tree species richness may also change with resolution (plot size) as tree species richness is not linearly related with the area measured (species-area curve) (Hubbell, 2001). This does not only complicate the development of reliable predictive models at a certain resolution, but also complicates the extrapolation of estimates at one resolution to larger areas, which

complicates successful mapping of pantropical tree species richness at high spatial resolution.

1.2.4 LIDAR FOR MAPPING DIVERSITY

Forest structure is, just as species richness, influenced by a complex interaction of historic, environmental, and human related variables; precipitation in the wettest month being the most important single predictor of plant height (Moles et al., 2009).

Forest structure traditionally measured in the field mainly comprised of four variables: canopy height, biomass, basal area/ha and tree density (Palace et al., 2015).

However, active remote sensing measuring techniques have revolutionized the way we look at canopy structure. With lidar remote sensing, for example, it is now possible to not only obtain information on canopy height, but also on the position and amount of plant material along the vertical axis of the canopy (Tang et al., 2012).

Palace et al. (2015) stressed that high resolution lidar data possesses vertical structure information that is inherently linked to ecological processes and forest dynamics that are exhibited in the structural properties of the forest.

This wealth of information on the vertical structure has been hypothesized to be a proxy for vertical niche occupancy and relate, to some extent, to tree species richness; thus, this information may be used for mapping tropical tree species richness.

Different tree species require different niches to grow and thus quantifying the occupation of the vertical forest niche space may relate to the number of different tree species present in an area.

Lidar data lends itself perfectly for studying the structural diversity as lidar instruments directly collect 3D measurements of both the horizontal and vertical vegetation structure (Lefsky et al., 2002). While a number of aspects of the vegetation structure have been derived and mapped from lidar data, they can generally be summarized to represent canopy height, canopy cover and vertical variation in foliage arrangement. Additionally, lidar measurements of forest structure can be used to estimate other biometric measurements such as basal area and tree density (Palace et al., 2015). The combination and diversity of forest structure attributes can be described at the individual tree, stand and ecosystem level, depending on the type of lidar instrument used. The structure of individual trees and small study sites is best measured with terrestrial lidar (Marselis et al., 2016), but high-density airborne lidar can also be used for studying the forest structure at the tree level (Coomes et al., 2017). At the landscape level, airborne lidar is now a standard tool for studying forest stand structure (Asner et al., 2009; Kent et al., 2015). Although not available spatially continuously, spaceborne lidar data from the Icesat-1 instrument provides the means to describe 3D forest structure and its diversity across large regions, such as the Amazon (Tang and Dubayah, 2017).

The use of lidar data for measuring forest structure and relating this to trees species diversity is currently of specific interest with the development of the Global Ecosystem Dynamics Investigation (GEDI). GEDI is a lidar instrument that was launched to the international space station in December 2018. It is the first spaceborne lidar instrument of its kind, specifically optimized to measure ecosystem

structure between 52 degrees latitude north and south from its vantage point on the International Space Station (ISS). The GEDI instrument has three lasers of which one is split into two beams resulting in four beams total. These laser beams are dithered back and forth creating a scanning pattern of 8 ground tracks. The between-track separation is 600 m at the equator and the along-track distance between lidar shots is 60 m (Dubayah et al., under review). The lidar footprints have a nominal footprint of ~25 m diameter, equivalent to the Land Vegetation and Ice Sensor (LVIS), which is the airborne predecessor of the GEDI instrument (Blair et al., 1999). The GEDI instrument will provide freely available lidar waveforms collected across the Earth and thus may be a source of particularly relevant information that can potentially be used for mapping tree species diversity across the tropics in case a significant relationship exists between the canopy structure metrics provided by GEDI and the tree species composition and diversity.

1.3 RESEARCH OBJECTIVES

The overall goal of my research is to explore the potential of full-waveform lidar data to characterize tree species diversity in the tropics. Three research objectives will help to reach this goal: (1) Assess whether full-waveform lidar data can be used to distinguish vegetation types with complex vertical structure along a successional gradient in Lopé National park, Gabon. (2) Explore the structure-diversity relationship in a tropical forest and savanna landscape across four study sites in Gabon. (3) Evaluate the existence of a pan-tropical relationship between vertical canopy structure and tree species richness. Finally, the gained insight are put into the

context of GEDI and its capability to aid mapping of tree species diversity across the tropics.

1.4 DISSERTATION OUTLINE

Chapter 2 first compares vertical canopy profiles derived from LVIS full-waveform airborne lidar data (collected with LVIS) with profiles derived from terrestrial lidar data. Then it evaluates the canopy structure of five distinct vegetation types in a forest-savanna mosaic and develops a model to classify the forest-savanna landscape in Lopé National Park into those five types.

Chapter 3 first establishes the relation between canopy height and tree species diversity across four study sites in Gabon. Then it develops a predictive relationship between canopy structure metrics additional to canopy height and evaluates the increase in model performance. Lastly, canopy structure models are used to predict tree species diversity and richness within and between the four study sites in Gabon using gridded LVIS products and ICESat lidar waveforms.

Chapter 4 first evaluates the differences in canopy structure across the tropics. Then, it summarizes the range in tree species richness and how this range changes by plot size. Finally, it assesses the relationship between vertical canopy structure and tree species richness at the local, regional and pan-tropical scale.

Chapter 5 synthesizes the outcomes of the three studies and details how the findings of this study may affect the future of tree species diversity mapping using spaceborne lidar data from GEDI.

2. DISTINGUISHING VEGETATION TYPES WITH AIRBORNE WAVEFORM LIDAR DATA IN A TROPICAL FOREST-SAVANNA MOSAIC: A CASE STUDY IN LOPÉ NATIONAL PARK, GABON

ABSTRACT

Tropical forest vegetation structure is highly variable, both vertically and horizontally, and provides habitat to a large diversity of species. The forest-savanna mosaic in the northern part of Lopé National Park, Gabon, has a large and complex variation in vegetation structure along a successional gradient. The goal of this research is to assess whether large footprint full-waveform lidar data can be used to distinguish successional vegetation types based on their vertical structure in this area. Eleven vegetation metrics were derived from the lidar waveforms: canopy height, canopy fractional cover, total Plant Area Index (PAI) and vertical profile of PAI. The PAI profiles from airborne waveform lidar showed good agreement with those from Terrestrial Laser Scanning, sampled at eight field plots across different vegetation types ($r^2 = 0.95$, RMSE = 0.63, bias = 0.41). The agreement further strengthened our confidence that lidar waveforms can be used to distinguish between the five

vegetation types, within the limits of the sampled structure, because TLS was known to provide distinct PAI profiles for these vegetation types. We then employed a Random Forest model, trained with 193 locations of known vegetation type, to classify the entire study area into five successional vegetation types (classification accuracy = 81.3%). The resulting predictive map revealed the overall spatial pattern of vegetation types across the study area. Our results suggest that lidar-derived vegetation profiles can provide valuable information on vegetation type and successional stage. This, in turn, can further help to improve habitat and biodiversity conservation and forest management activities.

2.1 INTRODUCTION

Increasing environmental and human pressures have led to broad-scale forest degradation, deforestation and fragmentation. This has had unprecedented impacts on the biomass stock, habitat characteristics and patterns in biodiversity (Bonan, 2008; Hansen et al., 2013; Naeem et al., 1999). However, the exact magnitude of forest change and its impacts on species composition, forest regrowth and successional pathways is unclear (Rockstrom et al., 2009; Schimel et al., 2013; Turner et al., 2003). Large area mapping of the three dimensional vegetation structure in tropical forests, especially in biodiversity and biomass hotspots, is of utmost importance to enable better understanding of these impacts (Jantz et al., 2015; Palace et al., 2015; Tattoni et al., 2012; Tews et al., 2004; Vierling et al., 2011; Wright and Muller-Landau, 2006). Applications of conventional passive optical imagery from air- and space- based platforms have been limited, as these data are not directly sensitive to

vertical vegetation structure. However, active remote sensing, especially from lidar instruments, has clearly demonstrated the ability to accurately measure both spatial and vertical vegetation structure, even over dense tropical rainforest (Bergen et al., 2009; Goetz et al., 2007; Huang et al., 2014; Swatantran et al., 2012; Turner et al., 2003; Whitehurst et al., 2013). Small footprint lidar data have been used to distinguish between successional vegetation types in temperate mixed-forests in the northwestern United States (Falkowski et al., 2009; Zimble et al., 2003). Temperate Eucalypt and tropical forest stands were distinguished and predicted in Australia in a similar study, using Plant Area Volume Density (PAVD) profiles derived from small-footprint lidar data (Fedrigo et al., 2018). Full-waveform lidar data also provide valuable information to quantify vertical vegetation structure, as shown in a number of studies in the last two decades with NASA's Land Vegetation and Ice Sensor (LVIS) (Blair et al., 1999; Tang et al., 2012; Tang and Dubayah, 2017). Thus, LVIS could theoretically also be used to distinguish successional vegetation types with distinct vertical structure. This is of great importance since the next spaceborne lidar mission, the Global Ecosystem Dynamics Investigation (GEDI), will deploy a full-waveform lidar instrument, designed to measure global vegetation structure, on the International Space Station (ISS) in late 2018. GEDI will collect billions of lidar waveforms similar to LVIS during its nominal two-year mission, sampling between 51.5° latitude North and South (Dubayah et al., under review; Stavros et al., 2017; Stysley et al., 2016). Hence, GEDI has a high potential to fill current observation gaps of tropical forest structure and provide unprecedented opportunities for ecological and

biodiversity studies (Jetz et al., 2016; Rose et al., 2015; Skidmore et al., 2015; Wolf et al., 2012).

Here, we test the use of vegetation structural parameters derived from full-waveform LVIS lidar data to map landscape level patterns of interior and exterior forest structure in a tropical forest-savanna mosaic in the north-west corner of Lopé National Park in Gabon. Cuni-Sanchez et al. (2016) demonstrated the possibility to distinguish the vegetation profiles of five main successional vegetation types in this forest-savanna mosaic using Terrestrial Laser Scanning (TLS). TLS indirectly estimates vertical Plant Area Index (PAI) profiles using multi-angular lidar observations, but only provides these at the plot level, thus producing data over limited spatial scales (Newnham et al., 2015). However, large footprint airborne waveform lidar, such as LVIS, enables large area wall-to-wall mapping of the vertical vegetation profile, but only from the near-nadir view and thus may be less sensitive to differences in canopy element angle distribution and foliage clumping (Hopkinson et al., 2004; Jupp et al., 2009). Currently, there is a paucity of studies validating the estimation of vertical PAI profiles from large footprint full-waveform airborne lidar with TLS derived profiles, especially in African tropical rainforests. Hence, validating the PAI profiles from LVIS data with those from TLS, in this study environment, is of high importance.

The goal of this research is to assess whether full-waveform lidar data can be used to distinguish vegetation types with complex vertical structure along a successional gradient. The remainder of this paper is structured as follows: first, the vertical

vegetation profiles from large footprint airborne lidar waveforms are validated against those from TLS. Then, field data are used to compare waveform-derived vegetation profiles between the main vegetation types. Subsequently, vegetation structure parameters, derived from the lidar waveforms, are used to spatially map the five main vegetation types. Then, an example application of this vegetation map for local conservation management of forest islands is demonstrated. Lastly, a discussion of the findings in light of the research goal and the implications for space-based observations of tropical forest structure from GEDI is provided.

2.2 METHODS

2.2.1 STUDY AREA & AVAILABLE DATASETS

Much of what we know about Lopé National Park and the vegetation of its forest-savanna mosaic comes from (White and Abernethy, 1997). The area is known for its exceptionally high species diversity in a successional gradient from savanna to complex tropical forest (White, 2001; White and Abernethy, 1997). Savanna fragments occur in this region as a remnant of long-term climate change, initially formed during an arid time period between 18000 and 12000 years ago. The annual rainfall is much lower (~1500 mm) than in other parts of this tropical region (~3000 mm) because of the position relative to the Massif du Chaillu. This climatic niche, in combination with regular fires, led to the forest-savanna mosaic Lopé is known for today (White and Abernethy, 1997). The vegetation has been categorized in five broad probable successional vegetation types: (1) savanna, characterized by long grass, shrubs, solitary trees and frequent fires. (2) Colonizing forest, which results

from the dispersal, settlement and growth of pioneer species in unburned savanna. (3) Monodominant Okoumé forest, where Okoumé (*Aucoumea klaineana* Pierre) trees, one of the initial colonizers along with Azobé (*Lophira alata* Banks ex C.F.Gaertn.) and Ozouga (*Sacoglottis gabonensis* (Baill.) Urb.), form tall closed canopy even-aged patches with an open understory (White, 2001; White and Abernethy, 1997). (4) Marantaceae forest (also known as Young Marantaceae forest), where canopy can be rather open and the understory filled with a very dense layer of herbaceous plants from the Marantaceae and Zingiberaceae families. (5) Mixed forest (also known as Mixed Marantaceae forest), which is a more mature stage in the succession where tree species are more diversified and the understory clearer (White, 2001). Climate change, forest fires and elephant activity are hypothesized to contribute to the retreating and advancing of the forest boundary over time, resulting in forest islands within the savanna (Ukizintambara et al., 2007). Two distinct types of forest islands exist: bosquets and gallery forests. Bosquets are forest fragments fully isolated within the savanna, showing evidence of past human presence in their vicinity, while gallery forests are riverine forest fragments that occur along water streams and are more connected to the continuous forest (White and Abernethy, 1997). The two types of forest have distinct species composition and conservation value because of their different origin (Ukizintambara et al., 2007).

Three types of data were available to study the vegetation in Lopé: field data, TLS data, and LVIS full-waveform lidar data. In this study, field information on “vegetation type” was obtained at 23 plot locations surveyed during two different

campaigns (ten 20x40m plots in 2013, and three 0.5ha and ten 1ha plots in 2016). The 2013 field data were collected coincidentally with the TLS campaign (Cuni-Sanchez et al., 2016). The 2016 vegetation data were extracted from a vegetation inventory dataset collected during the AfriSAR campaign and available through ForestPlots.net (Labrière et al., 2018; Lopez-Gonzalez et al., 2011, 2009). AfriSAR was a joint airborne and field campaign between the European Space Agency (ESA) and the National Aeronautics and Space Administration (NASA), with collaboration from the Gabonese Studies and Space Observations Agency (AGEOS), the German Aerospace Center (DLR) and the French Aeronautics Space and Defense Research Lab (ONERA) (Fatoyinbo et al., 2017). The field plots were set up to cover a gradient in biomass and succession. Three constraints were used to determine the plot locations: within-plot vegetation type homogeneity, topographic homogeneity, and within spatial coverage of previously acquired P-band radar data. The vegetation type was determined by a local botanical expert through a visual assessment, prior to plot set up, and in line with the literature (e.g. Okoumé and Sacoglottis dominating "Okoumé forest") (White and Abernethy, 1997). TLS data were collected in 2013 with the RIEGL VZ-400 for eight plots at 6 scanning locations in 20x40 m plots (Cuni-Sanchez et al., 2016). The TLS instrument had a nominal beam divergence of 0.35 mrad, operated in the infrared (wavelength 1550 nm) and had a pulse repetition rate of 300 kHz. Scans were taken in the four corners and in the middle of the two long (40 m) sides. At each location, scans were performed at an angular resolution of 0.06° in azimuth and zenith directions to ensure full hemispherical coverage (Cuni-Sanchez et al., 2016). LVIS collected full-waveform lidar data with 1042 nm wavelength and

nominal footprint size of approximately 20 m in March 2016 during the AfriSAR campaign. Ten flight strips covered a wall-to-wall area of 10x75 km over Lopé NP (LVIS swath of 1km; Figure 1). We studied a 10x30 km area within the wall-to-wall LVIS sampling area that comprised the forest-savanna mosaic (black box in Figure 1).

PAI (area of plant material, in m^2/m^2) profiles were extracted from the TLS point clouds to represent the vertical vegetation profile. The PAI profile was calculated in bins of 0.5m according to the methods described by (Calders et al., 2014; Cuni-Sanchez et al., 2016; Jupp et al., 2009). The plot-level TLS-derived PAI profile was calculated as the average from all TLS scans along the plot edges. This average PAI profile was then aggregated to different vertical bins (1, 3, and 5m respectively) to allow for comparison with LVIS data Figure 2. LVIS data were processed to retrieve ground elevation, vegetation height, canopy fractional cover, total effective PAI and the PAI profile at 1 m vertical resolution as part of the mission requirements from AfriSAR (Tang et al., 2014, 2012). PAI profiles from all LVIS waveforms within the plot boundaries were averaged to represent the plot level PAI profile. These PAI profiles were also vertically aggregated to 1, 3, 5 and 10 m vertical bins. Variation of the PAI was calculated at each height bin as the standard deviation of PAI from all TLS scans or LVIS waveforms in each plot at the given height. The PAI profiles of the TLS and LVIS instruments were then compared using R^2 , bias, Root Mean Squared Error (RMSE) and the Concordance Correlation Coefficient (CCC) (Lawrence and Lin, 1989). Agreement between LVIS and TLS PAI profiles would

suggest that LVIS data can be used to distinguish the successional vegetation types in Lopé, given that (Cuni-Sanchez et al., 2016) successfully used the TLS data to distinguish these same vegetation types.

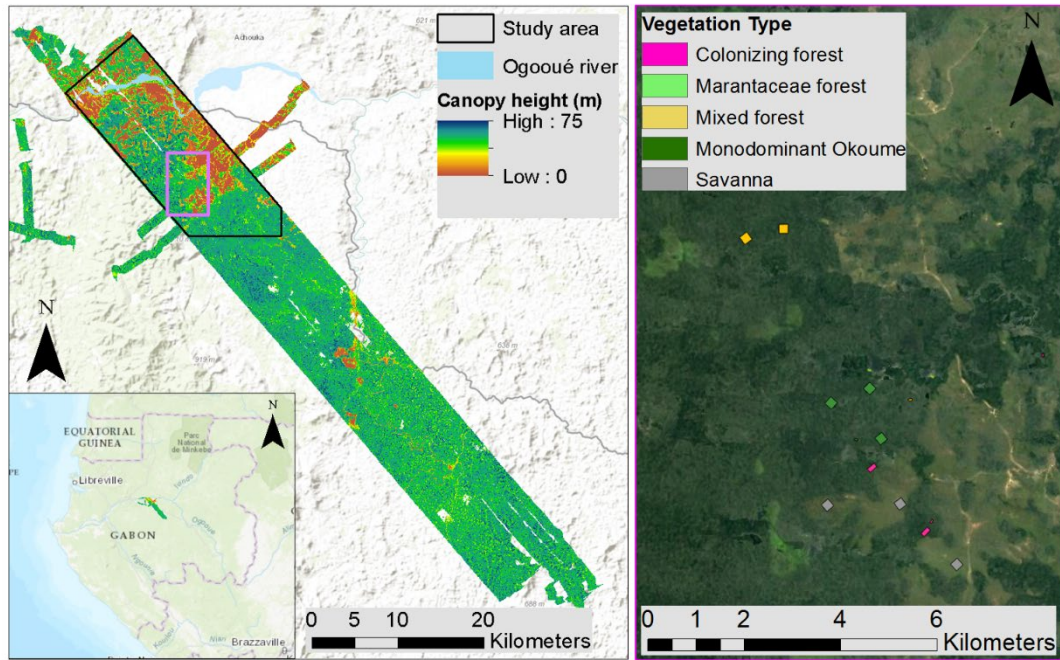


Figure 1: Inset on left: location of data collection in Gabon. Left: canopy height collected by the LVIS instrument and location of study area. Magenta outline shows location of right panel. Right: Location of subset of field plots in the Lopé forest-savanna mosaic.

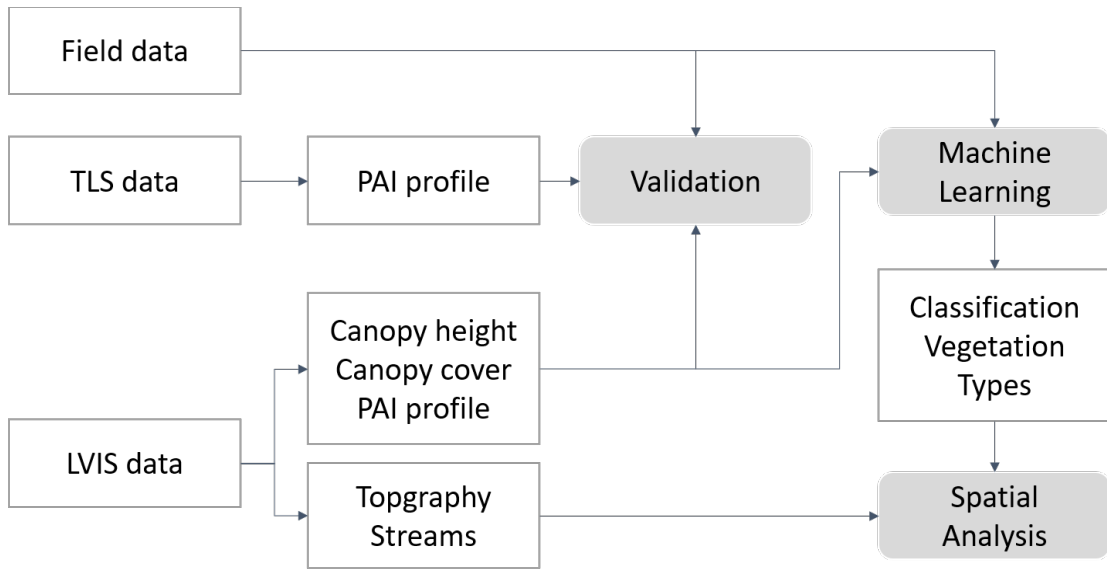


Figure 2: Workflow of study, grey boxes show main processes to reach study objectives.

2.2.2 LIDAR DATA PROCESSING

PAI (area of plant material, in m^2/m^2) profiles were extracted from the TLS point clouds to represent the vertical vegetation profile. The PAI profile was calculated in bins of 0.5m according to the methods described by (Calders et al., 2014; Cuni-Sanchez et al., 2016; Jupp et al., 2009). The plot-level TLS-derived PAI profile was calculated as the average from all TLS scans along the plot edges. This average PAI profile was then aggregated to different vertical bins (1, 3, and 5m respectively) to allow for comparison with LVIS data Figure 2. LVIS data were processed to retrieve ground elevation, vegetation height, canopy fractional cover, total effective PAI and the PAI profile at 1 m vertical resolution as part of the mission requirements from AfriSAR (Tang et al., 2014, 2012). PAI profiles from all LVIS waveforms within the plot boundaries were averaged to represent the plot level PAI profile. These PAI profiles were also vertically aggregated to 1, 3, 5 and 10 m vertical bins. Variation of the PAI was calculated at each height bin as the standard deviation of PAI from all TLS scans or LVIS waveforms in each plot at the given height. The PAI profiles of the TLS and LVIS instruments were then compared using R^2 , bias, Root Mean Squared Error (RMSE) and the Concordance Correlation Coefficient (CCC) (Lawrence and Lin, 1989). Agreement between LVIS and TLS PAI profiles would suggest that LVIS data can be used to distinguish the successional vegetation types in Lopé, given that (Cuni-Sanchez et al., 2016) successfully used the TLS data to distinguish these same vegetation types.

2.2.3 VEGETATION TYPE CLASSIFICATION

We gridded canopy height, canopy cover fraction, total PAI and eight 5 m bins to describe the PAI profile (Tang et al., 2016) at a 25x25 m spatial resolution. This was to facilitate analysis of the structural variability within and between classes and allow for vegetation classification. Information on vegetation type was available as reference data for 193 pixels, covered by the 23 field plots. For each pixel the single-pixel value of each metric (11 total) was extracted. A mean, standard deviation, and coefficient of variation were also calculated for each pixel, using a 3x3 window around the pixel to account for high local variability within vegetation types (adding 33 metrics). Two training datasets were created based on the 193 pixels, with 11 (pixel-based) and 44 (neighbor-based) vegetation metrics respectively. These training datasets were used to build two Random Forest classification models, with the randomForest package in R (Breiman, 2001). The resulting models were used to predict the vegetation class of all pixels in the 10x30 km study area, using the majority vote for each pixel, resulting in two spatial maps (pixel- and neighbor-based) of vegetation classes for the study area. Accuracy of the random forest classification was expressed with the out-of-bag estimate of the error rate, the confusion matrix of the out-of-bag classifications, the kappa value and omission and commission errors.

The classified vegetation map was further refined ad-hoc by focusing on characteristics of forest islands and the location and extent of Colonizing forest to provide further insight in potential applications of the created product. Forest islands within the savanna are known to be either one of two types, bosquet or gallery forest. The vegetation structure of the bosquets and gallery forests may appear mostly

similar to that of colonizing forests, but they can be distinguished mainly depending on the proximity to water streams (Ukizintambara et al., 2007; White and Abernethy, 1997). Therefore, spatial data on the hydrology would allow classification of forest islands within the savanna. A Digital Terrain Model (DTM) was created from the LVIS ground elevation, to assess the hydrological network. Flow direction and flow accumulation (Jenson and Domingue, 1988) were calculated in ArcGIS 10.4 and all cells with a weight greater than 30 were extracted (Tarboton et al., 1991). Flow accumulation was used to generate a stream order map with streams up to the 6th order. A forest-savanna map was derived from the vegetation classification created with the Random Forest model to extract forest islands from the savanna. All forested cells were clustered based on adjacent pixels in a four cell window to ensure pixel groups represented physically continuous (adjacent) forest. The large forest islands that formed the continuous forest were eliminated and the actual forest islands were classified as gallery forest or bosquet according to the following classification rule: forest islands overlapping with the location of a 4th or higher order stream, or directly adjacent to the Ogooué River, were classified as gallery forest. The cut-off value for water-carrying streams was determined by expert knowledge of the local river flow system. All remaining forest islands were identified as bosquet. The savanna edge was also determined from the vegetation type map and all Colonizing forest adjacent to savanna (eight cell window) was classified as edge-Colonizing forest. All Colonizing forest cells away from the edge were classified as inner-Colonizing forest.

2.3 RESULTS

2.3.1 PAI PROFILE VALIDATION

Cumulative PAI profiles at 1, 3 and 5m vertical intervals from LVIS and TLS showed good overall correspondence (5m: $r^2 = 0.95$, RMSE = 0.63, bias = 0.41, CCC=0.94) (Figure 3, Chapter I.1 TLS-LVIS comparison multiple scales). A slight bias towards relatively higher LVIS PAI values occurred in some higher canopy strata. Standard deviation of PAI increased higher in the canopy. Total PAI from TLS measurements is generally lower than from LVIS.

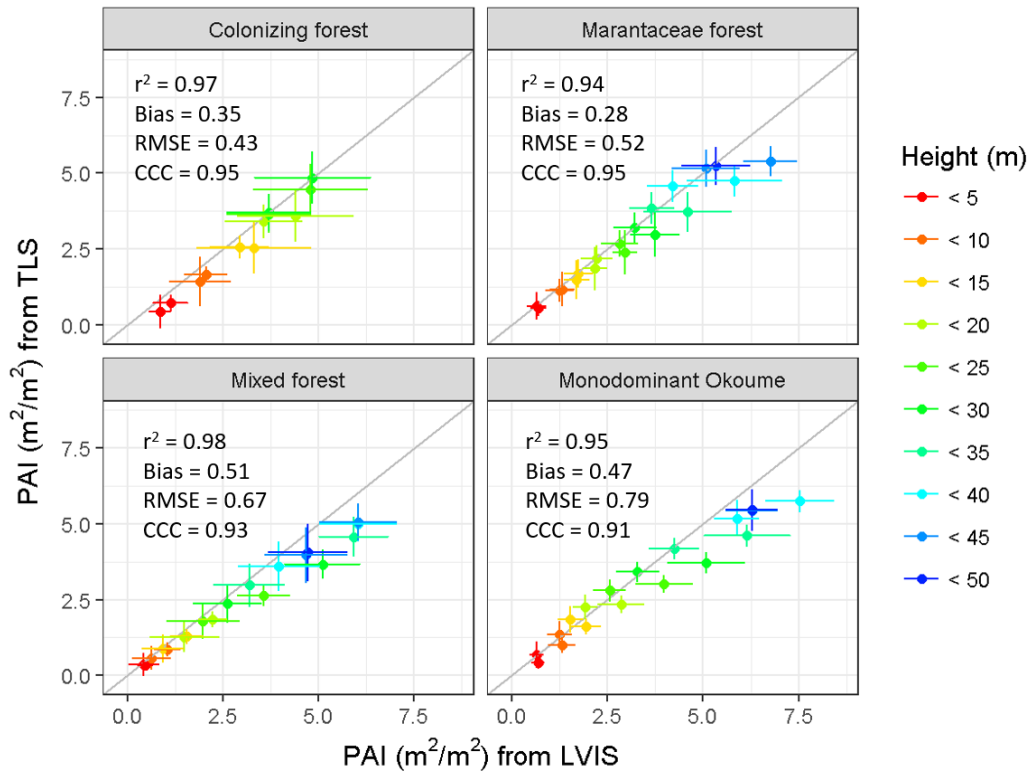


Figure 3: Cumulative Plant Area Index (PAI) profiles from Terrestrial Laser Scanning (y-axis) and airborne lidar data (x-axis) in vertical intervals of 5m for eight plots covering four vegetation types. Each point represents the average cumulative PAI up to indicated height (m), calculated as the mean of all TLS scans around the plot edge (TLS) or all waveforms within the plot boundary (LVIS). Whiskers indicate PAI standard deviation of the group of scans (TLS) or waveforms (LVIS) at indicated height. 1:1 line shown in grey.

2.3.2 VEGETATION STRUCTURE VALIDATION

The five vegetation types were roughly distinguishable by their vegetation structure summarized by the PAI profile (10 m vertical interval), total PAI, canopy height and/or canopy cover Figure 4. More details on the structure of each vegetation type were obtained by studying the PAI profile at higher vertical resolution (5 m) and by observing the mean and standard deviation of the vegetation metrics within a 3x3 window of the observed location (Chapter I.2 Vegetation structure metrics: Figure 25 - Figure 28). Savanna was distinguished by low total PAI, canopy height and canopy cover. Monodominant Okoumé forest showed the highest canopy cover fraction (ranging between 0.93 – 0.99). Marantaceae forest, Mixed forest and Colonizing forest occurred with more similar characteristics. However, Colonizing forest generally had lower canopy height (mean height = 29m, compared to inner-forest mean height = 40 m). In Marantaceae forest, the PAI in lower strata (< 10 m) was generally lower than in Mixed forest (median of 0.81 compared to 1.73). Mixed forest occurred with the most diverse structure, highly variable at small spatial scale.

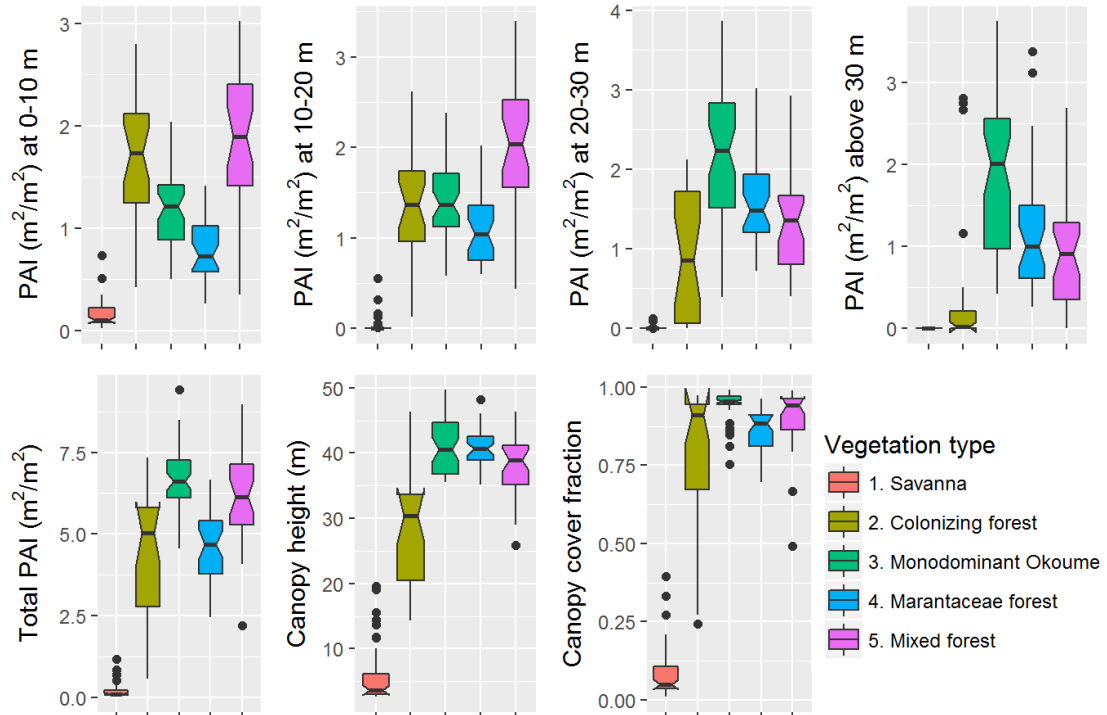


Figure 4: Seven metrics (y-axis) describing characteristics of the vegetation structure of five vegetation types (colored), ordered by successional stage. Each boxplot is composed of the pixel values within all field plots covered by one specific vegetation type.

High spatial variation in vegetation structure at the forest edge and in the inner forest was revealed by a Red-Green-Blue image combining information on three PAI profile layers at 10 m vertical interval (Figure 5). In the south-east part of the study area the forest edge showed a more gradual transition towards savanna, with high PAI in the lower strata (< 10 m). In the north-east area, the forest-savanna edge was more abrupt. Inner-forest vegetation structure showed high pixel-to-pixel variation, but more homogenous areas were distinguished at the landscape scale.

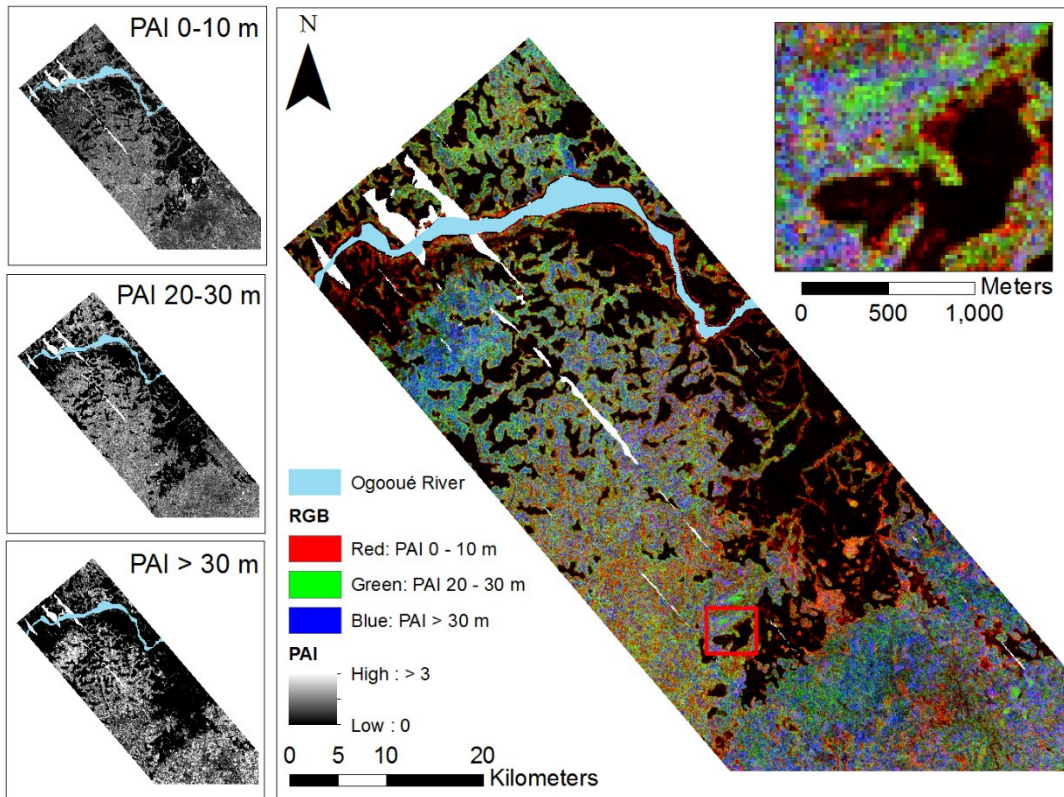


Figure 5: PAI values of different vegetation height layers (left) combined into a Red-Green-Blue image (right), showing the variation in vertical vegetation structure across the study area. Zoom-in on right (location indicated with red box) reveals high structural variation over small scale.

2.3.3 VEGETATION CLASSIFICATION AND SPATIAL ANALYSIS

Both pixel- and neighbor-based Random Forest classifications highlighted the spatial distribution of the vegetation types across the study area (Figure 6; Chapter I.3 Alternative classifications). The out-of-bag error estimate resulting from the pixel-based Random Forest model was 18.7%, with a κ -value of 0.76 (Table 1). The inner-forest in the center of the study area was dominated by Mixed forest, while the North and the South were dominated by Marantaceae forest. Colonizing forest occurred mainly along the savanna edge, but was also found in small formations in the inner

forest, often surrounded by Mixed forest. Monodominant Okoumé and Savanna, the two vegetation types with the most distinct vegetation structure (Figure 4), showed the lowest omission errors (13.7% and 4.0%, respectively). Omission errors were highest for Colonizing forest and Mixed forest (36.0% and 34.4%, respectively). Commission errors were roughly similar for all forest types, but lowest for savanna (Table 1). The extent of the Colonizing forest was largely over-estimated using neighbor-based vegetation metrics in the classification model (Chapter I.3 Alternative classifications). However, the apparent variation in inner-forest types from pixel-based classification (observable as the “pixelated” nature of the classification in Figure 6) was reduced by incorporating neighbor information. Overall, the neighbor-based approach provided a smoother, inner-forest classification with distinct vegetation patches of each type (Chapter I.3 Alternative classifications).

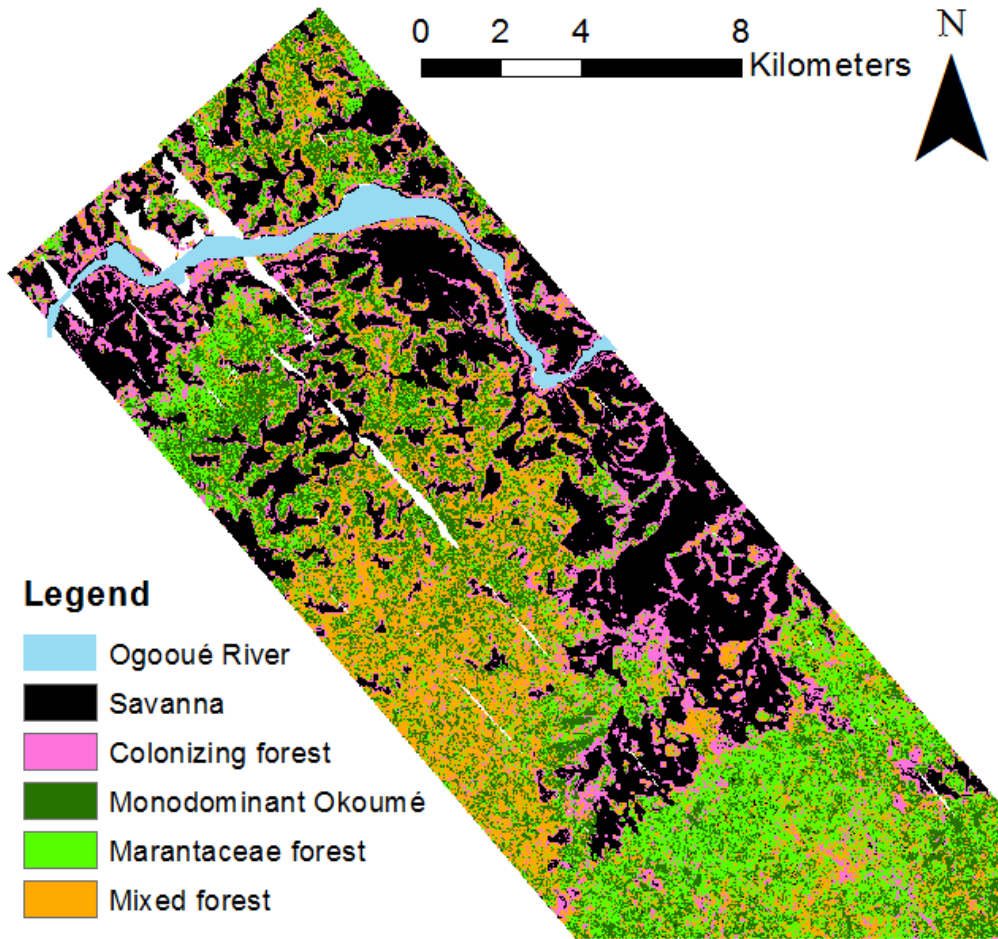


Figure 6: Classification of the five successional vegetation types in the study area, using pixel-based Random Forest classification.

Table 1: Confusion matrix of pixel-based random forest classification using a total of 193 training pixels. Out-of-Bag estimate of the error rate is 18.7%.

	Savanna	Colonizing forest	Monodominant Okoumé	Marantaceae forest	Mixed forest	Total	Omission error
Savanna	48	2	0	0	0	50	0.040
Colonizing forest	1	16	4	1	3	25	0.360
Monodominant Okoumé	0	1	44	3	3	51	0.137
Marantaceae forest	0	0	7	28	0	35	0.200
Mixed forest	0	1	4	6	21	32	0.344
Total	49	20	59	38	27	193	
Commission error	0.020	0.200	0.254	0.263	0.222		
$\kappa = 0.76$							

The forest edge in the south-east part of the study area was generally accompanied by a band of Colonizing forest, while the forest edge in the north-west part of the study area was often characterized by a sharp boundary between forest and savanna (Figure 7). Gallery forests were more abundant at the lower elevation than Bosquets.

Bosquets were generally smaller in size than gallery forests (Figure 7).

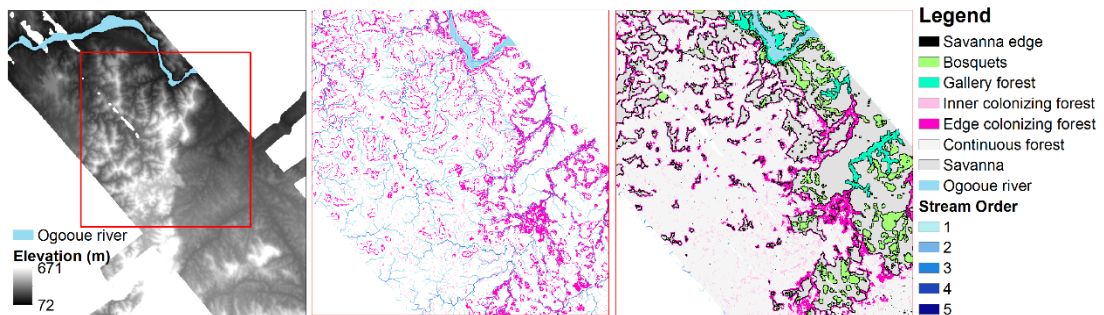


Figure 7: Digital Terrain Model (DTM) from LVIS data (left) showing zoom-in area of other figures (red box). Stream network (6th order stream is shown as the Ogooué river) overlain with inner- and edge-Colonizing forest (middle). DTM, stream order and forest island classification led to the discrimination between bosquets and gallery forest within savannas (right).

2.4 DISCUSSION

2.4.1 VEGETATION STRUCTURAL VARIABILITY

The cumulative PAI profiles derived from LVIS agree well with the TLS-derived cumulative PAI profiles at the plot level (Figure 3). However, there is a bias in the high strata (approx. > 35m), with the largest bias and lowest CCC value in monodominant Okoumé forest. This can be attributed to a combination of the attenuation and the occlusion effect, which the terrestrial laser instrument experiences from its ground-based perspective in tall, dense, forest (Calders et al., 2014; Hilker et al., 2010; Zhao et al., 2013). The larger observed variance in LVIS measurements compared to TLS measurements may be attributed to the different fields-of-view and angular sampling of the two instruments. LVIS measures the vegetation structure from (near-)nadir view and typically observes canopy gaps, while TLS measures the structure with a full hemispherical view and over a larger effective footprint size. Despite this, LVIS and TLS vertical PAI profiles show high correspondence at a plot scale of 20x40 m in these tropical forest ecosystems. This suggests that, within the limits of the site-specific range of vegetation structure sampled, we would be able to use the structural information to predict the five vegetation types spatially as opposed to using a data-mining approach to classify the area.

The tropical forest vegetation structure as derived from LVIS data in Lopé National Park is highly variable and complex. Savanna, on the other hand, portrays low variability, both horizontally and vertically, at the plot scale because of the sparser vegetation, consisting mainly of grass, shrubs and the occasional isolated tree (Figure

4; White and Abernethy, 1997). Savanna is therefore distinguished easily from all forest types and classified with the highest accuracy. Colonizing forest shows high variability in vertical structure, as the structure highly depends on the time since colonization (White and Abernethy, 1997). Its structure is distinct from that of the other vegetation types as a result of its characterizing lower canopy height and cover and relatively high PAI in the lower strata (Figure 4). Monodominant Okoumé forest was best distinguished by high canopy height and high cover, with lower spatial variability (Figure 4; Chapter I.2 Vegetation structure metrics), because of the roughly even-aged nature of the stands (White, 2001). Marantaceae forest and Mixed forest have rather similar structural characteristics, which is in correspondence with the literature and the known vegetation structure in these forest types (White, 2001). In our analysis, Marantaceae forest generally has a lower PAI in the lower strata (< 10 m) compared to Mixed forest, and was distinguished by this characteristic. However, this characteristic could seem slightly counter-intuitive at first sight as Marantaceae forests are known for their very dense herbaceous layer of Marantaceae up to 2-3m aboveground (White, 2001; White and Abernethy, 1997). That said, the LVIS signal from this thick herbaceous layer is most likely mixed with that from the ground, thus resulting in an erroneously lower-than-expected PAI for the lower strata (< 3 m). Moreover, because Marantaceae outcompete woody seedlings and saplings in light competition, this results in a "vertical vegetation gap" between ca. 2-3 to 10 m aboveground, which is accurately depicted by the low LVIS-derived PAI for the corresponding layer, which allows for correct classification of Marantaceae forest. Mixed forest tends to have a more complex spatial structure with high horizontal and

vertical variation. This high variability in the LVIS-derived vegetation characteristics complicates successful differentiation of this vegetation class.

2.4.2 VEGETATION CLASSIFICATION

The high spatial variability of the forest structure becomes apparent when combining multiple layers of structural information into one image (Figure 5) showing a high small-scale diversity (from pixel-to-pixel) within the continuous forest. This spatial variation is reduced by classifying each pixel into one of the five vegetation types. Our classification results (overall accuracy 81.3%) can be discussed in the context of studies with similar aims and methods. However, note that these studies utilize small-footprint lidar data and are not focused in the tropics. Falkowski et al. (2009) and Zimble et al. (2003) report a classification accuracy >90%, using small-footprint lidar data to distinguish single- and multi-layered temperate forests in the inland western United States. Zimble et al. (2003) used one decision rule to classify two types of forest while Falkowski et al. (2009) used two random forest models to classify, respectively, seven and six structurally different successional classes with class accuracies ranging between 63.64% to 100%. Fedrigo et al. (2018) used PAVD profiles and random forest models to classify a temperate forest stand in two vegetation types: Eucalypt and rainforest. They reported an overall accuracy of 84%. Our classification accuracy falls within the range of previously reported accuracies, given that a different type of lidar data were used and that our study site – with 4 tropical forest types and savanna – is structurally more complex than those from the published studies.

Errors occur in our classification of the four forest types as their vertical vegetation structure is not completely different from each other, providing a fundamental limitation for solely using vertical structure information. Colonizing forest and Mixed forest show the lowest classification accuracy (both roughly 65%). The vertical structure of Colonizing forest can be variable as it depends largely on the time since colonization and growth rate and Mixed forest structure may resemble that of Monodominant Okoumé and Marantaceae forests, possibly because succession can lead to vegetation compositions with similar structures as those types. This variation leads to lower classification accuracy of the aforementioned two types. The neighbor-based classification was not considered reliable in this forest-savanna mosaic, because Colonizing forest could not be depicted accurately with neighbor-based metrics as it generally only occurs as a narrow band along the forest edge. A hierarchical approach, using pixel-based estimates for the classification of savanna and Colonizing forest and neighbor-based values for inner-forest type classification was tested additionally (Chapter I.3 Alternative classifications). The results showed a classification with more homogenous inner-forest vegetation classes while maintaining the narrow band of Colonizing forest (Chapter I.3 Alternative classifications). However, additional field data would be needed to allow for sufficient validation of such a hierarchical approach, which is beyond the scope of this study.

Two main assumptions of the classification process need to be taken into account when interpreting the results. First, the training data are assumed to be a

representative sample of the population. However, the 193 pixels available for the classification process in this study were largely clustered because multiple pixels were available for each of the 23 field plots, thus, even though the pixels are within the assigned vegetation type, they may potentially misrepresent the population. The sample of Colonizing forest pixels (the class with highest classification error) was low compared to the other vegetation types because of the edge-driven process of savanna colonization by forest encroachment. Whereas 1-ha field plots were laid out in the other vegetation types, Colonizing forest plots were only 0.5-ha (100 m along the edge x 50 m across the edge towards the inner forest), because often the Colonizing forest does not extend further from the edge. Additional field data providing the location of specific vegetation types, distributed over a larger area, is expected to provide further insight in the variation of the vertical vegetation structure within and between vegetation types and would help to create better classification models. Secondly, during the classification process it was assumed that each pixel of the entire study area could be classified as one of these five vegetation types. However, in reality this may not be the case. Succession is a continuous process and vegetation within a successional gradient may be at a stage in between two vegetation types, complicating a successful classification. Also, the 25 m resolution of the data may have led to increased confusion at the forest edge. For example, if the forest edge falls in the middle of a pixel, the PAI values of this pixel will be lower than if the entire cell would have been forested, but higher than if the entire pixel would have been savanna. At the forest-savanna intersection it is therefore logical to see a band of one-pixel width, with a vegetation structure characteristic to Colonizing forest, even

if in reality Colonizing forest is absent and there is a sharp forest-savanna boundary. This also leads to the erroneous classification of Colonizing forest when using neighbor- rather than pixel-based information in the classification process (Chapter I.3 Alternative classifications).

We have shown that full-waveform airborne lidar data contains information that can be used to distinguish between different successional vegetation types from savanna to more mature tropical forest. In the future, it may be worth investigating forest with a continuous structural typology instead of, in this case, five vegetation classes. How to comprehensively derive information of forest structure from lidar waveforms in such a way should be studied. Combining the information describing the vegetation structure with the terrain below it (e.g. topography, soil type) can lead to further insight in landscape level forest function, species diversity or conservation value. The potential to classify forest islands into two classes with different species composition and conservation value, solely using lidar data, provides a promising application to nature conservation management (Ukizintambara et al., 2007). Previously, lidar waveforms have not been used widely to distinguish between different successional vegetation types, but with the upcoming launch of GEDI, it is increasingly important to explore applications such as this one, providing structural information for all tropical forests around the world and allowing for regional inter-comparison. This in turn may help to produce vegetation maps that are more policy relevant and thus lead to sounder ecosystem management.

2.5 CONCLUSION

The species composition of five successional vegetation types in a forest-savanna mosaic in Gabon results in distinct vegetation structures that can be expressed through Plant Area Index (PAI) profiles. Large-footprint (~25 m) full-waveform airborne lidar data provide accurate PAI profiles for each of the vegetation types, as was shown by validating the airborne lidar-derived profiles against in-situ measurements from terrestrial lidar data. Waveform lidar derivatives were then used to predict the spatial pattern of successional vegetation types, using a random forest model. This shows potential for characterizing successional vegetation types and identifying transition zones with data from the upcoming Global Ecosystem Dynamics Investigation (GEDI), a mission that will deploy a lidar instrument on the International Space Station that will collect similar waveforms across the globe.

ACKNOWLEDGEMENTS

This work is supported by NASA's Global Ecosystem Dynamics Investigation (GEDI, Grant NNL15AAO3C) and NASA's AfriSAR campaign. This work is also supported by NASA Headquarters under the NASA Earth and Space Science Fellowship Program – Grant 80NSSC17K0321. The authors express their gratitude to Mathias Disney, Andrew Burt, Aida Cuni-Sanchez (University College London), and J. Gonzalez de Tanago, Martin Herold (Wageningen University & Research), for their assistance with collecting and processing the TLS data. The authors also express their gratitude to Simon Lewis (University College London & University of Leeds) and the field research assistants of the Agence Nationale des Parcs Nationaux (ANPN) in Gabon for the collection of ground data on vegetation types in Lopé.

3. EXPLORING THE RELATION BETWEEN REMOTELY SENSED VERTICAL CANOPY STRUCTURE AND TREE SPECIES DIVERSITY IN GABON

ABSTRACT

Mapping tree species diversity is increasingly important in the face of environmental change and biodiversity conservation. We explore a potential way of mapping this diversity by relating forest structure to tree species diversity in Gabon. First, we test the relation between canopy height, as a proxy for niche volume, and tree species diversity. Then, we test the relation between vertical canopy structure, as a proxy for vertical niche occupation, and tree species diversity. We use large footprint full-waveform airborne lidar data collected across four study sites in Gabon (Lopé, Mabounié, Mondah, and Rabi) in combination with in-situ estimates of species richness (S) and Shannon diversity (H'). Linear models using canopy height explained 44 and 43% of the variation in S and H' at the 0.25 ha resolution. Linear models using canopy height and the Plant Area Volume Density (PAVD) profile explained 71% of this variation. We demonstrate applications of these models by mapping S and H' in Mondah using a simulated GEDI-TanDEM-X fusion height

product, across the four sites using wall-to-wall airborne lidar data products, and across and between the study sites using ICESat lidar waveforms. The modeling results are encouraging in the context of developing pan-tropical structure-diversity models applicable to data from current and upcoming spaceborne remote sensing missions.

3.1 INTRODUCTION

Spatial information on tree species diversity is important to enable effective conservation and biodiversity management (Turner et al., 2003), and allow for a better understanding of scale dependent relationships between forest composition and productivity (Luo et al., 2019). Information on the local presence, absence, and diversity of tree species has traditionally been collected through in-situ forest inventories (Rios-Saldaña et al., 2018). Yet, those are time-consuming and expensive, which often prevents extensive and spatially representative coverage. Using such inventories in combination with remotely sensed data is one approach to overcome some of these limitations and relationships between remotely sensed environmental data (e.g. mean annual temperature, annual precipitation, dry season length, etc.) and vegetation species have been developed to map Amazonian and Global vascular plant diversity (Mutke and Barthlott, 2005; Ter Steege et al., 2003). However, these approaches incorporate little information on the vegetation itself and the data products are provided at spatial resolutions (typically ~100 km grid cells) not optimized for conservation management.

Huston (1979) pointed out that there is a large body of literature regarding the relation between forest structure and the diversity of different taxa, while only few studies explore the relation between forest structure and tree species diversity itself. It has since been argued on several occasions that vegetation structure could be related to tree species diversity via the occupation of vertical niche space. This is the basis of a potential structure-diversity relationship, and three hypotheses have been proposed to explain it (Gatti et al., 2017; Kohyama, 1996; Marks et al., 2016; Sheil and Burslem, 2003). First, the forest architecture hypothesis asserts that tree species in the tropics have a higher variance in adult tree height, corresponding to higher species diversity, than in the temperate zone where short-statured stands with trees of uniform adulthood-height demonstrate less diversity (Kohyama, 1996, 1993). Second, the gap-size frequency hypothesis, proposed by Sheil and Burslem (2003), asserts that canopy gaps provide a range of light conditions to which different species can be more or less specialized. A changing gap-size frequency distribution would allow different light regimes and thus different species to grow (Denslow, 1980). Third, the height-diversity hypothesis asserts that a higher forest volume provides greater niche space and would thus likely result in a higher species diversity. A significant positive relation between tree species diversity and tree height, as a proxy for forest volume, was found, explaining a limited amount of the variance in tree species diversity within the USA and on a global scale (Gatti et al., 2017; Marks et al., 2016). However, Givnish (2017) objects that the use of just forest height leads to deceptive results since it only provides information on the potential trait space (volume below canopy) that could be occupied, but does not measure the occupation of this trait

space. In this study we address this caveat by exploring the relation between tree species diversity and the vertical vegetation structure; hereafter referred to as the structure-diversity relationship.

Active remote sensing systems, such as lidar and radar provide information on the vertical arrangement of the canopy (Bergen et al., 2009). Specifically, full-waveform lidar data provide detailed and more direct measurements of the vertical profile of canopy elements, which is available from the previously earth-orbiting GLAS instrument onboard of the ICESat satellite (Tang and Dubayah, 2017), NASA's airborne Land Vegetation and Ice Sensor (LVIS) (Blair et al., 1999; Tang et al., 2012) and the recently launched Global Ecosystem Dynamics Investigation (GEDI). Such vertical canopy profiles providing information on the amount of plant material (leaves and branches) along the vertical axis have previously been used to distinguish successional vegetation types in a tropical savanna-forest mosaic in Gabon (Marselis et al., 2018). We expect that these profiles also provide information on the occupation of available niche space between the forest floor and the top of the canopy and can be related to tree species diversity. The implication is that mapping tree species diversity in the tropics may then be feasible due to the launch of GEDI, which will collect billions of lidar waveforms between 51.6° latitude north and south of the equator, in December 2018. Data from other spaceborne missions, such as the TanDEM-X radar and the ICESat-2 lidar missions, may aid in the mapping process by gap-filling GEDI data products, and increasing the spatial resolution and extent of these data (Lee et al., 2018; Qi and Dubayah, 2016).

The overall goal of this study is to explore the structure-diversity relationship in a tropical forest and savanna landscape in Gabon. In particular, we seek to understand whether vertical canopy structure explains more of the variance in tropical tree species diversity than canopy height alone and whether we can use this relationship to predict tree species diversity across Gabon using different remote sensing datasets.

The remainder of the study is organized as follows: (i) we first create linear models to relate canopy height from lidar waveforms to tree species diversity from field data; (ii) we then use two sets of metrics describing the vertical canopy profile and relate these to the tree species diversity; (iii) we then compare the performance of these models; (iv) and use the height model to predict tree species diversity from LVIS canopy height and from GEDI-TanDEM-X fusion height products in Mondah; (v) and we apply the vertical structure model to gridded LVIS products and vertical profiles derived from ICESat to predict tree species diversity across and between four study sites.

3.2 METHOD

In this section, we introduce the four study sites and the datasets used. We explain the processing of the field and lidar data, followed by the model development for testing the structure-diversity relationship and the application of these models to create spatial predictions of tree species diversity from different active remote sensing data products.

3.2.1 FIELD DATA PROCESSING

Gabon was chosen as case-study location because of the large availability of field and remote sensing datasets collected during the AfriSAR campaign (Fatoyinbo et al., 2017). The four study sites in Gabon are referred to as Lopé, Mondah, Mabounié and Rabi (Figure 8) and each site includes multiple sampling plots. All sites fall within the general classification of tropical *terra firme* broadleaf forest, but have different species compositions and assemblies, disturbance history, and management regimes. The plots in this analysis are distributed between savanna, successional, degraded and old-growth tropical forest and are situated in climate zones with different annual precipitation and temperature (Table 2; Chapter II.1 Dataset details). In Lopé twelve stem mapped field plots (nine 1 ha, three 0.5 ha) were established in 2016 as part of the AfriSAR campaign. These data are available through ForestPlots.net (Labrière et al., 2018; Lopez-Gonzalez et al., 2011, 2009). The Mondah data were also collected during the AfriSAR campaign in 2016 and are publicly available through the NASA Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) (Fatoyinbo et al., 2018). The site comprises fifteen 1 ha stem mapped plots, thirteen of which are coincident with the LVIS lidar data. We excluded three of the thirteen plots from the Mondah site that consisted of a high percentage of unidentified trees (Chapter II.1 Dataset details). Twelve 1 ha field plots were established in Mabounié in 2012 (Labrière et al., 2018), of which 10 are coincident with LVIS lidar data and suitable for this study. A 25 ha plot comprises the Rabi site. These data were collected by the Smithsonian Conservation Biology Institute, National Zoological Park, and the Forest Global Earth Observatory (ForestGEO), and are available on

request through the ForestGEO website¹. From the Rabi plot, we selected the thirteen non-adjacent hectares and considered them as separate plots for the analysis. Subsampling kept this study site from dominating the models, as the Rabi site consisted of a total sampled area twice as large as the other sites (Figure 9a). Field data collection at the four sites was performed by different people and organizations, which led to datasets with varying characteristics (in terms of e.g., plot layout and minimum DBH measured). Details about the study sites and field datasets are described in (Chapter: II.1 Dataset details). For consistency among the sites we only included trees with $DBH \geq 10$ cm in our analysis. The availability of stem maps (in Lopé, Mondah, and Rabi) and subplots (in Mabounié) enabled testing of the structure-diversity relation at different resolutions. This was necessary because species diversity and plot size are not linearly related (species-area curve, described by MacArthur and Wilson (1967) and no optimal resolution has been identified for the structure-diversity relation. Smaller plots were created by subdividing each original plot into smaller squares or rectangles to create 5 spatial resolutions: 1, 0.5, 0.25, 0.0625 and 0.04 ha (Figure 9b). Mabounié data were only used at the 1 ha and 0.04 ha resolution, due to the absence of stem maps. At each resolution only plots with at least one identified tree were included in the analysis. This resulted in respectively 41, 64, 128, 481 and 935 plots for the resolutions 1.0 ha through 0.04 ha. Tree species diversity was then quantified for each of the plots at all resolutions using two

¹ <https://forestgeo.si.edu/sites/africa/rabi>

variables: the Shannon diversity (H') and tree species richness (S) expressed as the total number of tree species per area (Morris et al., 2014).

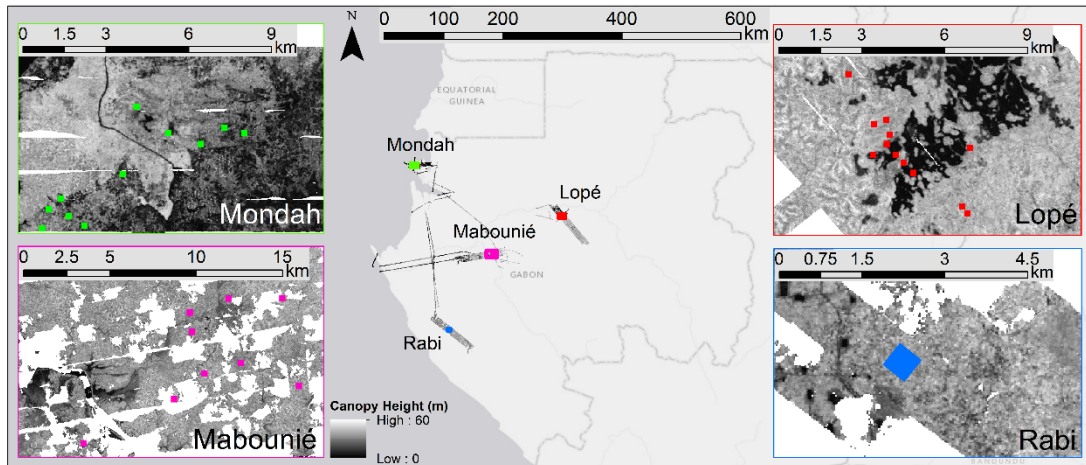


Figure 8: Field and lidar datasets from four regions (Mondah, Mabounié, Rabi and Lopé) in central and west Gabon are used in this study. LVIS lidar acquisitions are displayed as gridded canopy height (m). Insets show distribution of field plots across each study site, markers are not to scale of field plot size.

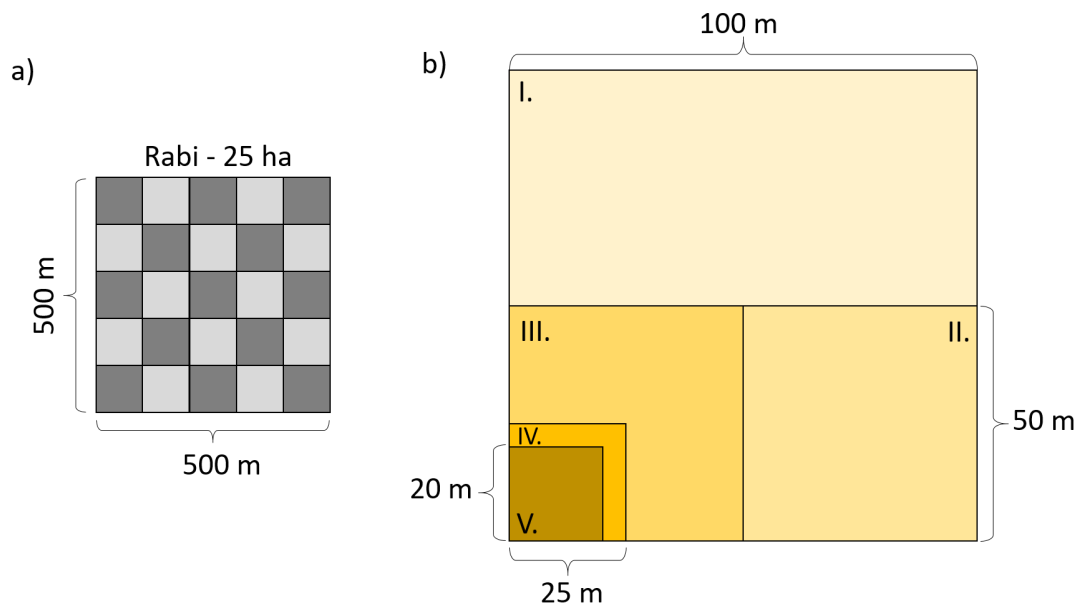


Figure 9: (a) Thirteen 1 ha plots (dark grey) were selected from the 25 ha Rabi plot and included in this study. (b) Subdivision of 1 ha plots to allow for analyzing the

structure-diversity relationship at 5 spatial resolutions led to plots of sizes I. (100x100 m), II. (100x50 m), III. (50x50 m), IV. (25x25 m) and V. (20x20 m).

Table 2: Information for each of the study sites regarding the total sampled area, temperature and precipitation, the tree density, species richness and Shannon diversity calculated including only trees with DBH \geq 10 cm, at the 1 ha resolution.

Name	Total Area (ha)	Mean Annual Temperature (°C)	Mean Annual Precipitation (mm/y)	Total no. of trees	Tree density (trees/ha)			Total no. of species	Species richness (S) (No. species/ha)			Shannon diversity (H') (H'/ha)		
					Range	Mean	Std		Range	Mean	Std	Range	Mean	Std
Lope	10.5	26-33 ¹	1500 ¹	3140	9-501	308	174	118	2-54	32	21.6	0.64-3.08	2.15	0.90
Mondah	10	25 ²	3000-3500 ²	2368	26-453	237	173	139	7-75	32	21.3	1.46-3.58	2.60	0.69
Mabounié	10	26 ²	2030 ²	3537	222-444	354	61	183	44-68	54	7.7	2.90-3.61	3.27	0.23
Rabi	13	24-28 ³	2300 ³	6056	231-689	466	119	211	55-101	82	13.5	3.35-4.04	3.70	0.18

¹White (2001), ²Labrière et al. (2018), ³Lee et al. (2006).

3.2.2 LIDAR DATA PROCESSING

Full-waveform lidar data were collected using the LVIS instrument with a 1 km swath width, a 1064 nm wavelength and a nominal footprint diameter of ~20 m across all study sites as part of the AfriSAR campaign in February and March 2016. The following metrics were derived from all lidar waveforms using mature, validated and published algorithms: Relative Height 100 (RH100), Plant Area Index (PAI) at 5 m vertical resolution forming the Plant Area Volume Density (PAVD) profile and cumulative PAI at 5 m vertical resolution (Blair et al., 1999; Marselis et al., 2018; Tang et al., 2012). These footprint-level metrics are publicly available as part of the AfriSAR deliverables through the NASA ORNL DAAC (Tang et al., 2018). RH100 was used to represent canopy top height. We chose the PAVD and cumulative PAI profile because they describe the vertical vegetation structure as a biophysical property that can also be derived from data from other instruments such as ICESat and GEDI, which enables transferability of the models. Additionally, they have previously been used to understand structural and functional differences between vegetation communities (Cuni-Sanchez et al., 2016; Decuyper et al., 2018; Marselis et al., 2018). For each plot, the average of all waveform metrics and the standard deviation of Canopy Height (RH100_sd) were calculated using all waveforms with the footprint center located within the plot boundary.

3.2.3 DATA ANALYSIS

The field and lidar data processing resulted in five datasets, one for each spatial resolution, comprised of n samples, with n equivalent to the number of plots at the specific resolution. Each sample had one value for S, H' and each of the lidar metrics. Four sets of lidar metrics were created to test the structure-diversity relationship (Table 3). We built linear models using metric sets 1 through 4 at each of the five spatial resolutions. Variable importance could not be rigorously assessed for the individual metrics due to collinearity between the metrics describing the vertical profile, since the existence of plant material higher in the canopy is dependent on the existence of material lower in the canopy. For each model, we performed leave-one-plot-out cross-validation. All plots within a 1 ha plot were kept aside in the cross-validation approach for models built at the 0.5, 0.25, 0.0625 and 0.04 ha resolutions, to avoid overestimation of model performance by the model being trained towards the local variation within the 1 ha plot. For each model, we calculated the R-squared (R^2) and the Root Mean Squared Difference, as a percentage (RMSD%) of the mean observed S and H', from the cross-validated predictions to evaluate model performance (Piñeiro et al., 2008). Bias was evaluated as the intercept of the linear relation between predicted (x-axis) and observed (y-axis) S and H'. An intercept not significantly different from 0 indicated an unbiased prediction. The slope of the linear relation indicated the consistency of the predicted and observed variables. A slope significantly different from 1 indicated no consistency (Piñeiro et al., 2008). We used 95% confidence intervals around the R^2 to evaluate whether metric sets 3 and 4

(canopy structure) led to significantly better diversity predictions than sets 1 and 2 (canopy height).

We used the canopy height and standard deviation of canopy height model at the 0.25 ha resolution to predict S and H' for Mondah from two datasets: 1) the canopy height product from LVIS gridded at 50 m and 2) a simulated GEDI-TanDEM-X fusion height product at 20 m resolution (Lee et al., 2018). TanDEM-X is a twin X-band SAR satellite mission from which the HH polarization was used for the fusion product. GEDI simulations were created from the LVIS waveforms using the GEDI simulator, considering 50% cloud cover (Hancock et al., 2019; Lee et al., 2018). We resampled the GEDI-TanDEM-X product to 50 m resolution and computed mean and standard deviation of forest height for each cell.

We used the PAVD (starting at P10) and canopy height model at 0.25 ha resolution to predict S and H' for all four regions. We applied these models to 1) LVIS data products gridded at 50 m resolution and 2) ICESat waveforms collected within and between the four study sites. We used all ICESat waveforms collected between 2004 and 2006. P5 was omitted as it has a low accuracy due to the interference of slope with vegetation close to the surface within the nominal footprint of 50 m.

Table 3: Sets of metrics used in linear models predicting H' and S.

Metric set	Model name	Explanatory variables
1	Height	RH100
2	Height and standard deviation of height	RH100 + RH100_sd
3	PAVD profile + height	P5*, P10, P15, P20, P25, P30, P35, P40, P45, P50**, RH100
4	Cumulative PAI profile + height	CP5***, CP10, CP15, CP20, CP25, CP30, CP35, CP40, CP45, CP50, RH100

*P5 = PAI between 0 and 5 m aboveground, subsequently P10 = the PAI between 5 and 10 m etc.

** P50 = PAI between 45 to 50 meters above ground, in case vegetation is higher than 50 meters this value indicates the PAI above 45 meters.

*** CP5 = Cumulative PAI between 0 to 5 m aboveground, subsequently CP10 = the PAI between 0 and 10 m, etc.

3.3 RESULTS

We first provide the results of the linear models using canopy height alone to predict S and H'. Then we provide the results of the models using the vertical canopy profile metrics to predict S and H'. Last, we show the spatial predictions of S and H' created using the calibrated and validated models.

3.3.1 MODELING RESULTS USING CANOPY HEIGHT

The cross-validated linear models using canopy height (set 1) explains a limited percentage of the variance in H' and S. At the 1.0 ha resolution, 28% and 35% of the variation in H' and S is explained ($R^2=0.28$ and $R^2=0.35$). At the 0.25 ha resolution $R^2=0.43$ and 0.44 (Figure 10; Chapter II.2 Model performance). Incorporating information on canopy height variation (set 2) increases R^2 to 0.58 and 0.52 at the 0.25 ha resolution. However, the R^2 confidence intervals comparing the model performance using metric set 1 and 2 are overlapping at all spatial scales, suggesting that adding information on the variance in canopy height within a plot does not necessarily lead to better predictive models (Figure 10; Chapter II.2 Model performance).

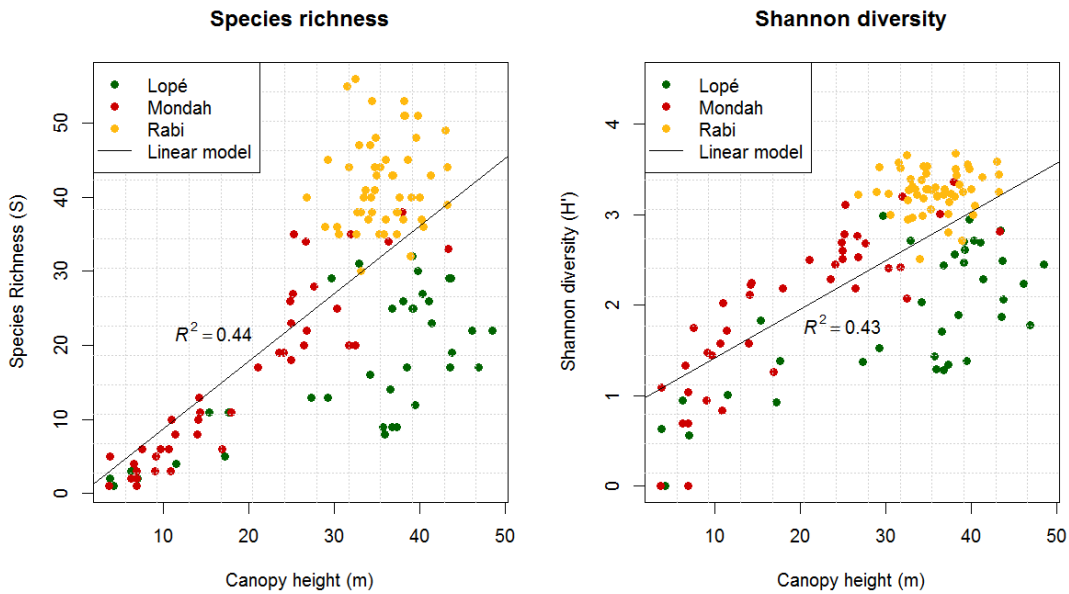


Figure 10: Canopy height explains up to 44% of the variation in tree species richness and 43% of the variation in Shannon diversity across the study sites in Gabon (cross-validated results).

3.3.2 MODELING RESULTS USING CANOPY STRUCTURE

Including information on the vertical canopy profile (set 3) increases the explained variance in both H' and S to 71% at the 0.25 ha resolution, a significant improvement in model performance (Figure 11; Chapter II.2 Model performance). Metric sets 3 and 4, including two different expressions of the vertical canopy profile, generally lead to similar results and only the results of set 3 are discussed in the remainder of the paper (SI2). Our results show the best model performances for the structure-diversity relation (highest R^2 , lowest RMSD% and unbiased predictions) at the 0.25 ha resolution, for both H' and S models (Figure 11; Figure 12; Chapter II.2 Model performance).

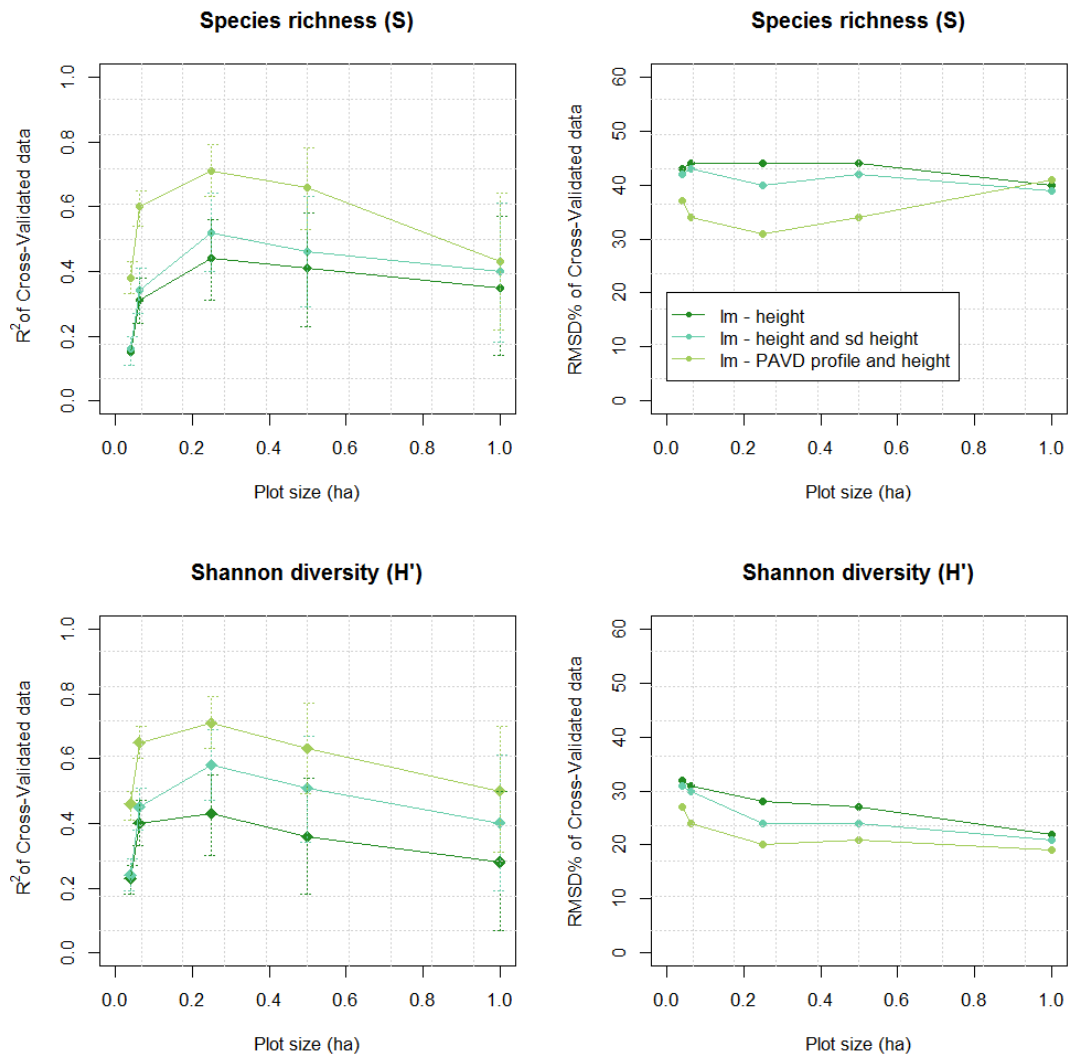


Figure 11: Linear models using height alone as a predictive variable generally show the lowest R^2 and highest error (RMSD %). Adding variation in canopy height (RH100_sd) to the model leads to slightly better, but not statistically different, model performance (see overlapping error bars on the R^2 plots). Incorporating information on canopy structure significantly improves the S and H models at the 0.04 to 0.25 ha resolutions, when compared to using height alone. Error bars are larger at larger plot size because of smaller sample sizes.

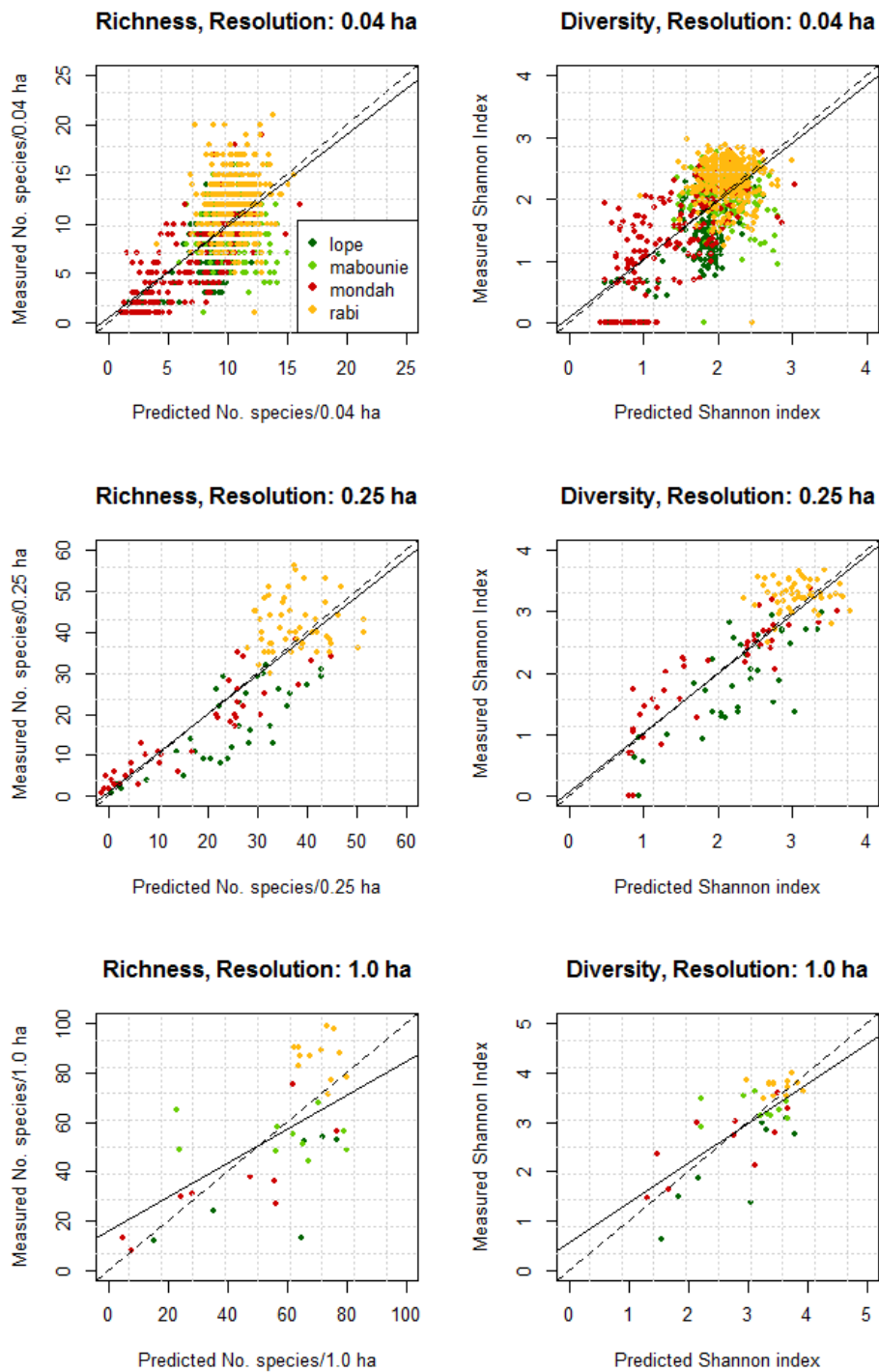


Figure 12: Predicted vs. observed species richness and Shannon diversity resulting from the models using metric set 3 (PAVD profiles and height), points are colored by study site. 1.0 Ha resolution models are significantly biased, 0.25 ha models show highest accuracy and statistically unbiased predictions for both species richness and Shannon diversity.

3.3.3 PREDICTIVE MODELING RESULTS

The 0.25 ha linear models using metric set 2 (model performance for H' : $R^2 = 0.58$, $\text{RMSD} = 24\%$; and S ; $R^2 = 0.52$, $\text{RMSD} = 40\%$) were used to map H' and S in Mondah from 1) LVIS gridded products and 2) GEDI-TanDEM-X gridded products (Figure 13). GEDI-TanDEM-X derived predictions for both species richness and Shannon Diversity are consistent with the LVIS predictions ($R^2 = 0.68$ and $R^2=0.66$) but overall biased slightly low (Figure 13).

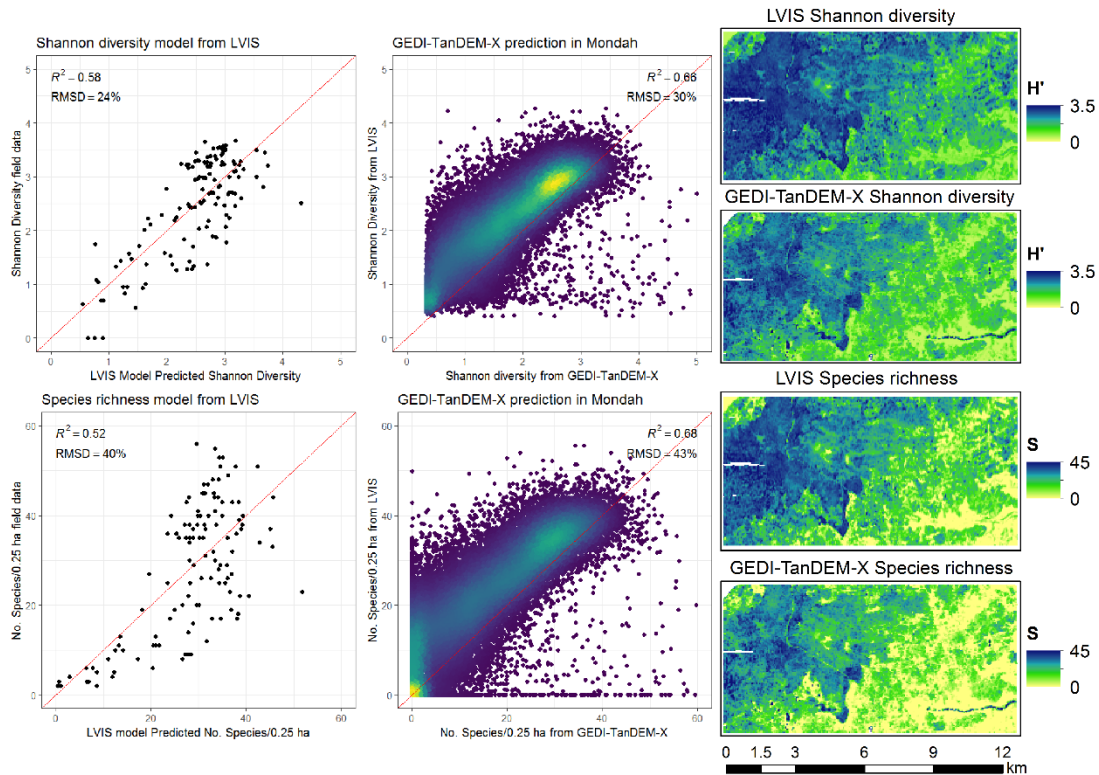


Figure 13: Shannon diversity and species richness predicted in Mondah using canopy height (RH100) and standard deviation of canopy height (RH100_sd). Predictions created from gridded LVIS data products and a simulated GEDI-TanDEM-X fusion product. Points in center panel are colored by density.

Spatial predictions of H' and S for all study sites were created from the LVIS wall-to-wall data products and ICESat waveforms using the 0.25 ha linear models with the PAVD profile starting at P10 to P50 and canopy height (model performance for H' : $R^2 = 0.71$, RMSD = 20%; and S ; $R^2 = 0.71$, RMSD = 31%) (Figure 14). The ICESat footprints overlapping with the wall-to-wall lidar show ICESat predictions of S and H' are slightly higher than predictions from LVIS. Shannon diversity predictions (with $R^2 = 0.66$ and RMSD = 26%) are more accurate than the species richness predictions (with $R^2 = 0.62$ and RMSD = 38%).

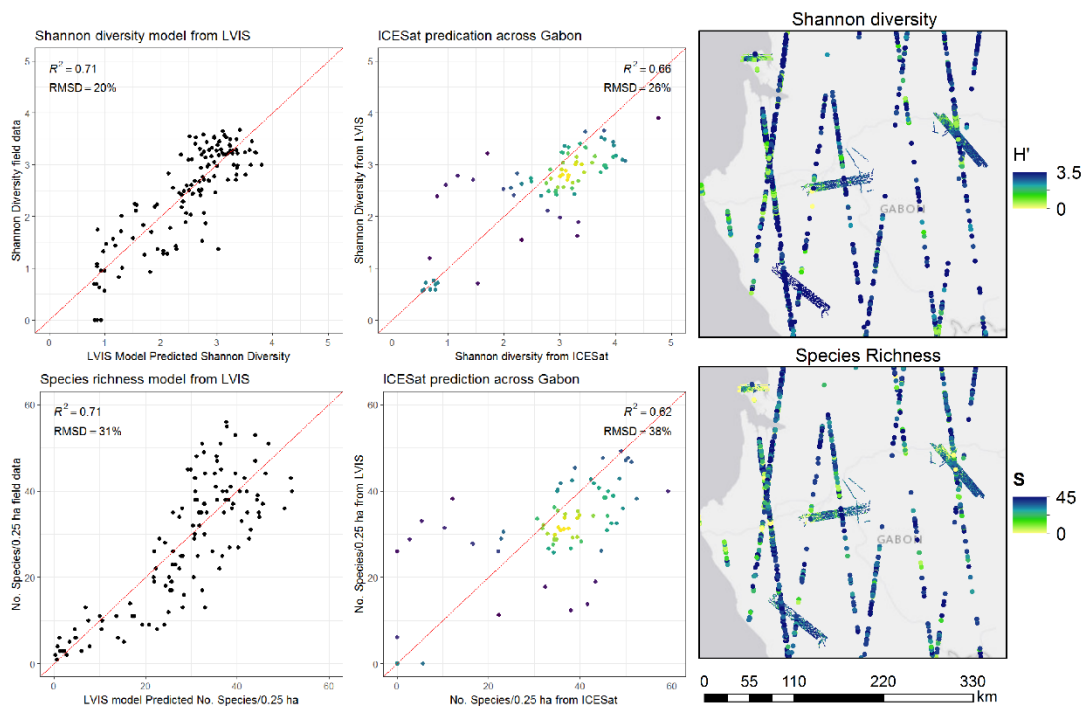


Figure 14: ICESat data enabled predictions of H' and S within and between the study sites. ICESat richness predictions were biased slightly high compared to LVIS predictions. Points in center panel are colored by density.

3.4 DISCUSSION

Our study presents regional results relating canopy height and vertical canopy structure, derived from active remote sensing data, to tree species diversity in African tropical forest. We first discuss the modeling results and the relationship between canopy structure and tree species diversity. Then we discuss some limitations to our approach and close with an outlook on the applications of the structure-diversity models.

3.4.1 MODEL PERFORMANCE

Canopy height has been hypothesized to be related to tree species diversity by assuming canopy height is a proxy for canopy volume, representing the available niche space. Our study shows a significant relationship between canopy height and the Shannon index and species richness; affirming that canopy height explains a limited percentage of the variation in tree species diversity. These results are in line with previous results found in the USA and globally by (Gatti et al., 2017; Marks et al., 2016). We expected that including information on the variance in canopy height within a plot might lead to better predictive models as the combination of metrics provides a more accurate proxy for forest volume than canopy height alone. Our models show an increase in R^2 across all resolutions, but the improvement is not statistically significant. We then hypothesized that including information about the occupancy of vertical space, implemented by including the PAVD profile in the canopy height model, would lead to better predictions of tree species diversity. The PAVD profile is a direct measure of the degree of occupation of vertical niche space (divided in 5 m height bins) by plant material (PAI index) and thus resembles

occupied niche space instead of available niche space. Our results support the hypothesis as we find a statistically significant improvement in model performance at the 0.04 – 0.25 ha resolutions when including the PAVD profile. At larger plot sizes (0.5 and 1.0 ha) the R^2 values increased but not significantly. We tested the structure-diversity relation at different resolutions because neither species richness nor Shannon diversity is linearly related to plot size (Hill, 1973; MacArthur and Wilson, 1967). Our models performed best at the 0.25 ha resolution (a decrease in performance was found for both smaller and larger plot sizes). We consider two potential explanations for this phenomenon. The first one is related to the quality of measurement: large plots tend to have greater sampling errors caused by a smaller sample size, while smaller plots are less precise due to a higher influence of geolocation error (though minimized by (Labrière et al., 2018), a small misalignment between plot and lidar data may still result in poorer model performance). The second potential explanation is the maximum natural variation in the structure-diversity relation at an intermediate resolution. In a small plot only a limited number of trees could possibly grow due to space constraints, reducing the total variability or upper boundary of tree species diversity. By contrast, the amount of different tree species may get close to a certain maximum in a large plot (as species-area curves suggest) while similar canopy structure may be found as in smaller plots. In our study, it is at an intermediate resolution that the largest variation in tree species/plot was found, as indicated by the highest coefficient of variation in tree species richness at the 0.25 ha resolution (Chapter II.3 Coefficient of variation). This trend is consistent with the trend of canopy structure with its maximum coefficient of variation at the 0.25 ha

resolution. Yet, it is difficult to decouple the two factors due to the limited data sets in this study, and additional research would be needed to better identify the ecological mechanisms behind this phenomenon. Nonetheless, the modeling results for predicting Shannon diversity and species richness follow the same trend - for both diversity variables the PAVD profile with height provides a significantly better model than using height alone at the 0.25 ha resolution – strengthening our confidence in the existence of a structure-diversity relationship across these study sites in Gabon.

3.4.2 LIMITATIONS

Time lag between field and lidar data collection was minimized, but we had to accept a maximum time-lag of five years (Rabi) which may have affected the strength of the structure-diversity relationship. The new census of the Rabi field data, which is under collection at the time of writing, may provide more insight on tree species diversity change over time and the effect of time-lag in the structure-diversity relation.

The diversity metrics used to establish the structure-diversity relation rely heavily on the accuracy of species identification in the field. Tree species identification in the tropics can be very complex and highly trained local botanists are needed to provide the most accurate identification. In most of our study sites (Rabi, Mabounié and Lopé) the percentage of trees of unknown species was low, but in Mondah this percentage was much higher because the species identification was carried out by botanists with knowledge of vascular plant taxonomy in Gabon, but not specifically for the Mondah region (Chapter II.1 Dataset details). Slik et al. (2015) showed that higher percentages of unidentified trees can affect model outcomes, which should be

kept in mind when evaluating our results and future research could assess the specific impact of unidentified trees on our models, for example by re-inventory of the trees in Mondah.

In our study, we were not able to include small trees because only two field inventory datasets included trees with $DBH \geq 1$ cm, thus we included only trees with $DBH \geq 10$ cm to allow for consistency across all study sites. However, Fricker et al. (2015) showed that the inclusion of small trees in Barro Colorado Island (Panama) improved their model performance (relating species richness to various metrics describing terrain, hydrology and canopy structure) from $R^2 = 0.25$ when using trees with $DBH \geq 10$ cm to $R^2 = 0.35$ when using trees with $DBH \geq 1$ cm. More information on small trees will be needed to assess their specific impacts on the models created in this study.

The plots included in the analysis were distributed across different vegetation types: savanna, forest with different degrees of degradation, successional forest and low-disturbance old-growth tropical forest. It is likely that the encountered structure-diversity relationship is driven partly by this gradient in forest types. More undisturbed old-growth tropical forest plots would be needed to study the structure-diversity relationship within old-growth forest. Additionally, it is yet unclear how the structure-diversity relationship holds or changes across different biogeographic regions, with changing limitations to tree growth, and different climatic niches that may change the canopy structure and the use of vertical niche space. Future research should focus on the transferability of the structure-diversity models to other regions

and continents to establish the potential of using this method pan-tropically or potentially across multiple biomes. Lastly, studying how the structure-diversity relationship changes between the local-scale, such as in a large plot like BCI (Fricker et al., 2015; Wolf et al., 2012), and at a regional scale, such as developed here and in (Mao et al., 2018; Robinson et al., 2018), will be important to ultimately evaluate the structure-diversity relationship pan-tropically.

3.4.3 APPLICATIONS & FUTURE RESEARCH

The commissioning of new satellite missions, providing information on vertical canopy structure, is an important incentive to explore the potential applications of the structure-diversity relationship. Here we demonstrated the application of the developed models to wall-to-wall LVIS data products to create predictive maps of diversity and richness in the four study sites. The spatial area covered by these wall-to-wall products is limited due to the airborne nature of the data and data gaps occur as the laser energy does not penetrate clouds, which are often prevalent over Gabon. But, the real strength of our modeling approach lies in its potential to be applied to data from other instruments such as ICESat, GEDI, and TanDEM-X. The ICESat predictions overlapping with the LVIS wall-to-wall products showed S and H' predictions around the 1:1 line. The scatter around the 1:1 line is partly caused by the model accuracy, the time-lag between these two data products (collected 10-12 years before the LVIS data) and the accuracy of the ICESat data products. But, the major drawback for using ICESat is the sparse sampling, leaving tens of kilometers between tracks over Gabon. On the contrary, the recently launched GEDI lidar mission will provide much denser lidar sampling with between-track spacing of 600 m and along-

track spacing of 60 m. However, the nominal footprint size (~22 m diameter) is close to the smallest resolution tested here and our modelling results showed a higher predictive error compared to the models that can be applied directly to ICESat data. On the other hand, fusion of the GEDI waveforms with other active remote sensing data, such as from TanDEM-X radar, has the potential to change this resolution, fill in data gaps and provide wall-to-wall data products allowing for the application of models developed at different resolutions (Lee et al., 2018; Qi and Dubayah, 2016). It may then also be possible to build structure-diversity models directly on fused GEDI-TanDEM-X data products when these data become available more widely after GEDI's data collection. So far, it has been shown that canopy height can be retrieved using GEDI-TanDEM-X fusion, but, if we would be able to extract more information on the vertical canopy profile using GEDI and TanDEM-X data fusion, or fusion with another interferometric sensor such as on the BIOMASS mission, it may be possible to accomplish true wall-to-wall mapping of tree species diversity in the pan-tropics and beyond.

3.5 CONCLUSION

Canopy height alone, used as a proxy for niche volume, can predict a limited percentage of the variation in tree species diversity across a savanna-tropical forest landscape in Gabon, Africa. Including information on the vertical canopy structure, used as a proxy for vertical niche occupation, derived from large-footprint full-waveform lidar data improved these models significantly. This structure-diversity relationship and the models developed here show potential for application to data

from active remote sensing satellites such as ICESat, TanDEM-X, BIOMASS and GEDI. Further research is encouraged to study the structure-diversity relationship across the tropics to establish the potential of mapping pantropical tree species diversity using measurements of canopy structure.

ACKNOWLEDGEMENTS

This work is supported by NASA Headquarters under the NASA Earth and Space Science Fellowship Program – Grant 80NSSC17K0321, under the NASA New Investigator Program - Grant 80NSSC18K0708, under NASA's Global Ecosystem Dynamics Investigation (GEDI, Grant NNL15AAO3C) and NASA's AfriSAR campaign. The AfriSAR campaign was conducted in Gabon in collaboration with l'Agence gabonaise d'études et d'observation spatiale (AGEOS) and the Agence Nationale des Parcs Nationaux (ANPN). Research permission was granted by the Centre National de la Recherche Scientifique et Technologique (CENAREST - permit numbers AR0011/17/MESRSFC/CENAREST/CG/CST/CSAR and AR0024/10/MESRDT/CENAREST/CG/ CST/CSAR) and the Agence Nationale des Parcs Nationaux (ANPN). Lidar data collection in Lopé, Mondah, Mabounié and Rabi was carried out by the NASA Land Vegetation and Ice Sensor (LVIS) team led by Bryan Blair, Michelle Hofton and David Rabine and hosted by AGEOS (Ghislain Moussavou, Tanguy Gahouma), we express our sincere gratitude to their team for their effort to collect these data.

We are grateful to ANPN and AGEOS for the provision of logistical support that facilitated both field work and lidar data collection, specifically to Flore Koumba Pambo Josue Edzang Ndong and David Lehmann from ANPN for their support. The field data collection in Lopé was funded by ESA through the AfriSAR campaign, coordinated by Simon Lewis and guided by Nicolas Labrière, hosted by ANPN and the University of Stirling at the SEGC field station, we would like to specifically thank Carl Ditougou, Pacôme Dimbonda, Arthur Dibambou, Edmond Dimoto, and

Napo Milamizokou. The field data collection in Mondah was funded by NASA through the AfriSAR campaign and coordinated and guided by John Poulsen, hosted by ANPN. The field data collection in Mabounié was led by Nicolas Barbier and performed in collaboration with the Missouri Botanical Garden (Tariq Stevart) and Golder Associates and we would specifically like to thank P. Ploton, V. Droissart, and Y. Issembe for their help and expertise. The field data collection in Rabi was funded through Shell Gabon and the Smithsonian Conservation Biology Institute. The data were collected under guidance of Alfonso Alonso, Hervé Memiaghe, David Kenfack and Pulcherie Bissiengou. This is contribution No. 171 of the Gabon Biodiversity Program. We express our sincere gratitude to everyone involved in these field campaigns and for making the data available for this study.

4. EVALUATING THE POTENTIAL OF FULL- WAVEFORM LIDAR FOR MAPPING PAN- TROPICAL TREE SPECIES RICHNESS

ABSTRACT

Mapping tree species richness across the tropics is of great interest for effective conservation management to prevent widespread species extinction. In this study, we evaluate the potential of full-waveform lidar data for mapping tree species richness across the tropics by relating measurements of vertical canopy structure, as a proxy for the occupation of vertical niche space, to tree species richness. First, we evaluate the characteristics of the vertical canopy structure across 15 study sites using simulated full-waveform lidar data and relate these findings to in-situ tree species information. Then, we develop structure-richness models at the local, regional and pan-tropical scale at three spatial resolutions (1.0, 0.25 and 0.0625 ha) using Poisson regression. The results show only a weak structure-richness relation at the local scale (within 25-50 ha plots). At the regional scale a stronger relationship was found between canopy structure and tree species richness across different tropical forest types; such as across our study sites in Gabon, across Central Africa and in South America (R^2 ranging from 0.49-0.54, RMSE ranging between 18-57%, at 0.0625 ha resolution). A weaker but unbiased and significant relationship was found at the pan-

tropical scale, including data across four continents ($R^2 = 0.41$ and $RMSE = 39\%$, 0.0625 ha resolution). The 0.0625 ha resolution (25 x 25 m square plots) corresponds well to footprint-level measurements collected with the Global Ecosystem Dynamics Investigation (GEDI) launched in late 2018. Our results may thus serve as a basis for future development of structure-richness models to map tree species richness using vertical canopy structure information from GEDI lidar data. Yet such effort is contingent on the availability of a larger set of field reference data in South America and South-East Asia. Future research could also support the use of GEDI canopy structure data in frameworks using environmental and spectral information for modelling tree species richness across the tropics.

4.1 INTRODUCTION

Tropical forests are known for their high tree species diversity. Current estimates find at least 40,000 different tree species across the tropical region, but this number may be even higher than 53,000 species, in contrast to the 124 tree species growing in temperate forests (Slik et al., 2015). Trees are the backbone of tropical forest ecosystems. They play a crucial role in regulating global and local climate (e.g. by absorbing carbon dioxide from the atmosphere and by regulating water availability and runoff); providing a habitat to a variety of mammals, insects, and other organisms, and supporting the livelihoods of indigenous communities (Watson et al., 2018). Unfortunately, 35% of pre-agricultural global forest cover has been lost over the past 300 years, largely due to increasing human pressures on the environment, and 82% of the remaining forest is estimated to experience some degree of degradation

(Watson et al., 2018). Current extinction rates are estimated to be at least 1000 times higher than background extinction rates (Pimm et al., 2014), and it was recently estimated that in the Amazonian tropics alone around 25% of the tree species are threatened for extinction (Ter Steege et al., 2015).

The scientific community has called for bolder science in conservation strategies to enable effective management of the Earth's forests and allow for better conservation of our natural ecosystems (Watson et al., 2016). The Convention of Biological Diversity (CBD) and GEOBON have developed a list of important variables aiming to provide quantitative information on biodiversity to reach the Aichi biodiversity targets 2020 (Pereira et al., 2013; Skidmore et al., 2015). Among these Essential Biodiversity Variables (EBVs) is the need for mapping taxonomic diversity (Pereira et al., 2013). In this paper we focus on the use of active remote sensing techniques to meet this requirement for mapping taxonomic tree species diversity in the tropics.

Some of the pioneering work quantifying the number of tree species across the globe dates back about 15 years ago (Kier et al., 2005; Mutke and Barthlott, 2005). Kier et al. (2005) mapped the number of different tree species by ecoregion based on field inventories, and Mutke and Barthlott (2005) created a continuous raster quantifying tree species richness for 10,000 km² cells. More recently, remotely sensed data has become an essential part of mapping tree species richness in the tropics by providing consistent information across large scales (Keil and Chase, 2019). This use of remote sensing data can be categorized in two ways: first, remote sensing data is used to measure species, community or ecosystem-scale patterns directly. Second,

biodiversity patterns can be modeled indirectly using remote sensing data as predictive variables in combination with field observations (Anderson, 2018). Regression analysis using spectral data and field reference data is probably the most widely used method to map species diversity (Rocchini et al., 2016). For example, Foody and Cutler (2006) related a limited set of reference tree species richness data from the Danum Valley in Malaysia to Landsat TM information from the six non-thermal wavebands using a neural network approach ($n=10$, $r=0.69$). Schäfer et al. (2016) mapped tree species richness in a tropical montane forest in the Taita Hills, Kenya, using a segmentation and clustering method based on spectral reflectance information from 129 spectral bands resulting in $R^2 = 0.5$. Even though these methods have been progressively developing over the last decade, they are not yet operational for mapping tree species richness due to either insufficient accuracy or high spatial resolution across the tropics.

Recently, canopy structure components measured using active remote sensing systems have been proposed as potentially useful for mapping tree species richness (Marselis et al., 2019; Robinson et al., 2018; Wolf et al., 2012). Several studies have shown promising results using this method in different parts of the tropics: Plant species richness was related locally to canopy height and topography derived from airborne lidar data in Barro Colorado Island, Panama (Wolf et al., 2012) and along an elevation gradient in Costa Rica (Robinson et al., 2018). Marselis et al. (2019) showed spatial mapping of tree species richness in different parts of Gabon, Africa, which was possible using information on canopy height and the vertical canopy

structure derived from full-waveform airborne lidar data. The growing use of canopy structure data for estimating tree species richness could potentially become operational at a pan-tropical scale with the rapidly increasing availability of spaceborne canopy structure information derived from the Global Ecosystem Dynamics Investigation (GEDI), a full-waveform spaceborne lidar system (Dubayah et al., under review). GEDI is expected to provide over 10 billion measurements of vertical canopy structure across the temperate and tropical forests between 2019 and 2021.

The main goal of this study is to evaluate the efficacy of full-waveform lidar for mapping tree species richness across the tropics. We address this goal by studying the following: First, we compare characteristics of the vertical canopy structure of tropical forests across the world. Then, we evaluate the differences in species richness and species-area curves across the different study sites. Subsequently, we evaluate the potential of developing local, regional and pantropical structure-richness relationships, relating canopy structure metrics to tree species richness measurements from the field at three spatial resolutions. Last, we discuss the potential of full-waveform lidar data from GEDI for mapping tree species richness across the tropics using structure-richness relationships.

4.2 BACKGROUND

In this section we evaluate the existing literature on the biogeographical history, tree species diversity and canopy structure across the Earth's tropical regions to provide a framework for the potential existence of one or more structure-richness relationships.

Tropical forests are spread over five main biogeographical regions. The Neotropics in central and south America, tropical forest on the African mainland, tropical forests on the island of Madagascar, the Sundaland tropics in South-East Asia and the tropical forests in the region of Papua-New Guinea and Australia. Tree species diversity of the largest regions is ranked, in decreasing order; the Neotropics, Sundaland and lastly the African tropical forest (Corlett and Primack, 2011). Historical biogeographical processes are known to be the most important driver of tree species diversity patterns at the continental scale (Keil and Chase, 2019). Tree species composition in the Neotropics evolved to be the most distinct of the three, as Central and South America were separated from Africa and South-East Asia about 70 million years ago when they broke off from the Gondwana continent. The floristics in Africa and South-East Asia are more similar to each other, although Africa only has 10-20% of the species in families shared between the two regions. The drought resulting from the latest ice age is expected to have been the primary cause of the reduced diversity in Africa, and also resulted in the characteristic African savanna-forest ecosystem (Corlett and Primack, 2011).

Early spatial analyses of the vascular plant species richness by Barthlott et al. (2005), Kreft and Jetz (2007) and Mutke and Barthlott (2005) used field data to estimate the distribution of species richness across the globe. According to their analyses, in general, the hotspots of vascular plant species richness can be found in Central America, the eastern Amazon, along the south-east coast of Brazil and the east coast of Madagascar, in the northern part of the island of Borneo, and along the northern

slopes of the Himalaya's, stretching east into south-east Asia (Mutke and Barthlott, 2005). Three maps of vascular plant diversity derived by (Kreft and Jetz, 2007) using different methods show alternative distributions with lower diversity along the east coast of Brazil and in south-east Asia's mainland. Both these analyses were based on field data, in which the spatial distribution of the available field data was a limiting factor in extrapolating species richness between field sites and across data-scarce areas. This extrapolation can be aided by known relationships between diversity and environmental factors. In general, tropical tree species diversity increases with increasing precipitation, forest stature, soil fertility, time since catastrophic disturbance and rate of canopy turnover and decreases with seasonality, latitude, and altitude (Givnish, 1999). At regional scales, topography was shown to be an important predictor in tree species diversity along an elevation gradient in Costa Rica with an elevation range of 33-2903 m, along a ~30 km transect (Robinson et al., 2018). At the local scale, within 25 – 50 ha plots, soil nutrients explained 36-51% of tree species richness in study sites in Colombia, Ecuador and Panama (John et al., 2007). Factors influencing tree species richness on a global scale differ from those affecting spatial patterns of species richness at regional or local scales. Large-grain studies show historical biogeography processes are more important, whereas smaller plot-scale studies have pointed to a strong role of environmental variables (Kreft and Jetz, 2007).

The list of variables that can be correlated with tree species richness at different scales is long (Keil and Chase, 2019). Moreover, the strength of the relationship

between a variable and tree species richness may also change with resolution (plot size) as tree species richness is not linearly related with the area measured (species-area curve) (Hubbell, 2001). This inconsistency complicates the development of reliable predictive models at a certain resolution, but also complicates the extrapolation of estimates at one resolution to a larger area, which complicates successful mapping of pantropical tree species richness at high spatial resolution.

Similarly to species richness, forest structure is influenced by a complex interaction of historic, environmental, and human related variables; precipitation in the wettest month being the most important single predictor of plant height at the global scale (Moles et al., 2009). Forest structure traditionally measured in the field is mainly comprised of four variables: canopy height, biomass, basal area/ha and tree density (Palace et al., 2015). However, active remote sensing techniques have revolutionized the way we look at canopy structure (Newnham et al., 2015). With lidar remote sensing, for example, it is now possible to not only obtain information on canopy height, but also on the position and amount of plant material along the vertical axis of the canopy (Tang et al., 2012). Palace et al. (2015) stressed that the high resolution lidar data possesses vertical structure information that is inherently linked to ecological processes and forest dynamics that are exhibited in the structure properties of the forest.

This wealth of information on the vertical structure has been hypothesized to be a proxy for vertical niche occupancy and to relate, to some extent, to tree species richness (Marselis et al., 2019). Different tree species require different niches to grow

and thus approximating a quantification of the occupation of that niche space may relate to the number of different tree species present in an area. In this study, we address the potential scale and resolution dependencies by exploring the relationships between vertical canopy structure and tree species richness at three spatial resolutions (0.0625, 0.25, and 1.0 ha) and across three spatial scales (local, regional and pan-tropical).

4.3 METHODS

In this study we address the structure-richness relation in *terra firme* forest in the tropical region between 23.5° N & S. We compiled a comprehensive field and lidar dataset covering savanna, colonizing forest, old-growth tropical forest and forests under different degrees of degradation. We first discuss the available field, airborne lidar, and simulated GEDI data. Subsequently, we explore the vertical canopy structure and the range of species richness across all study sites. Finally, we outline the development of local, regional and pan-tropical structure-richness models.

4.3.1 FIELD DATASETS

All field datasets used in this study have been previously collected and published and have coincident airborne lidar data available. We used 15 datasets: one in Australia, two in South-East Asia, six in Africa, three in South America and three in Central America (Figure 15). Each field dataset is indicated with a three letter code and contains information on tree location, species and Diameter at Breast Height (DBH) and. The field information was used to calculate the reference values of species richness used to evaluate the structure-richness relationships.

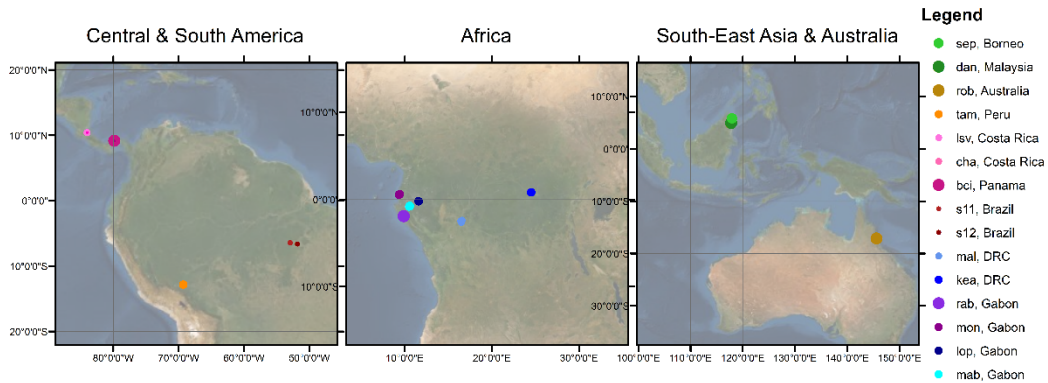


Figure 15: Location of field sites across the three continents, colors of each study site are consistent throughout paper. Gridlines indicate 10° intervals in longitudinal and latitudinal directions. The size of the place markers represents the relative size of the total sampled area.

All datasets were collected by different organizations and research teams resulting in different data characteristics (Table 4, Chapter III.1 Field data characteristics). Four datasets consisted of one large plot of 25 ha (*rob*, Australia and *rab*, Gabon) or 50 ha (*dan*, Malaysia and *bci*, Panama). The other eleven datasets consisted of multiple (3-21) smaller plots with sizes ranging from 0.16 ha to 4.0 ha.

Table 4: Information on the original plot size, the amount of total area sampled in the field and the source of the data which is either a website where the data is published and/or a publication in which the data is described.

Country	Project name	No. native plots	Total area (ha)	Source
Oceania				
Australia	<i>rob</i>	1	25	(Bradford et al., 2014)
South-East Asia				
Malaysia	<i>dan</i>	1	50	https://forestgeo.si.edu/sites/asia/danum-valley
Borneo	<i>sep</i>	9	36	(Jucker et al., 2018)
Africa				
DRC	<i>mal</i>	21	21	(Bastin et al., 2015)
DRC	<i>kea</i>	19	19	(Kearsley et al., 2013)
Gabon	<i>rab</i>	1	25	(Memiaghe et al., 2016)
Gabon	<i>lop</i>	11	9.5	(Labrière et al., 2018)
Gabon	<i>mon</i>	10	10	(Fatoyinbo et al., 2017)
Gabon	<i>mab</i>	10	10	(Bastin et al., 2015; Labrière et al., 2018)
South America				
Peru	<i>tam</i>	6	6	(Boyd et al., 2013)
Brazil	<i>s11</i>	9	1.44	https://www.paisagenslidar.cnptia.embrapa.br/
Brazil	<i>s12</i>	19	4.8	https://www.paisagenslidar.cnptia.embrapa.br/
Central America				
Costa Rica	<i>lsv</i>	12	6	https://tropicalstudies.org/carbono-project/
Costa Rica	<i>cha</i>	3	1.5	
Panama	<i>bci</i>	1	50	https://forestgeo.si.edu/sites/neotropics/barro-colorado-island

In this study we assessed the structure-richness relationship at three spatial resolutions (1.0, 0.25, 0.0625 ha) because of the non-linear relationship between number of tree species (species richness (S)) and sample area. These plot sizes were selected because square 1.0 ha plots (100x100 m) are often-used in forestry, especially in biomass studies, square 0.25 ha (50x50 m) because this seemed to be the most optimal resolution to describe the structure-diversity relationship in Gabon (Marselis et al., 2019) and 0.0625 ha because this is a standardized size close to the

GEDI footprint size, relevant when using the structure-richness relationship to map tree species richness with GEDI data. The datasets are used at one, two, or all three spatial resolutions to assess the structure-richness relationships, depending on the availability of stem maps or subplots (Chapter III.1 Field data characteristics). For each of the field sites we calculated the total number of tree species in the entire dataset and for each plot at each plot size (Table 5). Only live trees with a diameter at breast height (DBH) ≥ 10 cm were included, to create consistency among the datasets.

Table 5: The total number of species identified for each study site and the average (\bar{x}) and standard deviation (s) of the species richness for each of the three plot sizes expressed in last three columns as $\bar{x} \pm s$ (including only live trees with DBH ≥ 10 cm).

Country	Project Name	Total No. species	Total sampled area (ha)	Species richness 1.0 ha	Species richness 0.25 ha	Species richness 0.0625 ha
<i>Oceania</i>						
Australia	<i>rob</i>	205	25	98 \pm 10	56 \pm 8	27 \pm 5
<i>South-East Asia</i>						
Malaysia	<i>dan</i>	430	50	117 \pm 13	51 \pm 7	19 \pm 4
Borneo	<i>sep</i>	517	32	102 \pm 22	53 \pm 11	-
<i>Africa</i>						
DRC	<i>mal</i>	116	10.5	37 \pm 11	20 \pm 7	-
DRC	<i>kea</i>	232	19	50 \pm 23	24 \pm 13	10 \pm 6
Gabon	<i>rab</i>	234	25	84 \pm 8	42 \pm 6	17 \pm 4
Gabon	<i>lop</i>	118	9.5	32 \pm 22	17 \pm 10	8 \pm 4
Gabon	<i>mon</i>	146	10	32 \pm 15	15 \pm 9	7 \pm 5
Gabon	<i>mab</i>	196	10	55 \pm 8	-	-
<i>South America</i>						
Peru	<i>tam</i>	517	6	171 \pm 13	70 \pm 9	24 \pm 5
Brazil	<i>s11</i>	91	1.44	-	-	17 \pm 3
Brazil	<i>s12</i>	135	4.8	-	-	16 \pm 4
<i>Central America</i>						
Costa Rica	<i>lsv</i>	216	6	-	48 \pm 8	19 \pm 5
Costa Rica	<i>cha</i>	81	2	58	28 \pm 5	13 \pm 4
Panama	<i>bci</i>	220	50	87 \pm 8	42 \pm 6	17 \pm 3

4.3.2 LIDAR DATASETS

Each of the field data sets had coincident discrete return airborne laser scanning (ALS) data, or full-waveform lidar data from the Land Vegetation and Ice Sensor (LVIS), collected over the field plots within 5 years of the field data collection. We used the GEDI simulator (Hancock et al., 2019) to create lidar waveforms from the ALS data over the field plots. In this way, all lidar information could be processed in a consistent way across all study sites ensuring a reliable inter-comparison of canopy structure metrics derived from the waveforms and allowing for easy transfer of the developed models to future on-orbit GEDI data. Lidar waveforms were simulated with a 22 m ground footprint (Gaussian distribution of laser energy, $\sigma = 5.5$ m). Lidar waveform locations were determined by filling each field plot, using the original field plot size and shape, with footprint center locations 6.25 m from the plot edge and 5 m between footprint center locations (Figure 16). In this way, a reliable measure of canopy structure could be acquired for each plot by averaging lidar metrics from all waveforms in the plot, instead of using single waveforms in the plot center and evaluating structure-richness relationships based on such potentially biased or unrepresentative waveforms. The following information was extracted from each simulated lidar waveform using mature and published algorithms: Canopy height (expressed as the 98th percentile of the relative height metric; RH98), total Plant Area Index (PAI), and Plant Area Index at a 1 m vertical resolution (Drake et al., 2002; Hancock et al., 2019; Marselis et al., 2018; Tang et al., 2012). The 1 m vertical profile was aggregated into a 10 m vertical profile, summing all PAI values in each 10 m vertical bin. The average of each of these metrics from all waveforms within

each plot were computed to represent the canopy structure for each plot at each spatial resolution.

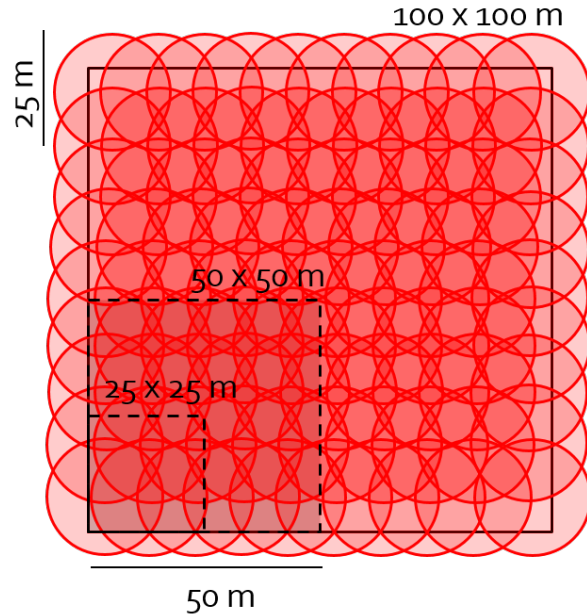


Figure 16: Illustration of simulated GEDI waveform layout. The GEDI waveforms (red circles) have a Gaussian energy distributed with $\sigma=5.5$ m, resulting in a roughly 22 m diameter footprint. Example of simulated footprint distribution locations in a 1.0 (solid outline), 0.25 and 0.0625 ha field plot (dotted outline).

4.3.3 CANOPY STRUCTURE ACROSS THE TROPICS

To evaluate the canopy characteristics across the different study sites we calculated the median Plant Area Volume Density profile (composed of PAI values for each 1 m vertical interval), using all simulated lidar waveforms for each study site. Additional to the median (50th percentile), we calculated the 10, 30, 70 and 90th percentiles of the PAI values in the same 1 m vertical bins, to provide a representative distribution of the canopy structure.

4.3.4 SPECIES RICHNESS ACROSS THE TROPICS

Aside from calculating total species richness at each study site and at all three plot sizes, we also created species-area relationships, calculating the mean and standard deviation of species richness for plot sizes ranging from 0.01 to 50 ha to assess how the species richness changes by plot size across the study sites across the tropics.

Each of the original field plots is filled with as many subplots as possible at each of these spatial resolutions (0.01, 0.0225, 0.04, 0.09, 0.16, 0.25, 0.36, 0.64, 1.0, 2.25, 4.00, 6.25, 9.00, 12.25, 16.0, 25.0, 50.0 ha) and we assigned each tree to a subplot at each resolution. We then calculated species richness at each plot size for each study site. Which plot sizes are used at each study site depended on the original plot size and the presence/absence of stem maps (Chapter III.1 Field data characteristics). We removed all plots in which more than 20% of the trees were not identified to the genus level. We visualized the mean and standard deviation of species richness for each plot size at each study site to evaluate the differences in species-area curves across the tropics.

4.3.5 STRUCTURE-RICHNESS ANALYSIS

To evaluate the existence of a relationship between vertical canopy structure and tree species richness across the tropics, we developed models at three scales: local, regional and pan-tropical, because many historical and environmental drivers of (tree) species diversity have stronger or weaker relations depending on the scale of observation (Gaston, 2000; Keil and Chase, 2019). Definitions of the scales are presented in the following sections.

Local analysis

The local analysis focused on the structure-richness relationship within large (25 or 50 ha) plots. With the local analysis we use data from field plots that are directly adjacent to evaluate the existence of a relationship between the number of different tree species (S) and the canopy structure expressed by canopy height (RH98) and the vertical canopy profile (PAI at 10 m height intervals). The local analysis was performed on data collected in *bci* (50 ha), *rab* and *rob* (25 ha). The other 50 ha plot (*dan*) was not found suitable for this analysis because the quality of the species identification was not consistent throughout the plot (Table 4; Chapter III.1 Field data characteristics). A Poisson regression model was used, because the species richness information consists of count data, to relate structure and species richness. We used 5-fold cross-validation, extracting 20% of the data at random in each fold as test data. We first performed feature selection on the training data, choosing the model with the lowest BIC score, and constructing the predictive model based on the same training data. We evaluated the model performance with R^2 , RMSE% and bias based on the predictions for the test data. The average and standard deviation of these metrics were recorded for each study site at each resolution.



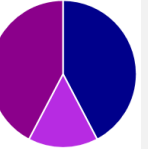

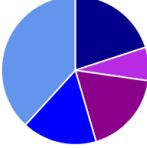
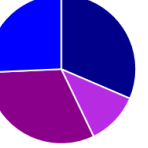
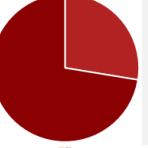
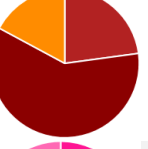
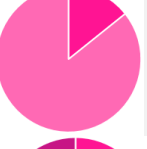

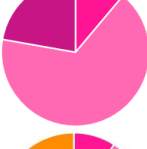
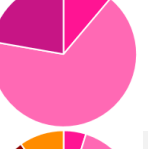
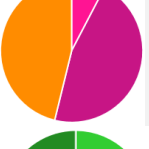

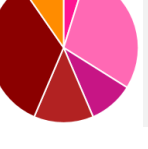
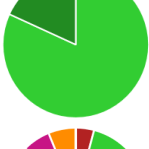
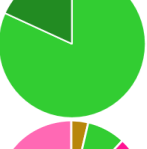
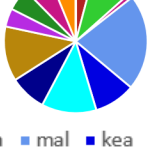

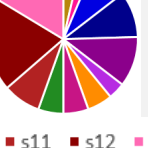
Regional and pan-tropical analysis

The regional analysis focused on the structure-richness relationship based on non-adjacent plots across study sites within the same country or continent. We evaluated

different combinations of study sites at three spatial resolutions (Table 6). To prevent the large plots from dominating the regional and pan-tropical analyses, we thinned their contribution to the regional and pan-tropical datasets. From the 25-ha plots we selected a 1-ha plots at each corner, and from the 50-ha plots we selected all corner 1-ha plots (4 total) and the middle plots along the long side of the plot (6 1-ha plots total). To avoid mixing of local and regional effects, we employed a Monte-Carlo simulation approach in which we drew different samples from the full regional dataset. In each Monte-Carlo run we drew a random sample selecting one plot at the given resolution from each original plot location (especially important at the 0.25 and 0.0625 ha resolution at which up to 16 plots exist at the location of each original 1 ha plot) and applied a cross-validation (80/20) or leave-one-out cross validation (if $n \leq 25$) approach. In the cross-validation we again performed a two-step approach in which we first selected features on the Poisson regression choosing the model with lowest BIC value (using the *bestglm* package in R), and then built the predictive model with the chosen variables. We applied the model to the test data and calculated the model performance statistics for each fold.

The pan-tropical analysis focused on the structure-richness relationship combining the information from all 15 study sites across all tropical regions, in other words, it was a special case of the regional analysis in which data from all sites was included (Table 6). Hence, the same methods were used as for the regional analysis.

Table 6: Datasets used for regional and pan-tropical analysis of the structure-richness relationships. Note that one region may not contain the same number of plots across all resolutions due to limitations in the availability of subplot and stem map information, limiting the use of data from some study sites to only one or two resolutions.

Region	1 ha resolution	0.25 ha resolution	0.0625 ha resolution
Gabon	31 	25 	26 
Africa	61 	55 	35 
Brazil			29 
South America			35 
Costa Rica		21 	21 
Central America		27 	27 
America	13 	33 	62 
South-East Asia	11 	11 	
Pan-tropical	97 	109 	110 

Legend: rob sep cha mal kea mab lop mon rab dan bci tam s11 s12 lsv

4.4 RESULTS

In this section we first show the canopy structure expressed by the vertical foliage profile for the different study sites. Then, we show the relationship between species richness and plot size across all study sites. These results are tied together in the results of the species-richness relationships at the local, regional, and pan-tropical scales.

4.4.1 VERTICAL FOREST STRUCTURE ACROSS THE TROPICS

The vertical canopy structure of forests, in terms of the distribution of plant material across the vertical axis is varies between regions across the tropics (Figure 17). Maximum canopy height in our study sites in the Neotropics and Central Africa average around 40 m and is slightly lower and Australia, while canopy heights in South-East Asia exceed 60 m. Many sites show a strong understory layer and a decrease in plant material through the canopy. Relative to the understory, the canopy layer may have a sharp decline in material (*sep*, Borneo; *dan*, Malaysia) or a steady decline along the vertical axis (*bci*, Panama; *rab*, Gabon; *mal*, DRC; *rob*, Australia). This pattern is exacerbated in degraded forests. In *s11*, *s12* (Brazil) and *mon* (Gabon), the bulk of the plant material exists close to the forest floor at ~5 m height, but remnant trees in some plots may reach to 40 m.

Other sites, especially undisturbed ones, have distinct canopy layers. In *tam* (Peru) and old-growth forests in *lsv* and *cha* (Costa Rica) we see multiple peaks of high density plant material across the vertical height. A multi-peak pattern is also strong in the profiles of *kea* (DRC) and *lop* (Gabon), reflecting the inherent structure of the

forest-savanna mosaic; showing two peaks in canopy material, one close to the ground ~5 m and one higher in the canopy, ~20-30 m. The less-disturbed *mab* (Gabon) forest shows high variability in canopy structure between the different plots, indicated by the wide shaded area in (Figure 17) remnant trees in some plots may reach to 40 m.

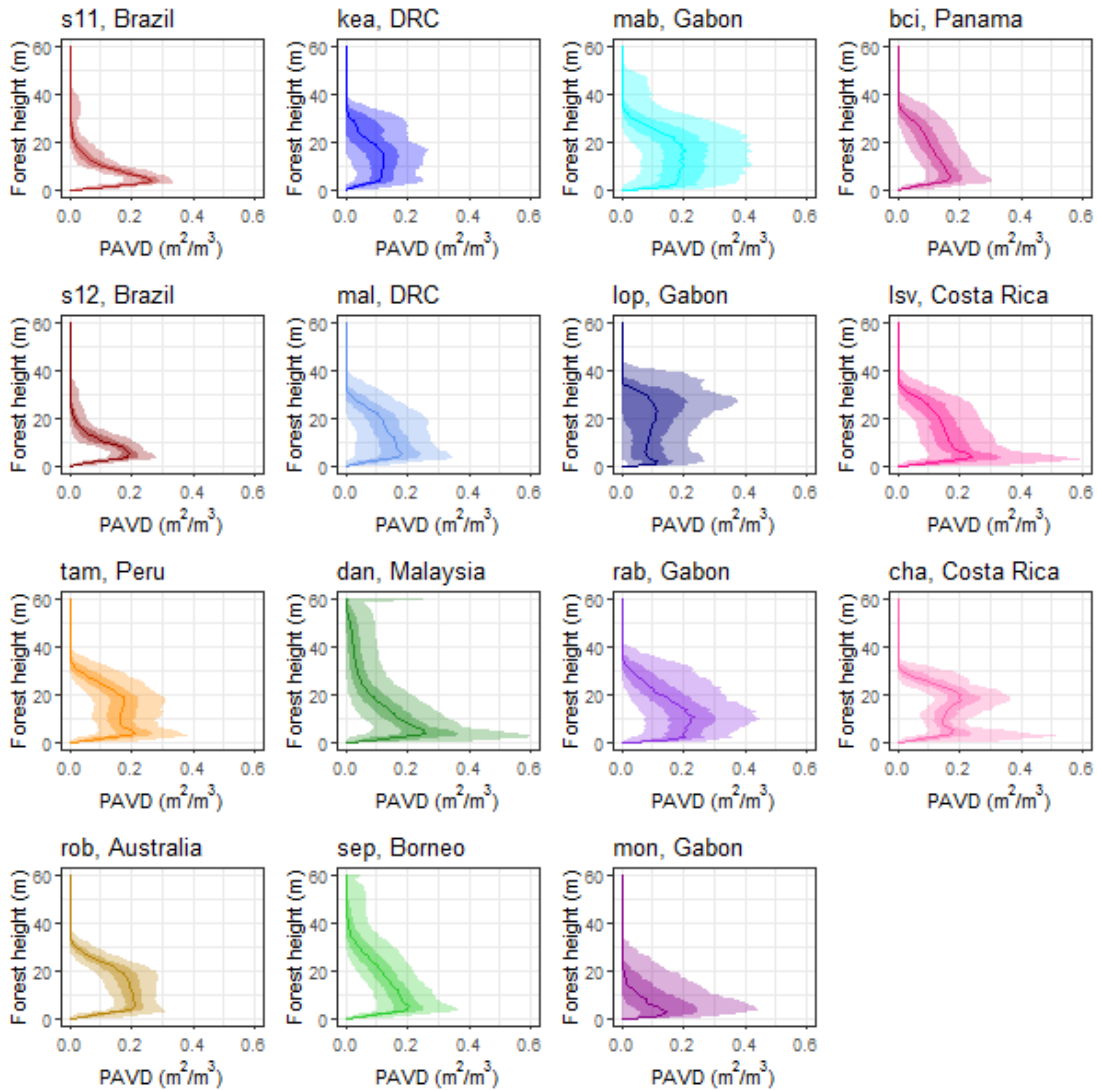


Figure 17: Canopy structure expressed as the Plant Area Volume Density profile (PAVD), expressing the Plant Area Index for each 1 m vertical bin, displayed as the median of all plots within each study site (solid line), the 30th-70th percentile (darker shaded area) and 10th-90th percentile (lighter shaded area).

4.4.2 SPECIES-AREA RELATIONSHIPS

The number of species increases with plot size, but the rate of increase varies across study sites (Table 5). For example, in *rob* (Australia) a 1 ha plot contains 82-117 species and 16-44 species in 0.0625 ha plots. *tam* (Peru) contains between 154-185 species/ha, and 11-35 species in a 0.0625 ha plot, similar to *rob*. This indicates that the adjacent 0.0625 ha plots in *tam* must contain species' compositions more different from each other than adjacent 0.0625 ha plots in *rob* (Australia), i.e. *tam* has a higher local species turnover rate than *rob*. The species-area curves vary in shape across study sites, with the highest total species richness in *tam* and lowest species richness in the African sites (Figure 18). Initial steep curves that decrease in steepness at larger plot sizes indicate a high local diversity but a lower regional diversity (e.g. when the area is increased, the same species are encountered).

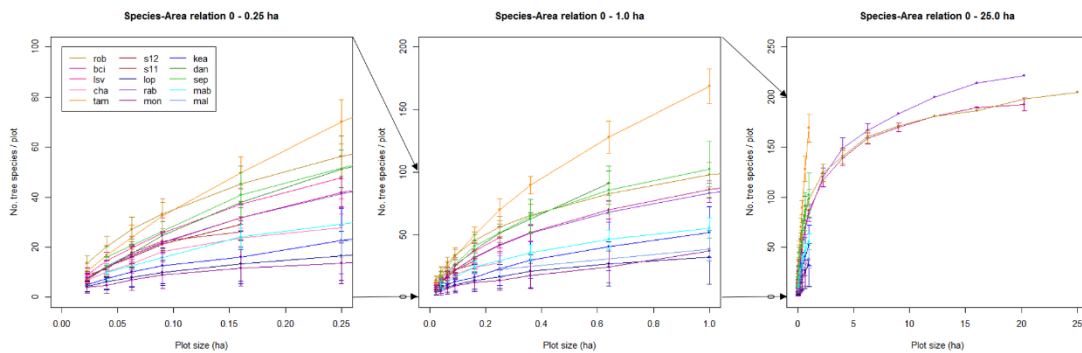


Figure 18: Relationships between tree species richness and area for each study site (note the change in y-axis across panels from left to right).

4.4.3 STRUCTURE-RICHNESS MODELS

Pulling together the information on tree species richness and canopy structure (RH98 and Total PAI), we can see that generally speaking, species richness increases with

increasing canopy height and increasing total Plant Area Index across the tropics
(Figure 19).

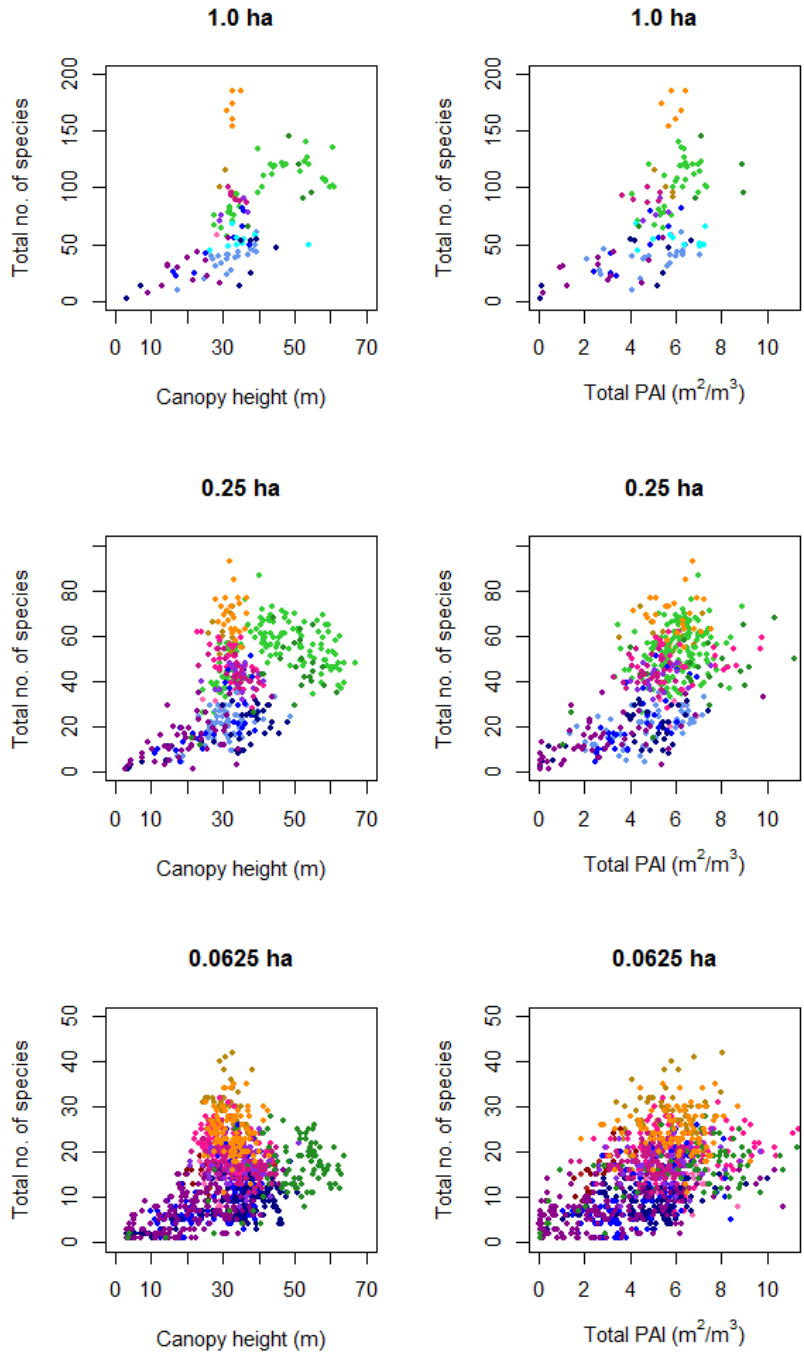


Figure 19: Relation between canopy height (left) and total plant area index across three spatial scales for all study sites across the tropics. Each point represents one plot at the specific resolution.

The cross-validation results of the local models reveal weak structure-richness relationships. Of the three superplots, only the models in *bci* are significant $\geq 95\%$ of the iterations, showing an evident relationship between the predicted and observed values (Figure 20). Even though species richness within all large plots can be predicted with a root mean squared error between 10-30% of the mean, the low RMSE% only indicates that the predictions at the local scale are around the mean species richness, and in *rab* and *rob* they are insensitive to the local variation in tree species richness (see example figures in Chapter III.2 Local model performance details).

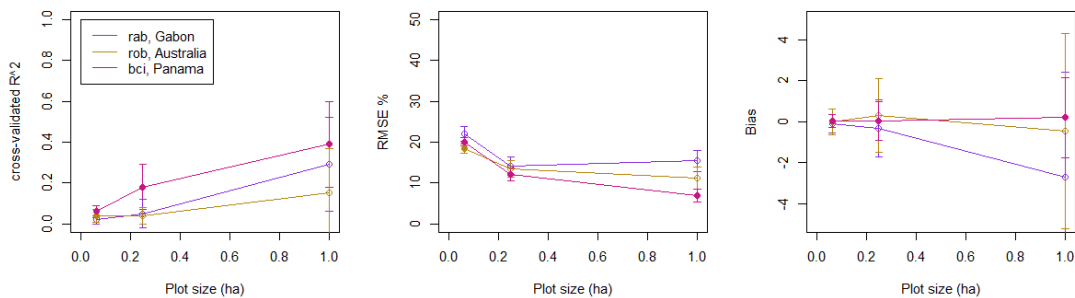


Figure 20: Cross-validated results of local structure-richness models. Open circles indicate less than 95% of the cross-validated models was significant. The models in Panama are significant across all plot sizes, whereas models in Gabon are insignificant at smaller plot size. The RMSE% is low for predictions at each study site, but little relationship exists between the predicted and observed data. All models are unbiased.

Regional structure-richness models generally show better performance at all spatial resolutions (Figure 21) when compared to the local models (Figure 20) in terms of the variance in species richness that can be explained with the canopy structure information (higher R^2 values). However, prediction error (as percentage of the mean species richness) is generally higher, partly due to the larger range in species richness

in these regional datasets. Regions of Gabon, Africa, Brazil and South America show the best model performance whereas regions including the Costa Rica datasets show much poorer performance (regions indicated with *costarica*, *centralamerica* and *america* (0.0625 and 0.25 ha)). Results from an analysis of the ecological distance (Bray-Curtis; Faith et al. (1987)) of the Costa Rica dataset show that, even though the species richness numbers in Costa Rica vary (Table 5), the plots share many of their species, i.e. the composition is similar. In the Gabon, Brazil and South-America datasets the variation in species richness is accompanied by a much larger variation in species composition (Chapter III.3 Ecological and structural distance).

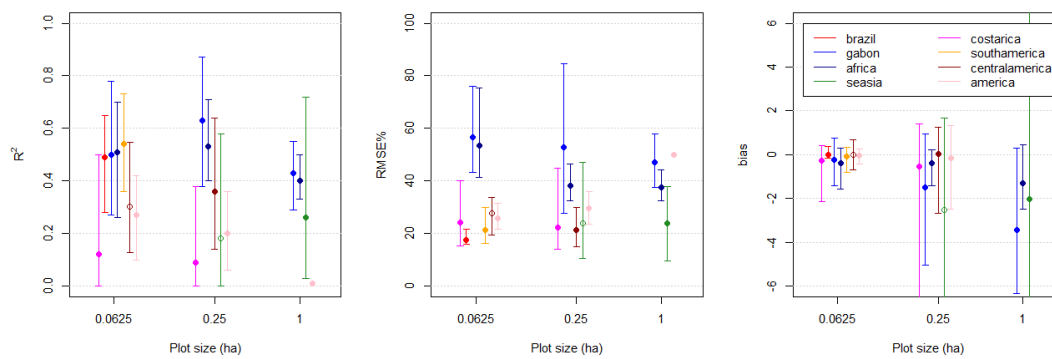


Figure 21: Cross-validated model performance of regional structure-richness models. Error bars indicate the 95% range of values for each metric.

Global structure-richness models show similar performance across all resolutions with mean R^2 ranging between 0.32 and 0.41 and RMSE% between 60 and 39%, respectively for the plot sizes from 1.0 and 0.0625 ha (Figure 22), indicating that around 41% of the variation in tree species richness can be explained using canopy structure metrics alone at the 0.0625 ha resolution at the global scale. Sites with extremely high values of observed species richness are generally predicted poorly (Chapter III.4 Detailed global modeling results).

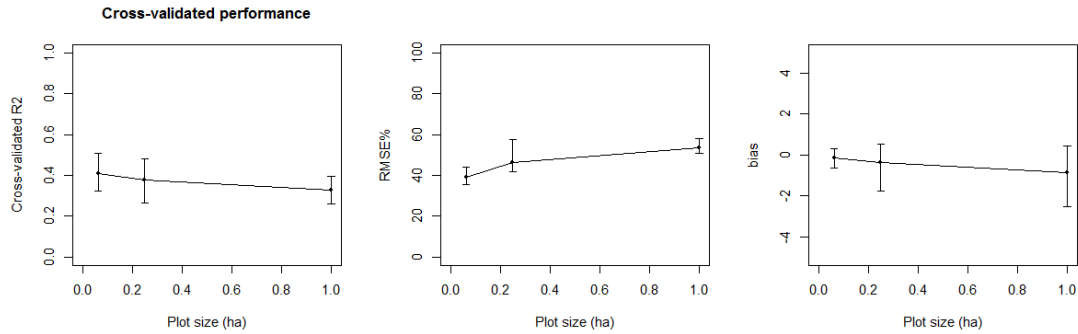


Figure 22: Cross-validated model performance at the global scale in terms of R2, RMSE%, and bias. Error bars indicate the range between which 95% of the performance values of the cross-validated models fall. All models are statistically significant and unbiased.

4.5 DISCUSSION

4.5.1 STRUCTURE-RICHNESS RELATIONSHIP ACROSS SCALES

In this study we explore the relationships between vertical canopy structure and tree species richness at different resolutions across the local, regional and pan-tropical scale. A variation in canopy structure between two plots in any given forest can be attributed to either: (1) the intra-variability of tree structure within the same species group or a (2) difference in structure caused by a difference in species composition. Similarly, a difference in the number of species per plot may either (1) reflect a difference in the composition between plots or (2) reflect the presence of a subset of the species from one plot in another. Based on the results in this study we argue that a significant structure-richness relationship exists between plots where differences in species composition have contributed to the difference in canopy structure and tree species richness based on the following findings:

At the local scale, within one big plot inside of one forest type, the variation in the canopy structure is determined largely by intra-variability within species (these superplots have a similar composition throughout the plot, see Chapter III.1 Field data characteristics). For example, an adult tree of species X may range in height from 20-40 m, so even though a different structure can be measured in two plots of similar composition, this difference in structure is not attributed to a difference in species composition. Furthermore, if a 20-m and 40-m tree of species X exist in the same plot, they could be predicted to be different species ($s=2$) based on the variation in structure, inflating the tree richness estimate. On the other hand, as the area observed increases in size it is more likely that the difference in structure is caused by a difference in composition. Individuals of most tropical forest species are spatially aggregated (Condit, 2000) so the composition between two adjacent plots is more similar than the composition between two plots farther away. This is the case for *bci*, where a 50 ha area was sampled and included in the local analysis, which led to successful prediction of species richness based on structure. Within the 25 ha's sampled at *rab* and *rob*, the variation in composition is smaller and no significant structure-richness relationships were found (Chapter III.3 Ecological and structural distance).

Further increasing in scale, we found that regions consisting of sites exhibiting large variation in species composition among plots show a much stronger structure-richness relationship (Chapter III.3 Ecological and structural distance). However, we note that model performance of the species-richness analysis differed drastically across regions

(either within a country or continent). The forest in *lsv*, Costa Rica, consist of largely similar species composition whereas the difference in composition is much higher in regions where the structure-richness models show better performance (South-America, Brazil, Gabon, Africa), continuing the trend from the local scale models that richness can be better predicted in areas with greater species turnover.

At the global scale we find a significant relationship between canopy structure and tree species richness across all spatial resolutions (plot sizes). At the higher resolution (0.0625 ha) this relationship appears to be stronger than at the lower resolution (1.0 ha), but no significant difference was found. However, the observed difference may be attributed to the lower sensitivity of species richness to rare species at smaller plot sizes. For example, *tam* (Peru) plots have very high species richness at the 1 ha resolution (Table 5), whereas at the 0.0625 ha resolution the species richness range for this site is between 22-50 species, which is still higher than most other sites but much less extreme than at the 1.0 ha plot size. Because the 1.0 ha plot size captures more rare species in each plot, the 1.0 ha global model predictions for *tam* contain outliers that are not present in the 0.0625-ha plot size (Chapter III.4 Detailed global modeling results). Rare species do not contribute much to the canopy structure, thus complicating the relationship between structure and richness at a scale in which they contribute largely to the species richness value.

4.5.2 LIMITATIONS

Here we will discuss the limitations to this study and describe the anticipated effects these limitations may have had on our results.

The quantity of the data is one of the main limitations. This research could be significantly improved by using more coincident lidar and field data to thoroughly evaluate the existence and strength of the structure-richness relationship across all tropical regions. However, we acknowledge the collection of such data is costly and time-consuming. Here, we were able to exploit fifteen independently collected existing datasets (Chapter III.1 Field data characteristics). However, there is quite a data gap, especially in the Amazon basin, the mainland of South-East Asia and New Guinea. Apart from the spatial representation problem, the low number of plots for certain regions attributes largely to the observed variability in model performance. Comparing the pan-tropical models (with $n \geq 97$ leads to much more stable model performance than for some regions with low number of plots). When the training dataset does not fully represent the range of structure in the full dataset, this may lead to highly erroneous precisions for some of the test plots. Especially with the use of Poisson regression such errors are exacerbated by the logarithmic link model, where errors can be exponentially higher when compared to using simple linear regression. On the other hand with linear regression, negative predictions are possible and the risk underestimating tree species richness at the high end of the spectrum is higher. Hence, we chose to use Poisson regression, knowing that it may lead to outlier predictions in certain cases that should be accounted for when operationalizing this method.

This study serves as a first attempt to study the pantropical structure-richness relationship and should be improved and further developed when more data becomes available. Additionally, the characteristics of each dataset differed widely because all data was collected by different people and institutions. We accounted for this as much as possible by using datasets only at reliable plot and subplot resolutions, including only trees ≥ 10 cm DBH and including only plots with a percentage of trees unidentified to the genus level smaller than 20%. Nonetheless, we acknowledge that the quality of the species identification may have affected our models. Although we tried to exclude plots with a high percentage of unidentified trees, it was also important to exploit the available data as much as possible and hence we chose the cut-off percentage of 20%. In the future, a genus-species relationship could be exploited to impute missing data and further increase the confidence of the structure-richness models.

The presence and absence of stem maps and subplots in each study site determined at which spatial resolutions the datasets could, and could not, be used. This resulted in the inclusion of different datasets for each region (Table 6). In this way it makes the inter-comparison of model performance in the same region at different resolutions unreliable because the models were not always built on the same data (plots and study sites), but we weighed this decision to maximize the largest possible variance in the structure-richness models.

4.5.3 FUTURE RESEARCH & APPLICATIONS

Our results provide confidence regarding the existence of regional and global species-richness models that may be exploited for mapping pan-tropical tree species richness. It appears that the most accurate predictions can be achieved at the regional scale when adequate data is available. However, in the absence of such data it may be of more immediate interest to further develop pan-tropical models which can explain ~40% of variation in tree species richness. At the time of writing, GEDI is collecting canopy structure information close to the smallest resolution tested here (0.0625 ha) and thus this data may be well suited for mapping tree species richness across the tropics. GEDI is a sampling mission in which lidar waveforms with 25 m diameter footprints are collected across 8 tracks (600 m between-track spacing, 60 m along-track spacing). GEDI gridded data products will have a 1 km² resolution in which the GEDI data samples are averaged to 1 km² values (Dubayah et al., under review). Our local scale models show that predictions of adjacent 0.0625 ha plots (or in the future, footprints) are on average correct, but they will not represent the local nuances in species richness within forests of uniform composition. We propose that the species richness predictions could potentially be used in a similar way as for gridded GEDI product and provide the average number of species/0.0625 ha within a 1 km² cell, as such information may still be of interest to local land managers. Given the variable species-area relationships it is not easy to translate a species richness at 0.0625 ha resolution to the expected number of tree species in 1 km². Also, we acknowledge that still only roughly 40% of the variance in species richness can be explained using just the lidar structure information and therefore we propose two research avenues of

potential interest: fusion with spectral data and using an environmental framework. Spectral data has previously been shown to predict some of the variance in tree species richness (Foody and Cutler, 2006; Schäfer et al., 2016) and may improve our models and allow for more accurate predictions of tree species richness across the tropics. Especially data from the hyperspectral HISUI (Matsunaga et al., 2013) instrument, that is soon to be launched to the International Space Station, may be highly relevant in such applications. Alternatively, we believe that the inclusion of structural data within previously developed bio-geographical frameworks to predict tree species diversity (Keil and Chase, 2019). Such frameworks could benefit from GEDI lidar data providing information on the occupation of the vertical niche space and likely improving predictions of tree species richness across the tropics.

4.6 CONCLUSION

In this study we evaluated the existence of local, regional and global relationships between vertical canopy structure and tree species richness in the tropics at three spatial resolutions: 1.0, 0.25, and 0.0625 ha. Our results show that canopy structure can explain a limited percentage of variation in tree species richness across the different regions. On a pan-tropical scale between 41% of the variation in tree species richness can be explained with the vertical canopy structure using one single predictive model at a plot size similar to the GEDI footprint resolution. A full set of regional structure-richness models are most likely to aid accurate pan-tropical species richness mapping, but the development of such models is contingent on the availability of sufficient field data across the tropics. Alternatively, canopy structure information from GEDI could be included in existing modeling frameworks, combining spectral, environmental and structural information to provide more accurate tree species richness predictions.

ACKNOWLEDGEMENTS

This work is supported by NASA Headquarters under the NASA Earth and Space Science Fellowship 414 Program – Grant 80NSSC17K0321. We express our sincere gratitude to the following people and institutions for collecting field and lidar data and providing this data for use in this research: Bryan Blair, Michelle Hofton and David Rabine and the LVIS team for collecting LVIS lidar waveforms over *lop*, *mab*, *mon*, and *rab* in Gabon funded through the NASA AfriSAR campaign led by Lola Fatoyinbo. ANPN and AGEOS for logistical support that facilitated field and lidar data collection in Gabon, specifically Flore Koumba Pambo, Josue Edzang Ndong and David Lehmann. We thank ESA for funding field data collection in *lop* through the AfriSAR campaign, Simon Lewis and Nicolas Labrière for organizing and guiding, ANPN and the University of Stirling at the SEGC field station for hosting. Specifically, Carl Ditougou, Pacôme Dimbonda, Arthur Dibambou, Edmond Dimoto, and Napo Milamizokou. We thank NASA for funding field data collection in *mon* through the AfriSAR campaign, John Poulsen for coordinating it and ANPN for hosting it. We thank Nicolas Barbier, Missouri Botanical Garden (Tariq Stevart), Golder Associates, P Ploton, V Droissart, and Y Issembe, for field data collection in *mab*. We thank Shell Gabon and the Smithsonian Conservation Biology Institute for funding, and Alfonso Alonso, Herve Memiaghe, David Kenfack and Pulcherie Bissiengou for guiding field data collection in *rab* (Contribution No. 171 of the Gabon Biodiversity Program). We thank the Sustainable Landscapes Brazil project supported by the Brazilian Agricultural Research Corporation (EMBRAPA), the US Forest Service, and USAID, and the US Department of State for acquiring the *sll* and

s12 field and lidar datasets. We thank the Smithsonian Tropical Research Institute for funding the *bci* field data and the Smithsonian ForestGEO Global Earth Observatory Network for publishing it. The *dan* plot is a core project of the Southeast Asia Rain Forest Research Partnership (SEARRP). We thank SEARRP partners, especially Yayasan Sabah for their support, and HSBC Malaysia and the University of Zurich for funding. We are grateful to the research assistants who are conducting the census, in particular the team leader Alex Karolus, and to Mike Bernados and Bill McDonald for species identifications. We thank Stuart Davies and Shameema Esufali for advice and training. We thank Robin Chazdon for collecting the *cha* field data in Costa Rica. We thank Jim Kellner, David Clark and Deborah Clark for their long term efforts collecting field data in La Selva Costa Rica. We thank Bryan Blair, Michelle Hofton and David Rabine for collecting LVIS data over *slv* and *cha* in La Selva, Costa Rica. We thank Oliver Phillips, Doreen Boyd, Timothy Baker Chris Hopkinson, Ross Hill, Rodolfo Vasquez and Abel Monteagudo for their efforts collecting both field and lidar data in *tam*, Peru. We thank Hans Verbeek, Elizabeth Kearsley and Pascal Boeckx for field data collection in *kea*. We thank the WWF for funding lidar data collection over *mal* and *kea*. We thank Jean-Francois Bastin, Charles de Cannière and Jan Bogaert for collecting field data in the DRC. We thank Reuben Nilus, David Burslem, Lan Qie and Simon Lewis for field data collection in *sep* and David Coomes for the provision of lidar data over *sep* and *dan*.

5. DISCUSSION

This chapter starts with a synthesis of the findings of this dissertation, followed by a summary of the limitation of the current study. Then, it describes the implications of these findings for future use on incoming spaceborne lidar data from the Global Ecosystem Dynamics Investigation (GEDI) mission. A set of final remarks conclude this dissertation.

5.1 SYNTHESIS

The goal of this dissertation was to explore the potential of full-waveform lidar for characterizing tree species diversity in the tropics. This is of particular importance in the light of the recently launched GEDI mission, which will collect billions of lidar waveforms across all tropical forests. If a significant relationship exists between canopy structure and tree species diversity then this data may be used for mapping tree species diversity across the tropics. Such maps may then be used to inform biodiversity conservation and management and to enable a better understanding of the relationship between forest diversity and productivity.

As a first step I verified that full-waveform lidar data provides accurate canopy structure profiles (in terms of Plant Area Index (PAI) binned along the vertical forest axis) by comparing them to those derived from terrestrial lidar data. In chapter 2 I also showed that successional vegetation types with different canopy structure can be distinguished and spatially classified using this information on the canopy structure along the vertical forest axis from full-waveform lidar data. Cuni-Sanchez et al.,

(2016) previously showed that forest types with different tree species composition in Lopé National Park have a distinct canopy structure. This relationship has been explained in a successional setting: the savanna has an open structure with a tall grassy layer and scattered short-statured trees. Following first tree colonization of the savanna, a canopy structure evolves that is denser near the ground and more open at the top. A limited set of canopy species may then overgrow the colonizing species, forming a thick even-height foliage layer with a monodominant tree species composition. After first tree-fall and gap formation in the canopy layer, a changing light regime on the forest floor enables new species to grow. Eventually, the forest develops to a diverse tree species composition with heterogeneous canopy structure, both spatially and vertically (Cuni-Sanchez et al., 2016; White, 2001; White and Abernethy, 1997). The differences in species composition in each successional stage are observable with full-waveform lidar data because they are expressed in the canopy structure, allowing for spatial mapping of the forest types using wall-to-wall lidar data products.

Different numbers of tree species are found across the different forest types in Lopé NP. It has previously been hypothesized that canopy height, as a proxy for niche volume, must to some extent relate to differences in tree species diversity (Gatti et al., 2017). Chapter 3 shows that this height-diversity relationship exists at a regional scale, across multiple study sites in Gabon: canopy height alone explains up to 44% of the variation in tree species diversity. Moreover, I expected that including information along the vertical axis of the forest, expressed by the Plant Area Index

(PAI) profile, would improve these tree species diversity predictions because these metrics contain information on the occupation of the niche volume. Our models explain up to 71% of the variation in tree species diversity across four study sites in Gabon when including this structural information. These results support the hypothesis that information on the occupation of vertical canopy space explains more of the variation in tree species diversity than canopy height alone.

In chapter 4, an additional eleven datasets were used to evaluate the structure-diversity relationship across the tropics. The results show that one single pan-tropical model, using information of the vertical canopy structure and canopy height, can explain up to 40% of the variation in tree species richness across the tropics. Continental scale models for South America and Africa show better results, explaining between 50-65% of the variation in tree species diversity. However, using canopy structure information to explain differences in tree species diversity at a small scale, such as within a 25 or 50 ha field plot, remains a challenge.

In summary, the methods and results presented in this dissertation answer to the main goal of the dissertation by: 1) validating the correct vertical canopy profile extraction from full-waveform lidar data, 2) distinguishing successional vegetation types by their vertical structure, 3) establishing the existence of a significant relationship between tree species diversity and canopy height, 4) supporting the hypothesis that information on vertical canopy structure explains more of the variation in tree species diversity than canopy height alone, 5) establishing significant relationships between canopy structure and tree species diversity in different biogeographical regions and 6)

across the tropics. These results are promising for mapping tree species diversity with GEDI lidar data, but it is important to highlight five constraints before interpreting the implications of the findings for application to spaceborne lidar data.

5.2 LIMITATIONS OF THE CURRENT STUDY

Differences in canopy structure driven by succession (such as in Lopé, Gabon), various degrees of degradation (such as in *mon*, *s11* and *s12*), and differences in species composition within old growth forests (such as in *tam*), can be distinguished with full-waveform lidar data. Such differences in canopy structure relate significantly to the tree species richness. However, in chapter 2 I found that it becomes increasingly difficult to use full-waveform lidar data to distinguish among different species compositions that have a similar canopy structure. Additionally, in chapter 4 I found that the relationship between canopy structure and tree species richness may be weak at a local scale, across plots of similar composition. These two limitations may complicate the use of the methods presented here and should be kept in mind when operationalizing the methods.

The results of this dissertation only include an analysis of forests on solid ground (*terra firme*). It is important to note that the models presented here are not tested for other types of forests that occur in the tropics; such as mangroves or inundated forests. Such forests contain different species compositions with trees specialized to withstand extreme conditions. For example, the upper Gabon estuary is known for its extremely tall trees (62.8 m), but largely uniform species composition (Simard et al., 2019). Hence, a relationship between canopy height and tree species richness

developed in a forest-savanna gradient is unlikely to produce accurate predictions in said mangrove forest.

The nonlinear relationship between number of species and sampled area (species-area relationship) was expected to have an effect on the development and application of the structure-diversity relationship. The results from chapters 3 and 4 show that smaller plots (≤ 0.25 ha) may have a stronger structure-diversity relationship than larger plots (> 0.25 ha). However, not every dataset could be used at each tested resolution, depending on the original field plot size and the absence or presence of stem maps or subplots. Therefore, it is not yet possible to make a conclusive statement on the optimal plot size for the structure-diversity relationship.

Additionally, the species-area relationship is different across all regions in the tropics. This affects both the quality of the structure-diversity models at different resolutions, and the upscaling of predictions made at one resolution to a different resolution. For example, in *tam* the species turnover rate is high, meaning that two adjacent small plots have very different species composition. This leads to high species richness numbers for large plots; at the 1.0 ha resolution, the species richness number in *tam* is far outside the ranges of species richness in all other sites. This leads to poor predictions of tree species diversity for *tam* at the 1.0 ha resolution. On the other hand, at a smaller plot size, the species richness range of the *tam* plots is well within the range of the other study sites, and *tam* predictions are no longer outliers.

Additionally, if a certain number of species is predicted for a 0.25 ha plot, then the species-area curve of that specific region would be needed to establish the number of

species at a larger plot (for example, a 1.0 ha plot). Such species-area curves are not available for all regions in the world, complicating the scaling of plot-level predictions. Thus, both the non-linear species-area relationship and the regional variation of this relationship are important considerations in the effective use of these models.

Between chapter 3 and 4 I explored both linear and Poisson regression models between canopy structure and tree species diversity, expressed as tree species richness (S) and Shannon index (H') in chapter 3 and as S in chapter 4. Both models appear to have their advantages and disadvantages but based on the results of this study I am not able to determine whether one or the other is more appropriate. Linear models can only well-represent if a linear relationship exists between the lidar metrics and tree species diversity metrics. However, this relationship may not be exactly linear, especially when studying areas of extremely high diversity as seen in chapter 4 at the 1.0 ha resolution for the *tam* study site. Therefore, I explored the use of a Poisson regression model using a log link function. This avoids negative diversity predictions (which are possible with the linear model) and allows for the capturing of a logarithmic relationship between explanatory variables and tree species diversity. However, in case the canopy structure in the training sample is not exactly representative of the measured canopy structure in the test dataset, this can lead to outlier predictions of diversity, greatly overestimating the diversity. Such large outlier predictions could be filtered relatively easily, but they do highlight the lack of a fully

representative set of canopy structure and tree species diversity information in certain regions.

5.3 FUTURE DIRECTIONS: LOOKING FORWARD TO DATA FROM GEDI

The GEDI instrument launched to the International Space Station (ISS) in December 2018 to acquire information on the spatial and vertical vegetation structure across all temperate and tropical forests. In a changing world with increasing human and environmental pressures on the natural systems, such information is of immediate importance to enable benchmark mapping of the world's forested ecosystems. This study proposes to extend the use of spaceborne full-waveform lidar data beyond GEDI's original mission goals; to study the use of full-waveform lidar data for mapping tree species diversity in the tropics. At the time this study was conducted, GEDI was mostly in the development and construction phase as it has been collecting science data only since April 2019 (Dubayah et al., under review). The incoming data was being processed at the time of this writing and therefore could not be included within the scope of this research. However, this study does provide valuable insights that can set the expectations for tree species diversity mapping with GEDI lidar data.

GEDI is a sampling mission, which means that the instrument collects data samples instead of a continuous wall-to-wall data product. The GEDI instrument contains three lasers, of which one is split into two beams, the coverage beams, creating four laser beams total. These beams are optically dithered across their tracks creating a total of eight tracks with 600 m across-track spacing at the equator. The along-track

spacing is 60 m, allowing for non-overlapping footprints at a nominal footprint diameter of approximately 25 m (Dubayah et al., under review). The footprint and resulting waveform characteristics are highly similar to those from the Land Vegetation and Ice Sensor (LVIS) used in this study. Hence, results from this study provide excellent guidance for the use of GEDI data for mapping tree species diversity.

Future research directions outlined in this chapter fall into two broad categories: (1) specifically following the line of research proposed in this dissertation using GEDI data alone, and (2) additional approaches, beyond the scope of this study, including other data source for mapping pan-tropical tree species richness.

5.3.1 MAPPING SUCCESSIONAL VEGETATION TYPES

The results from chapter 2 show that highly accurate vertical canopy structure profiles can be derived from LVIS waveforms. Under desirable atmospheric conditions GEDI will provide similarly accurate information on the vertical canopy profile. These profiles are processed and published as Level 2B GEDI data products (Dubayah et al., under review). However, the method to classify a study site in successional vegetation types, as shown in chapter 2, cannot be applied to GEDI directly. There are two main limitations: (1) to allow for spatial mapping of the vegetation types, the method requires a wall-to-wall data product at 25 m resolution, but GEDI only provides circular data samples of ~25 m diameter. Hence, the classification method could be applied to the GEDI waveforms, but the predictions will not create a wall-to-wall map. A classification map could potentially be produced

if GEDI data are fused with radar data to produce a wall-to-wall product of canopy structure, but this would be contingent on the availability of such data and the further development of such fusion methods. (2) Another limitation to applying the classification method is that the vegetation types studied in Lopé National Park are confined to a small area. As such, the developed method cannot be extrapolated to other regions with different forest types. Instead, one may prefer a data driven method in which data mining of GEDI waveforms could be used to indicate groups of similar canopy structure (Moran et al., 2018) to classify forested areas across the pantropical forest region.

5.3.2 ESTIMATING TREE SPECIES DIVERSITY IN GABON

The methods developed in chapter 3, estimating tree species diversity from structural data, can be applied most directly to incoming GEDI data. The 0.25 ha models have already been applied directly to waveforms from the ICESat-1 mission, demonstrating the considerable potential of applying 0.0625 ha models directly to GEDI waveforms. The modeling results show a reasonably strong relationship between canopy structure and tree species richness at the GEDI footprint level across the four study sites in Gabon ($R^2=0.6$, $RMSD = 34\%$; plot size 0.0625 ha). These 0.0625 ha structure-diversity models developed in chapter 3 can be applied directly to incoming GEDI Level 2B data products (Dubayah et al., under review). These GEDI data products will contain the same information as the LVIS data products and thus the methods should be directly applicable to GEDI waveforms to predict tree species diversity (H') and species richness (S) within the four study sites in Gabon from chapter 3. Such footprint-level predictions can be validated with predictions of S and

H' from the wall-to-wall LVIS data at the four study sites at the same spatial resolution. This will provide the most direct validation of GEDI predictions and demonstrate the applicability of these structure-diversity models to GEDI data.

Subsequently, I advise to study the development of a wall-to-wall product from these GEDI predictions. Inference methods can be used to create wall-to-wall data products for the average (and accompanying uncertainty) tree species richness/0.0625 ha samples within larger cells. Similar methods as those proposed to generate the 1 km gridded GEDI data products should be explored here as well (Patterson et al., 2019; Saarela et al., 2018). With these methods it should be possible to predict the 'average tree species diversity per 0.0625 ha within a 1 km cell'. Alternatively, fusion with e.g. Tandem-X radar data should be explored further to generate wall-to-wall canopy structure maps, on which the structure-diversity models can be applied directly to map tree species diversity at a higher spatial resolution.

5.3.3 ESTIMATING PAN-TROPICAL TREE SPECIES RICHNESS

It is of utmost importance to expand the calibration dataset used to develop the regional and pan-tropical structure-richness models. Currently, I used fifteen datasets across the tropics and it is important to note that the data included in the presented models is not necessarily representative of all *terra firme* forests across the pan-tropics. Contingent on the availability of more data, structure-richness models should be constructed for tropical forests grouped by biogeographical history as defined in Corlett and Primack (2011), leading to a set of reliable models to be used for predicting tree species richness across the tropics. Any such predictions based on real

GEDI data should undergo deep scrutiny and resulting data products should be thoroughly validated before being used to advise ecosystem and biodiversity management. Given the data-dependency for the development of these models, I also propose alternative methods to combine GEDI canopy structure data with other datasets to further enable pan-tropical tree species diversity mapping.

5.3.4 THE FUTURE IS FUSION

Data fusion between different sets of remote sensing data with information from GEDI may resolve a number of the limitations mentioned in this dissertation.

Resolving GEDI's spatial sampling limitation

Fusion of GEDI lidar data with Tandem-X may allow for wall-to-wall mapping of canopy structure at a high spatial resolution such as was performed for canopy height in chapter 3. Currently such methods are further developed to fully exploit the information on vertical canopy structure captured by TanDEM-X by fusion with GEDI structural data (Lee et al., 2018; Qi et al., 2019). This may lead to the development of a fused GEDI-TanDEM-X product representing not only canopy height but also vertical canopy structure at high (~25 m) spatial resolution across the tropics. The structure-diversity models mentioned here at the 0.0625 ha resolution could be applied to such a fused data product and result in mapping tree species richness.

Improving model performance by including additional information

Alternatively, data fusion with other remotely sensed information on different forest aspects could improve the models. For example, combining structural data from GEDI with multi- or hyper-spectral imagery may explain more of the variation in tree species diversity than canopy structure information alone. Spectral information by itself has shown to be useful for mapping variations in tree species diversity (Laurin et al., 2014) and it may be possible that such data explains a different portion of the variation in tree species diversity than the structural data does. Hence, fusion may improve the diversity estimates. Multi-spectral data fusion could be performed directly with information from Landsat images or even Landsat time series. The spectral information from Landsat may provide additional information on tree species, and time series information could add further information to this, given that seasonal differences between different forests could be captured and contribute useful information to the tree species diversity model. Upcoming hyperspectral data from the Hyperspectral Imager Suite (HISUI) mission has been identified by Stavros et al. (2017) as useful for deriving information on species, combining this with information on canopy structure may explain even more of the tree species diversity and build a path forward to pantropical tree species diversity mapping using spaceborne data from instruments on the ISS.

Last, diversity of tree species is a complex interaction of many different mechanisms and canopy structure explains only a limited percentage of the variation in tree species diversity. Hence, tree species diversity has previously been mapped across

different grains using a set of environmental variables (Keil and Chase, 2019). I believe that such a method could also benefit from the inclusion of canopy structural data. Once GEDI has completed its nominal two year mission, information on both canopy height and internal canopy structure will be available at a 1 km spatial resolution between 52° degrees N & S (Dubayah et al., under review). The structural data could be incorporated in existing frameworks developed using environmental data (Keil and Chase, 2019). As structure data provides information on the occupation of the vertical niche space across the tropics, the inclusion of canopy structure information may further explain variations in tree species diversity across the tropics and may thus be used to improve pan-tropical tree species diversity mapping.

5.3.5 IMPLICATIONS FOR ECOSYSTEM MANAGEMENT

A list of Essential Biodiversity Variables was generated to summarize the data needs for effective conservation and ecosystem management (Pereira et al., 2013). The development of information on the spatial distribution of taxonomic species diversity is among the variables of interest. Tree species diversity/richness maps across the tropics can be considered of high importance for biodiversity conservation. Such high-resolution maps providing consistent information across the entire tropics could: aid the prioritization of conservation areas, be used to estimate tree species loss after disturbance and may further the understanding of the interactions between biotic and abiotic environments over larger scales. Currently such maps do not exist and a call for bolder science in conservation strategies had been made to the community (Noss et al., 2012; Watson et al., 2016). This dissertation proposes such a bold idea, taking data from an instrument primarily developed for a different purpose (aboveground

biomass mapping) and exploiting the information for mapping tree species diversity. However, before the methods presented in this dissertation will be fully operational and able to provide data products that can be used directly by ecosystem managers, additional research as outlined in the previous chapters will need to be performed.

5.4 FINAL REMARKS

To conclude, this dissertation shows the potential of using full-waveform lidar data to advance pan-tropical tree species diversity mapping. This is the first study of its kind and this line of research is by no means complete enough to allow for direct mapping of tree species diversity across the entire tropics with incoming GEDI data. However, a significant relationship between vertical canopy structure and tree species diversity exists, both in biographical regions and across the entire tropics. Continuing this line of research by including additional calibration and validation data, incoming GEDI data products, additional environmental variables, and/or other remotely sensed data, is therefore highly recommended to ultimately enable high-resolution mapping of tree species diversity and to facilitate effective conservation management of our valuable tropical ecosystems.

I. SUPPLEMENTARY MATERIAL FOR CHAPTER 2

I.1 TLS-LVIS COMPARISON MULTIPLE SCALES

Comparison of the cumulative PAI profiles between TLS and LVIS with a vertical resolution of 3 m (Figure 23) and 1 m (Figure 24). Cumulative PAI profiles are provided at the plot level for eight field plots (20x40 m) collected in the four successional forest types (2 plots in each). Cumulative PAI profiles from LVIS and TLS show good correspondence in all plots, but with a bias in the higher strata (approx. > 35 m).

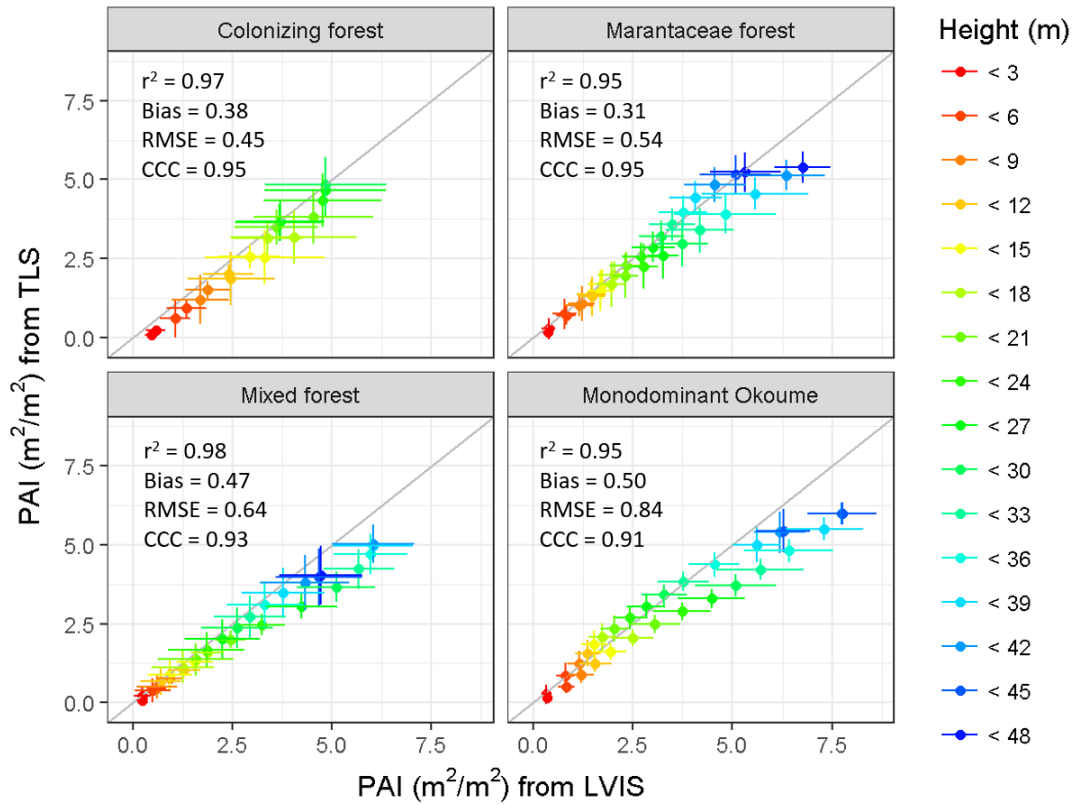


Figure 23: Cumulative PAI profiles from TLS (y-axis) and LVIS (x-axis) at 3 m vertical interval. Overall correlation results of all 8 plots: $r^2 = 0.95$, RMSE = 0.65, bias = 0.41, CCC=0.94.

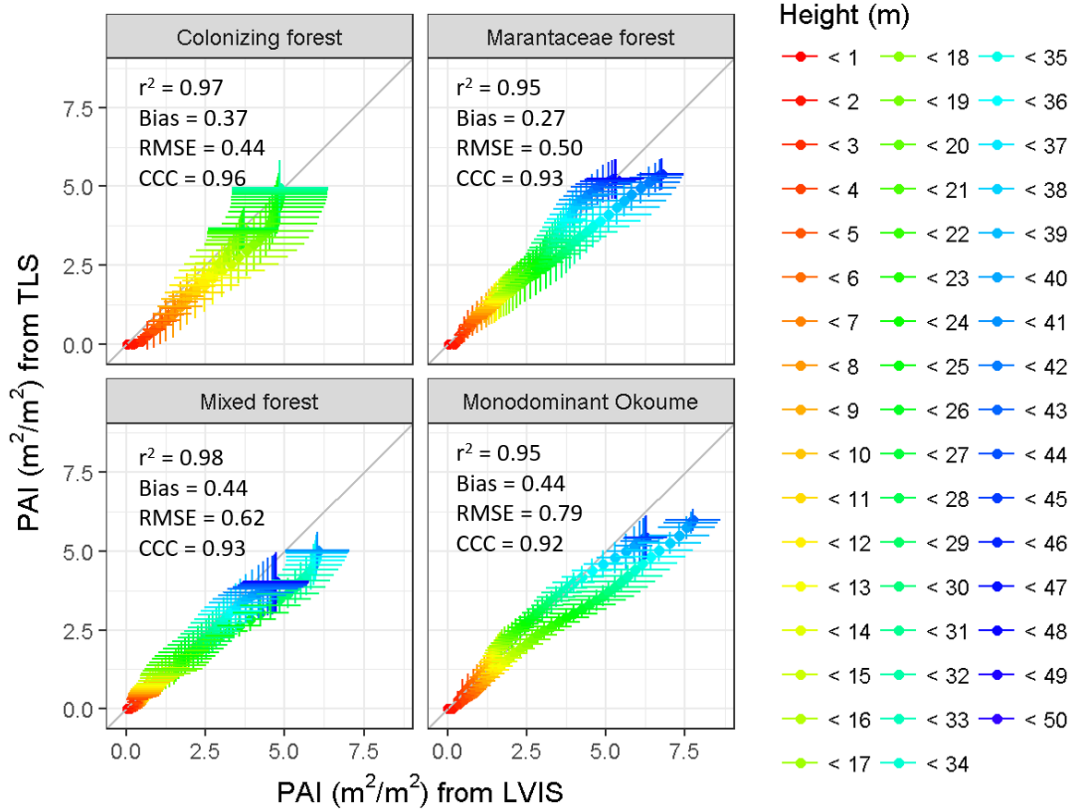


Figure 24: Cumulative PAI profiles from TLS (y-axis) and LVIS (x-axis) at 1 m vertical interval. At the 1 m interval profiles for the individual plots can be distinguished from the figure. Overall correlation results of all eight plots: $r^2 = 0.95$, RMSE = 0.61, bias = 0.38, CCC = 0.94.

I.2 VEGETATION STRUCTURE METRICS

Vegetation metrics showing aspects of the vertical structure of each vegetation type with vertical interval of the PAI profile of 5 m (instead of 10 m, Figure 4). The 5 m interval provides more detail on the specific structure of each vegetation type compared to the 10 m vertical interval (compare Figure 4 with Figure 25). Standard deviation of the PAI profile vegetation metrics are highest for Colonizing forest and Mixed forest in the lower strata (< 10 m) (Figure 26). This indicates that on a pixel-to-pixel scale, the vegetation structure of the lower strata is highly variable for these forest types. Standard deviation for canopy height, cover fraction and total LAI is lowest in Monodominant forest because of the homogenous nature of the vegetation structure in this forest type. Averaging the vegetation metrics for a 3x3 window reduces the variation of the structure for most vegetation types apart from Colonizing forest. This shows that Colonizing forest can occur with very different vegetation structure on a small scale (Figure 27) as the 3x3 window may not consist of solely Colonizing forest pixels since Colonizing forest patches are generally small and extend only along the forest-edge. The coefficient of variation (CV) does not provide any new information on top of the pixel- and neighbor-based mean and standard deviation of each metric (Figure 28).

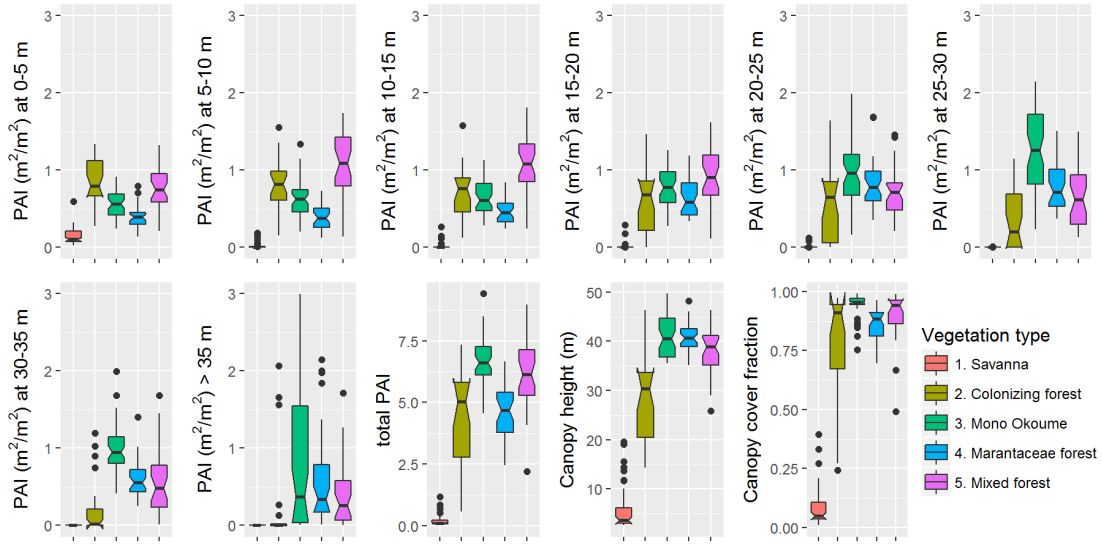


Figure 25: Eleven metrics (y-axis) describing characteristics of vegetation structure of the five vegetation types (colored), ordered by successional stage. Each boxplot is composed of the pixel values within all field plots covered by one specific vegetation type. PAI (m^2/m^2) profile is provided at a 5 m vertical interval.

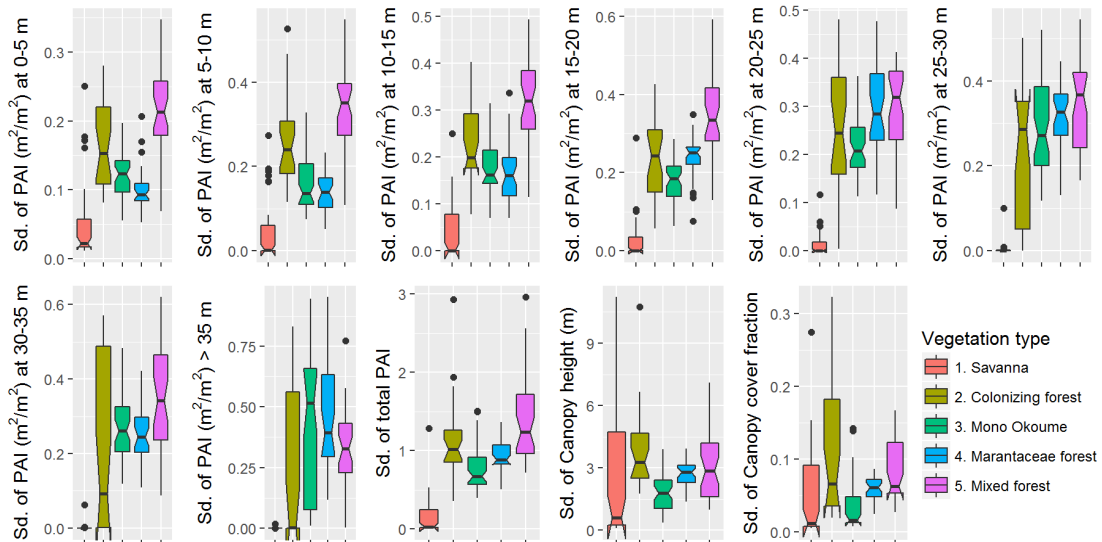


Figure 26: Standard deviation calculated from a 3x3 window around each pixel, providing information on the variation of each vegetation metric for each vegetation type (ordered by successional stage).

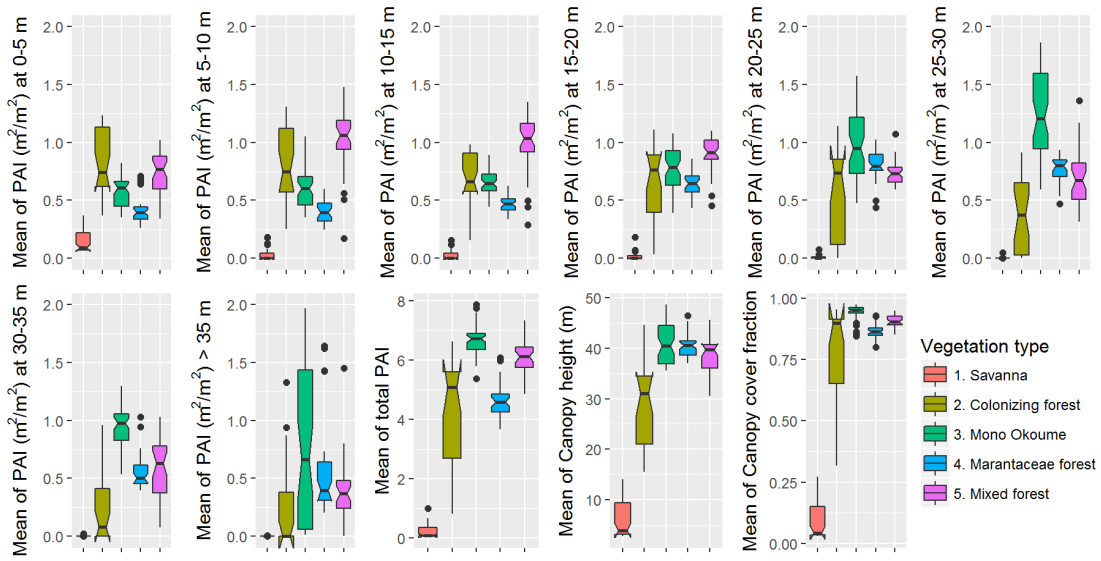


Figure 27: Mean of each structural metric calculated from 3x3 window around each pixel. Variation in metrics for each class is lower than when using single-pixel values.

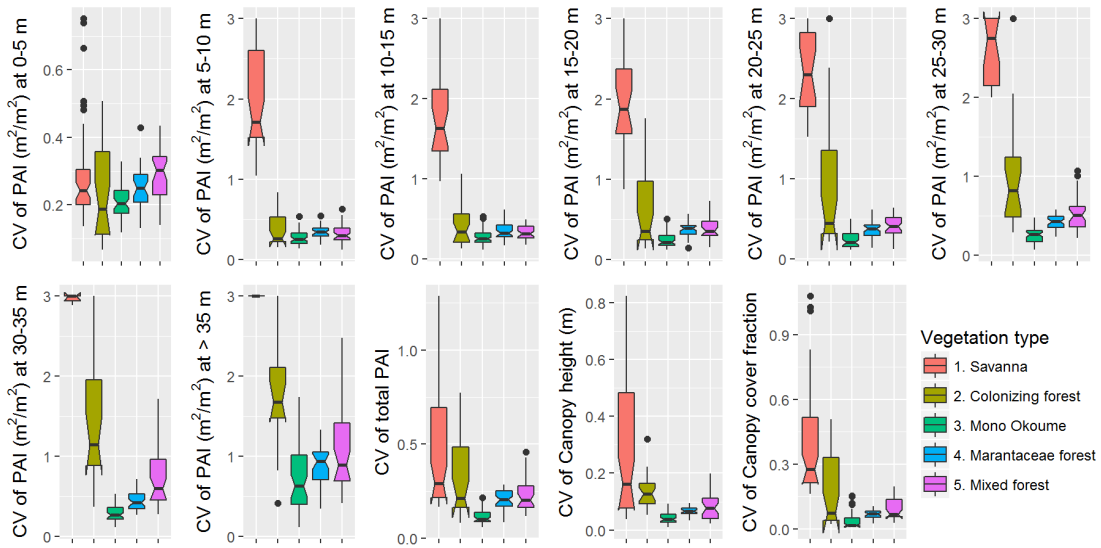


Figure 28: Coefficient of variation of each metric calculated from 3x3 window around each pixel.

I.3 ALTERNATIVE CLASSIFICATIONS

Three Random Forest models were explored: using only pixel-based metrics (Figure 6) using all vegetation metrics (including 3x3 pixel average, standard deviation and coefficient of variation, Figure 29a) and a hierarchical model using two-stage classification with different sets of metrics to classify savanna and Colonizing forest (stage 1 pixel-based) and then the three inner forest types (stage 2 neighbor-based, Figure S29b). The extent of Colonizing forest was largely overestimated using the neighbor-based model with all metrics (Figure 29a). Colonizing forest is known to only exist in thin bands along the savanna/forest boundary (White, 2001; White and Abernethy, 1997), but because of its high variation in structure and the high variation in average structure at a 3x3 pixel window along the forest edge, more cells were classified as Colonizing forest than what should realistically be the case (Figure 29a). However, using the average and standard deviation in the overall model helps classifying the vegetation types of the continuous forest in a more coherent way, showing continuous patches of specific forest types (Figure 29a) instead of isolated pixels (Figure 6). A two-stage classification approach may therefore be considered to improve the overall classification accuracy by representing the Colonizing forest in thinner bands, but dividing the inner forest into larger patches of similar type, providing a more realistic distribution of the vegetation types (Figure 29b). However, this classification approach is only shown here as a demonstration product because the resulting map would require further validation using field data that are yet to be collected.

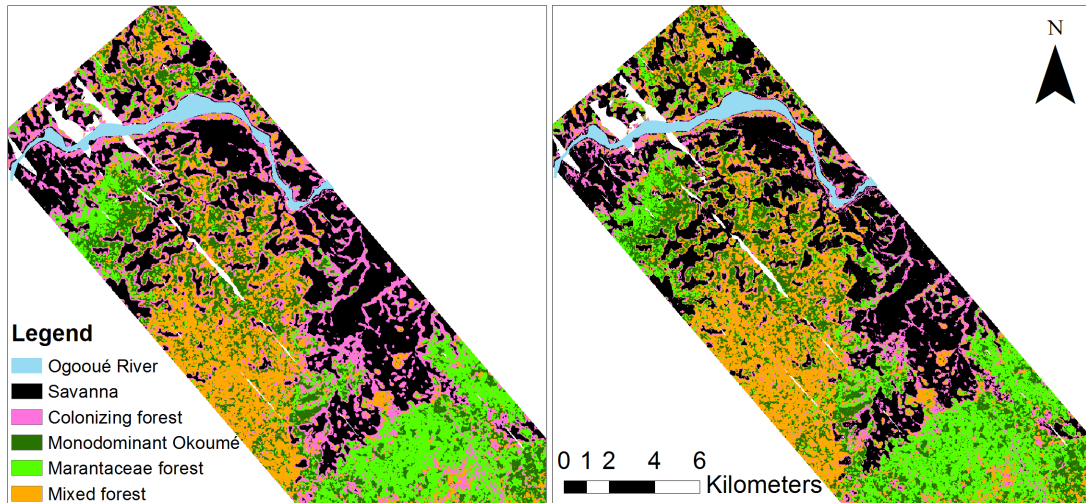


Figure 29 Left: (a) Classification of vegetation types with random forest model build on all vegetation metrics. Right: (b) Classification of vegetation types using two-stage classification. Savanna, Colonizing forest and ‘other forest’ were classified using pixel-based vegetation metrics, after which a random forest model with all metrics was used to re-classify ‘other forest’ to Monodominant Okoumé, Marantaceae forest and Mixed forest.

II. SUPPLEMENTARY MATERIAL FOR CHAPTER 3

II.1 DATASET DETAILS

II.1.1 STUDY SITE DESCRIPTION

Gabon was chosen as case-study location because of the large availability of field and remote sensing datasets collected during the AfriSAR campaign, a joint airborne and field campaign between the European Space Agency (ESA) and the National Aeronautics and Space Administration (NASA), with collaboration from the Gabonese Studies and Space Observations Agency (AGEOS), the German Aerospace Center (DLR) and the French Aeronautics Space and Defense Research Lab (ONERA), which aimed to collect data for the calibration and evaluation of forest structure and biomass retrieval algorithms of future satellite mission (Fatoyinbo et al., 2017). The four study sites in Gabon are referred to as Lopé, Mondah, Mabounié and Rabi (Figure 8). All are within the general classification of tropical *terra firme* broadleaf forest, but have different species compositions and assemblies, disturbance history, and management regimes. The plots in this analysis are distributed between savanna, successional, degraded and old-growth tropical forest and are situated in climate zones with different annual precipitation and temperature (Table 2).

Lopé National Park is located in central Gabon (Figure 8). The study site is mainly comprised of five successional vegetation types: Savanna (Sa), Colonizing forest (Cf), Monodominant Okoumé forest (MOf), Young Marantaceae forest (YMf) and Mixed Marantaceae forest (MMf) (White, 2001). The vegetation pattern is thought to

originate from the last ice age and today fire is the dominant factor maintaining this successional pattern at the forest edges (White and Abernethy, 1997). Twelve stem mapped field plots (nine 1 ha and three 0.5 ha) were established in 2016 as part of the AfriSAR campaign (Labrière et al., 2018) and the data are available through ForestPlots.net (Labrière et al., 2018; Lopez-Gonzalez et al., 2011, 2009). The plots are positioned to cover the five vegetation types (3 Sa, 3 Cf, 3 MOF, 2 YMF, and 1 MMF) and all are coincident with the LVIS lidar data. The Mondah study site is located on the Libreville peninsula, 20 km north of the densest urban area in Gabon (Figure 8) and characterized by coastal Guineo-Congolian forest dominated by *Aucoumea klaineana*. Mondah has been exploited over the last centuries for different resources, such as logging, bush meat, charcoal production and collection of non-timber forest products for traditional medicine. Deforestation to enable peri-urban land use, has increased largely over the last decades, leading to a 40% reduction of the forested area on the peninsula in the last 80 years (Walters et al., 2016). The study site comprises fifteen 1 ha stem mapped plots, thirteen of these are coincident with the LVIS lidar data and used in this study. The Mondah data were also collected during the AfriSAR campaign in 2016 and are publicly available through the NASA Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) (Fatoyinbo et al., 2018). The forest at the third site, Mabounié, is less disturbed than the Mondah forest as only part of the site underwent selective logging in the last decades (Fatoyinbo et al., 2018). Twelve 1 ha field plots were established in Mabounié in 2012 (Fatoyinbo et al., 2018), of which 10 are coincident with LVIS lidar data and suitable for this study. Rabi, the fourth study area, was logged for

Okoumé roughly 30 years ago and is now an active oil and logging concession (Memiaghe et al., 2016). A stem mapped 25 ha permanent forest plot was established between 2010 and 2012. The data were collected by the Smithsonian Conservation Biology Institute, National Zoological Park, and the Forest Global Earth Observatory (ForestGEO), and are available on request through the ForestGEO website². The plot was established to study long-term forest dynamics following the ForestGEO standard method including all trees with a stem diameter (DBH) ≥ 1 cm (Anderson-Teixeira et al., 2015).

II.1.2 FIELD DATA PROCESSING DETAILS

Field datasets for Lopé, Mondah and Rabi are comprised of stem mapped data for which information on tree species, DBH and tree location are available. In Rabi and Mondah trees with DBH ≥ 1 cm were included. In Lopé only trees with DBH ≥ 5 cm. In Mabounié DBH and species were recorded for each tree with DBH ≥ 10 cm, and each tree was assigned to a 20x20 m subplot within the 1 ha plot. For consistency across all study sites only trees with a DBH ≥ 10 cm were included in this study. For each plot, the original location was obtained in the field using (mostly handheld) GPS instruments. In this study, however, we used optimized plot locations as published by (Labrière et al., 2018). They refined plot geolocation by manually translating a stem map of largest trees with respect to a 1m-resolution Canopy Height Model (CHM; derived from discrete-return airborne lidar acquisitions) until a good agreement was found. As a consequence, this optimization should reduce the error component in our

² <https://forestgeo.si.edu/sites/africa/rabi>

structure-diversity models attributed to geolocation error. Most trees were identified to the species level, but in each site, there was a percentage of trees with either unknown species and/or genus. Trees of the same genus but unknown species were all conservatively assigned to the same species (e.g. *Anthocleista sp.*), only in Rabi unidentified species within genera were distinguished from each other in the field, (e.g. *Xylopia sp.1* and *Xylopia sp.2*). In Rabi 91.7% of the trees was identified to species-level, 98.4% to genus-level and 1.6% remained unidentified. In Lopé this was respectively 99.9%, 99.9% and 0.064%, in Mondah 44.4%, 85.1% and 14.9%, and in Mabounié 93.0%, 99.7% and 0.3%. Three of the Mondah plots were then excluded from the analysis, as respectively 64%, 42% and 32% of the trees in these plots were unidentified. This changed the identification of Mondah trees to 45.6%, 97.6% and 2.4% respectively and reduced the Mondah dataset to from thirteen to ten plots. The availability of stem maps enabled testing of the structure-diversity relation at different resolutions. This was necessary because species diversity and plot size are not linearly related (species-area curve, described by (MacArthur and Wilson, 1967) and no optimal resolution has been identified for the structure-diversity relation. Smaller plots were created by subdividing each original plot into smaller squares or rectangles to create 5 spatial resolutions: 1, 0.5, 0.25, 0.0625 and 0.04 ha (Figure 9(b)). From the 25 ha Rabi plot, the thirteen non-adjacent hectares were subsampled and considered separate plots for the analysis. The subsampling avoided this study site from dominating the models, as the Rabi site consisted of a total sampled area twice as large as the other sites (Figure 9(a)). Mabounié data was only used in the analysis at the 1 ha and 0.04 ha resolution, due to the absence of stem maps. At each resolution

only plots with at least one identified tree were included in the analysis. This resulted in respectively 41, 64, 128, 481 and 935 plots for the resolutions 1.0 ha to 0.04 ha.

Tree species diversity was then quantified for each of the plots at all resolutions using two variables: the Shannon diversity (H') and tree species richness (S) expressed as the total number of tree species per area (no. of species/area) (Morris et al., 2014).

The two metrics we used to describe species diversity (H' and S) are commonly found in the literature, but other metrics, such as the Simpson indices for species diversity and species evenness and Fisher's alpha are also prevalent (Morris et al., 2014; Wilsey and Potvin, 2000). We chose species richness as it is an important metric in the Aichi biodiversity targets³ and easy to interpret and can be directly used to advice biodiversity conservation management. We chose the Shannon's diversity index complementary to species richness as it takes into account the abundance of each species. In our study we found that our predictive models for Shannon diversity generally demonstrated a slightly better performance (Figure 11; Chapter II.2 Model performance). This is to be expected because canopy structure is not likely to be largely affected by the presence or absence of one individual of a species (rare species occurrence), whereas the species richness number will be directly influenced by the occurrence of such rare species.

³ <https://www.cbd.int>

II.2 MODEL PERFORMANCE

Results of model runs for Species richness and Shannon diversity with different sets of explanatory variables (set 1-4, Table 3). Results of linear models using 4 different sets of variables at 5 spatial scales predicting Shannon diversity H' (Table 7) and Species richness (Table 8). Res indicates resolution (plot size) in m^2 . RMSD provided as a fraction, value to be multiplied by 100 to show RMSD%. Bias refers to intercept of linear relation between measured vs. cross-validated predicted richness/diversity, consis refers to the slope of this linear relationship. Bias and consis are NA if values did not differ significantly from 0 and 1 respectively, indicating no significant bias or inconsistency between measured and predicted variables. Last two column provide the R^2 confidence interval calculated using the CI.Rsq function in the Psychometric library in R. Models using the cumulative PAI profile (set 4) generally show a similar performance compared to models using the PAVD profile (set 3). However, we chose to use set 3 for our applications because of the following two reasons: 1) when an 'erroneous' number occurs anywhere in the PAVD profile (for example this will carry over to all subsequent bins of the cumulative PAI profile, and may result in erroneous diversity predictions. 2) when the PAI value of the last height bin in which vegetation was detected was smaller than the order of 1/1000 this would have resulted in a PAI of 0 for the respective bin and also 0 in the cumulative PAI profile. In the PAVD profile a value of 0 has a much smaller impact (difference between, e.g., 0.001 or 0), whereas in the cumulative profile that difference would be much larger, say 4.661 vs. 0, leading to erroneous predictions of diversity.

Table 7: Results for models predicting Shannon diversity.

MODEL INFORMATION				MODEL PERFORMANCE				R ² CONF. INTERVAL	
SET #	SET NAME	VARS	RES	R ²	RMSD	BIAS	CONSIS	R ² MIN	R ² MAX
1	h10000	height	10000	0.28	0.22	NA	NA	0.07	0.50
	h5000	height	5000	0.36	0.27	NA	NA	0.18	0.54
	h2500	height	2500	0.43	0.28	NA	NA	0.30	0.55
	h625	height	625	0.40	0.31	NA	NA	0.33	0.47
	h400	height	400	0.23	0.32	NA	NA	0.18	0.27
2	hsd10000	height and sd height	10000	0.40	0.21	NA	NA	0.19	0.61
	hsd5000	height and sd height	5000	0.51	0.24	NA	NA	0.34	0.67
	hsd2500	height and sd height	2500	0.58	0.24	NA	NA	0.47	0.69
	hsd625	height and sd height	625	0.45	0.30	NA	NA	0.38	0.51
	hsd400	height and sd height	400	0.24	0.31	NA	NA	0.19	0.29
3	lsum10000	PAVD and height	10000	0.50	0.19	NA	NA	0.31	0.70
	lsum5000	PAVD and height	5000	0.63	0.21	NA	NA	0.49	0.77
	lsum2500	PAVD and height	2500	0.71	0.20	NA	NA	0.63	0.79
	lsum625	PAVD and height	625	0.65	0.24	NA	NA	0.60	0.70
	lsum400	PAVD and height	400	0.46	0.27	NA	NA	0.41	0.50
4	clsum10000	cum. PAI and height	10000	0.47	0.20	NA	NA	-0.04	0.32
	clsum5000	cum. PAI and height	5000	0.63	0.21	NA	NA	0.50	0.77
	clsum2500	cum. PAI and height	2500	0.61	0.23	NA	NA	0.51	0.72
	clsum625	cum. PAI and height	625	0.59	0.26	NA	NA	0.54	0.65
	clsum400	cum. PAI and height	400	0.45	0.27	NA	NA	0.40	0.50

Table 8: Results for models predicting species richness

<i>MODEL INFORMATION</i>				<i>MODEL PERFORMANCE</i>				<i>R² CONF. INTERVAL</i>	
<i>SET #</i>	<i>SET NAME</i>	<i>VARS</i>	<i>RES</i>	<i>R²</i>	<i>RMSD</i>	<i>BIAS</i>	<i>CONSIS</i>	<i>R² MIN</i>	<i>R² MAX</i>
1	<i>h10000</i>	height	10000	0.35	0.4	NA	NA	0.14	0.57
	<i>h5000</i>	height	5000	0.41	0.44	NA	NA	0.23	0.58
	<i>h2500</i>	height	2500	0.44	0.44	NA	NA	0.31	0.56
	<i>h625</i>	height	625	0.31	0.44	NA	NA	0.24	0.38
	<i>h400</i>	height	400	0.15	0.43	NA	NA	0.11	0.20
2	<i>hsd10000</i>	height and sd height	10000	0.40	0.39	NA	NA	0.18	0.61
	<i>hsd5000</i>	height and sd height	5000	0.46	0.42	NA	NA	0.29	0.63
	<i>hsd2500</i>	height and sd height	2500	0.52	0.40	NA	NA	0.40	0.64
	<i>hsd625</i>	height and sd height	625	0.34	0.43	NA	NA	0.27	0.41
	<i>hsd400</i>	height and sd height	400	0.16	0.42	1.26	0.86	0.11	0.20
3	<i>lsum10000</i>	PAVD and height	10000	0.43	0.41	15.88	0.69	0.22	0.64
	<i>lsum5000</i>	PAVD and height	5000	0.66	0.34	NA	NA	0.53	0.78
	<i>lsum2500</i>	PAVD and height	2500	0.71	0.31	NA	NA	0.63	0.79
	<i>lsum625</i>	PAVD and height	625	0.60	0.34	NA	NA	0.54	0.65
	<i>lsum400</i>	PAVD and height	400	0.38	0.37	NA	0.93	0.33	0.43
4	<i>clsum10000</i>	cum. PAI and height	10000	0.46	0.38	NA	NA	-0.07	0.23
	<i>clsum5000</i>	cum. PAI and height	5000	0.68	0.33	NA	NA	0.56	0.80
	<i>clsum2500</i>	cum. PAI and height	2500	0.62	0.36	NA	NA	0.52	0.72
	<i>clsum625</i>	cum. PAI and height	625	0.53	0.36	NA	NA	0.47	0.59
	<i>clsum400</i>	cum. PAI and height	400	0.37	0.37	NA	0.94	0.32	0.42

II.3 COEFFICIENT OF VARIATION

We calculated the Coefficient of Variation (CV, $CV = \text{standard deviation}/\text{mean}$) for species richness (S) and two summary canopy structure metrics (Canopy height and Total PAI). Figure 30 shows how the CV change across the sampled resolutions. The highest CV can be found at the 0.25 ha resolution, indicating that at the 0.25 ha resolution the variables are showing the highest dispersion around the mean value. Note that for species richness the CV is the same at both 0.25 ha and 0.5 ha resolution, but lower at all other resolutions.

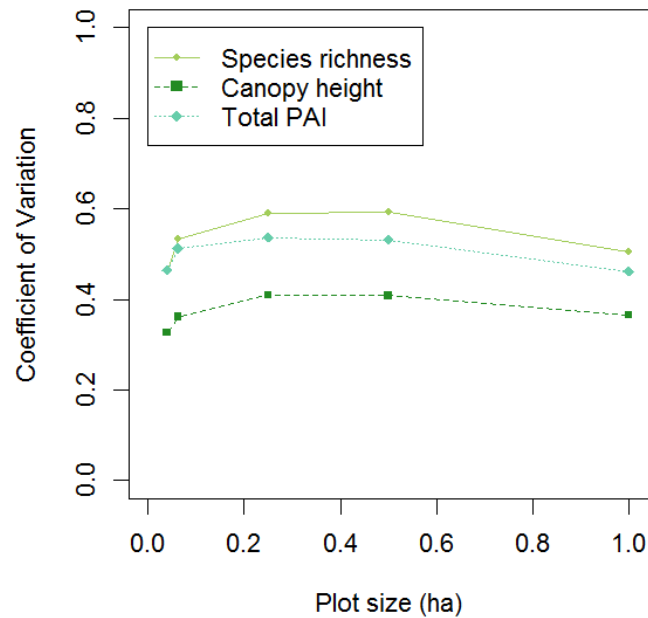


Figure 30: Coefficient of Variation for Species richness, Canopy height and total PAI across all sampled plot sizes.

III. SUPPLEMENTARY MATERIAL FOR CHAPTER 4

III.1 FIELD DATA CHARACTERISTICS

Here we provide information on the characteristics of each field and accompanying lidar dataset.

lop (Lopé, Gabon)

The *lop* field dataset in Gabon was collected as part of the AfriSAR campaign. The study site is located in Central Gabon in the northern section of the Lopé National Park. The field dataset was collected in 2016 and is published (Labrière et al., 2018). The monthly minima and maxima temperature ranges between 20-23 °C and 26-33 °C, respectively. Mean annual precipitation is ~1440 mm/year. The dataset consists of 12 field plots, distributed over five vegetation types: Savanna (3), Colonizing forest (3), monodominant Okoumé forest (3), young Marantaceae forest (2) and mixed Marantaceae forest (1). The three colonizing forest plots are 0.5 ha (100x50 m), all other plots are 1.0 ha (100x100 m). Each plot was stem mapped and all trees with $DBH \geq 5$ cm were recorded. Because of the presence of stem maps all 1.0 ha plots were used at all three spatial resolutions in this study, while the 0.5 ha plots were only used at the 0.25 and 0.0625 ha resolutions (Labrière et al., 2018). All plots had overlapping airborne lidar data collected in 2016 with the LVIS full-waveform lidar instrument, with a footprint spacing of ~8 m and a denser sampling in areas with overlapping flightlines (Tang et al., 2018).

mon (Mondah, Gabon)

The *mon* field dataset in Gabon was collected as part of the AfriSAR campaign in 2016 and this data is publicly accessible through the NASA Distributed Active Archive Center (Fatoyinbo et al., 2018). The mean annual temperature in this study site is ~25 °C, with a precipitation ranging between 3000-3500 mm. The forests across the different field plots have experienced different degrees of disturbance, ranging from slightly disturbed to highly disturbed, given their proximity to Gabon's capital Libreville (Walters et al., 2016). The data consists of 15 1 ha field plots, for which each tree with DBH \geq 1 cm was located, identified and measured. Only 12 of the field plots had overlapping lidar data, collected in 2016 with the LVIS lidar instrument during the AfriSAR campaign (Tang et al., 2018). Of these 12 plots only 9 had a species identification quality high enough to be included in this study (> 80% of the trees identified to the genus level). Given the presence of stem maps, each plot could be used at all three spatial resolution by creating subplots and using the tree location to assign them to the respective subplots.

mab (Mabounié, Gabon)

The *mab* study site is located in west-central Gabon. This study site has experienced a lower degree of forest degradation than the *mon* study sites. The mean annual temperature lays at 25 °C and mean annual precipitation at roughly 2030 mm. Wet and dry seasons occur. The study site consists of 12 field plots of 1 ha collected in 2012. For each of the plots, all trees with DBH \geq 10 cm were assigned within 20x20 m subplots and identified to the species level. Each of the plots was coincident with

airborne lidar data collected in 2016 with a the LVIS full-waveform lidar instrument, (Tang et al., 2018). The plots from this study site were only used at the 1.0 ha plot size because the native subplot size which was unsuitable for creating subplots of 25x25 (0.0626 ha) or 50x50 m (0.25 ha).

rab (Rabi, Gabon)

The Rabi study site in Gabon is located in southwestern Gabon. The region experiences an annual rainfall of approximately 2300 mm/year and a mean annual temperature between 25-28 °C (Labrière et al., 2018). The data was collected through a partnership of the Smithsonian Institute and Gabon's National Center for Scientific and Technological Research (CENAREST). All tree measurements were collected over a two year time period between June 2010 and June 2012. A 25 ha plot was stem mapped and for each tree with $DBH \geq 1$ cm, the species, size and location were recorded. Due to the presence of the stem map, this study site could be used at all spatial resolutions. For the regional scale analysis, we included only the four 1 ha plots at each corner of the square 500 x 500 m plot to reduce the mixing of the local and regional relationships between canopy structure and tree species richness. The lidar data was collected in 2016 with the LVIS full-waveform lidar instrument during the AfriSAR campaign (Tang et al., 2018).

mal (Malebo, DRC)

The *mal* dataset was collected in 2011 in the western part of the DRC. The mean annual temperature lays around 24.9 °C and the mean annual precipitation is measured to be 1587 mm. The dataset consists of 32 1-ha field plots in which each

tree of $DBH \geq 10$ cm is located inside a 50x50 m subplot. Therefore, this dataset was only used for analyses at the 0.25 and 1.0 ha resolutions. Lidar data was collected in 2014 coincident with 21 of the plots of which all had a high species identification quality and are included in this paper.

kea (Central Congo Basin, DRC)

The *kea* dataset was collected in the UNESCO Man and Biosphere reserve in YGB in the Democratic Republic of Congo and published in Kearsley *et al.* (2013). The average annual precipitation is 1762 mm/year with a constant temperature of ~ 25 °C. The forest plots cover different types of forest ranging from moist evergreen rainforest, to transition forest to swamp forest. In this study we did not include the plots located in swamp forest as this is a *terra firme* forest analysis only because we expect the relationship between canopy structure and tree species richness in such ecosystems to be different. Nine 1 ha stem mapped plots for which all trees with a $DBH \geq 10$ cm were located, identified and measured were included in this analysis. The plots were used in the analyses at all three spatial resolutions (Kearsley *et al.*, 2013). The coincident lidar data was collected in 2014.

s11 (SFX 2011, Brazil)

In 2011 a dataset consisting of nine 40 x 40 m field plots was collected in the São Félix do Xingu region located in the State of Pará in Brazil⁴. The region experiences a mean annual temperature of 25.1 °C and a mean annual precipitation around 1972 mm.

⁴

<https://www.paisagenslidar.cnptia.embrapa.br/geonetwork/srv/por/catalog.search#/metadata/c141496b-1fe3-40aa-926b-d2c88903df97>

The plots are located in a linear fashion, originally meant to cover areas where ICESat waveforms were collected. The type of forest ranges from old-growth forest to highly disturbed through (multiple) fire(s) and/or logging. The data collection was carried out through the Paisagens Sustentáveis (Sustainable Landscapes) program funded through the US Agency for International Development and the US Department of State. Each tree with a DBH ≥ 10 cm was measured, identified and located within each field plot, allowing this dataset to be used only at the 0.0625 ha resolution. Eight of the nine plots were used in this paper, as these met species identification quality requirement. The plots are located along a north-south transect. The US Forest Service in collaboration with the Brazilian Agricultural Research Corporation (Embrapa) collected coincident lidar data over all field plots in 2012 with a ALTM 3100 scanner, and a resulting point density of 30.1 points/m². Both datasets are available through the Sustainable Landscapes data portals. Field identifications were done by parataxonomists and in general the identifications were not checked by professional botanists nor were voucher specimens taken in most cases. We are unable to evaluate the data quality for identifications. We used them primarily to estimate wood density.

s12 (SFX 2012, Brazil)

The *s12* dataset was a follow-up field data collection on the *s11* dataset. This dataset has the same characteristics as the *s11* dataset, but it was collected in 2012 about 100

5

<https://www.paisagenslidar.cnptia.embrapa.br/geonetwork/srv/por/catalog.search#/metadata/a3bb4f79-ef32-4b28-8295-2d963a1044e5>

km west of the original dataset, but with the same spatial distribution of the field plots along ICESat data collection lines⁶. The dataset consists of 22 field plots of 40x40 m, including all trees with a DBH \geq 10 cm. These plots were also only used at the 0.0625 ha resolution due to their native plot size. 21 plots were used in the analysis, as there was one plot that did not meet the 80% identified to genus-level requirement.

tam (Tambopata, Peru)

The *tam* field dataset was collected in lowland Amazonian rainforest in Peru in 2011. The annual rainfall in this region is \sim 2250 mm/year and the mean annual temperature is measured to be 25.4°C. The study site lays at the boundary of a nature reserve where sustainable use of forest resources is granted. Data were originally collected across seven plots, but one of these was located in a swamp area and not included in this study. The plots are 100 x 100 m (1 ha) and each tree with a DBH \geq 10 cm was located, identified and measured. The airborne lidar data was acquired in 2009 with a Leica ALS50 II System. This dataset has a point density of 2.1 points/m² (Boyd et al., 2013).

bci (Barro Colorado Island, Panama)

The 50 ha permanent plot in Barro Colorado Island, Panama, is by far the best studied site included in this study. This site is the only one for which analysis on the relationship between canopy structure and tree species diversity have been carried out previously by other scientists (Wolf et al., 2012). The study site consists of a 500 x

⁶

<https://www.paisagenslidar.cnptia.embrapa.br/geonetwork/srv/por/catalog.search#/metadata/43b6c844-fc8d-42c4-9e54-0946b7e101e8>

1000 m plot on the Barro Colorado Island with a mean annual precipitation of ~2580 mm and a mean annual temperature is 26.5 °C. The plot was first established in 1980 and has since been measured every five years. All these data are available through the ForestGEO data portal⁷. In this analysis we used the data from the 2010 survey. Each tree with a DBH \geq 1 cm is marked, located, identified and measured. The data from the bci study site were used at all spatial resolutions because of the availability of a stem map. The lidar data were collected over BCI in 2015.

lsv (La Selva, Costa Rica)

The *lsv* dataset was collected as part of the CARBONO project at the La Selva Research Station in Costa Rica. Mean annual temperature is 25.9 °C, mean annual precipitation approximately 4035 mm. The study site consists of well-preserved old-growth tropical forest. The field study project focused on the long-term monitoring of tropical rainforest productivity at the landscape scale⁸. Field data has been collected in 18 50 x 100 (0.5 ha) plots for twenty-one consecutive years between 1997 and 2017. In this study we used the data from the 2009 census as this one was the closest in date to the lidar data collection. All trees were identified and stem-mapped and the data were used in this study at the 0.25 and 0.0625 ha resolution. The lidar data were collected with the LVIS instrument in 2009 coincident with all 18 field plots.

⁷ <https://forestgeo.si.edu/sites/neotropics/barro-colorado-island>

⁸ <https://tropicalstudies.org/carbono-project/>

cha (La Selva 2, Costa Rica)

The *cha* dataset was also collected in the same region as the *lsv* dataset, near the La Selva biological station in Costa Rica. The dataset consists of two 0.5 ha plots and one 1 ha plot. All trees with a DBH ≥ 5 cm were located, identified and measured. The plots were used at three or two spatial resolutions, depending on the size of the original plot. The same lidar dataset collected by LVIS in 2009 coincided with these field plots.

rob (Robson Creek, Australia)

The *rob* study site is comprised of a 500 x 500 m (25 ha) plot located in Northern Queensland, Australia (Bradford et al., 2014). Mean annual temperature is ~ 20.6 °C. The plot is considered a supersite and was established as part of the Terrestrial Ecosystem Research Network (TERN). The data are available through the Australian Supersite Network portal⁹. Mean annual precipitation in the region is measured at 1587 mm (Bradford et al., 2014). All trees with a DBH ≥ 10 cm were located, identified and measured, allowing the dataset to be used for the analyses at the 0.0625, 0.25 and 1.0 ha resolutions. Lidar data were collected with a Riegl VZQ560 airborne lidar over the study site in 2012 by the AusCover facility of TERN.

⁹ <https://supersites.tern.org.au>

dan (Danum Valley, Malaysia)

The *dan* dataset comprises a 50 ha (500 x 1000) plot in the Danum Valley conservation area in Malaysia. The data is available through the ForestGeo website¹⁰. Mean annual temperature at the study site is an estimated 25.5 °C, mean annual precipitation approximately 2307 mm. The plot was established in 2011 and all trees of DBH \geq 10 cm were located, identified and measured. However, the tree species identification of this field plot is still ongoing and therefore this plot was not used in the local scale analysis even though it comprised a large continuous area but not every adjacent 1 ha plot had the required level of species identification (> 80% of the trees identified to the genus level). The six corner 1 ha plots and the middle 1 ha plots along the long side of the plot were considered for the regional and global analysis. Of these six plots, two of them had the required tree species identification quality and were included in the regional and global analysis. The lidar data collection was funded through the National Environmental Research Council (NERC) and collected in 2014 with a Leica ALS50-II lidar system. The lidar dataset had a point density of 7.3 points/m².

sep (Sepilok, Borneo)

The *sep* dataset consists of nine 4 ha plots (200 x 200 m) located in the Sepilok Forest Reserve north-east of Borneo. The mean annual precipitation in this region is around 2929 mm and mean annual temperature between 26.7-27.7 °C (DeWalt et al., 2006).

¹⁰ <https://forestgeo.si.edu/sites/asia/danum-valley>

These field data were collected in 2014. From each original 4 ha field plot, one 1 ha plot (or one 0.25 ha plot, depending on the resolution of the analysis) was selected at random during the Monte Carlo simulations to avoid the confusion of the local and regional variation in canopy structure and tree species diversity. The *dan* lidar dataset covered seven of the nine field plots. These seven were included in the analysis in this paper.

Data processing

The data processing resulted in a variable number of plots at each resolution for each dataset. These numbers are represented in Table 9.

Table 9: Number of plots included at each spatial resolution for each dataset with a percentage of trees identified up to genus level > 80%.

Country	Project name	No. native plots	Total area (ha)	# of 1.0 ha plots	# of 0.25 ha plots	# of 0.0625 ha plots
Oceania						
Australia	<i>rob</i>	1	25	25/4	100/16	400/64
South-East Asia						
Malaysia	<i>dan</i>	1	50	2	8	36
Borneo	<i>sep</i>	9	36	36	144	-
Africa						
DRC	<i>mal</i>	21	21	21	62	-
DRC	<i>kea</i>	19	19	9	35	140
Gabon	<i>rab</i>	1	25	25/4	100/16	399/64
Gabon	<i>lop</i>	11	9.5	8	37	140
Gabon	<i>mon</i>	10	10	9	29	93
Gabon	<i>mab</i>	10	10	10	-	-
South America						
Peru	<i>tam</i>	6	6	6	24	96
Brazil	<i>s11</i>	9	1.44	-	-	8
Brazil	<i>s12</i>	19	4.8	-	-	22
Central America						
Costa Rica	<i>lsv</i>	12	6	-	36	144
Costa Rica	<i>cha</i>	3	1.5	1	8	32
Panama	<i>bci</i>	1	50	50	179	726

Simulated GEDI waveforms were created from each airborne lidar dataset using the GEDI waveform simulator (Hancock et al., 2019). The simulator takes discrete return airborne laser scanning data, or LVIS lidar waveforms, and makes them look like GEDI waveforms. The waveforms were processed using the gediMetric software, which extracts metrics of interest from the simulated waveforms. In this study we used the rhReal98 metric to reflect canopy height, and the hiLAI profile for represent the PAI profile as this method provides a profile is most similar to profiles derived from the LVIS waveforms verified and used in (Marselis et al., 2019, 2018).

III.2 LOCAL MODEL PERFORMANCE DETAILS

Example plots of observed vs. predicted tree species richness are provided here to illustrate the meaning of a low R^2 values accompanied by low RSE%. This model performance indicates that the species richness predictions (y-axis) are around the mean observed in the superplot (x-axis), but there is little sensitivity to the local variation in tree species richness within the plot (Figure 31).

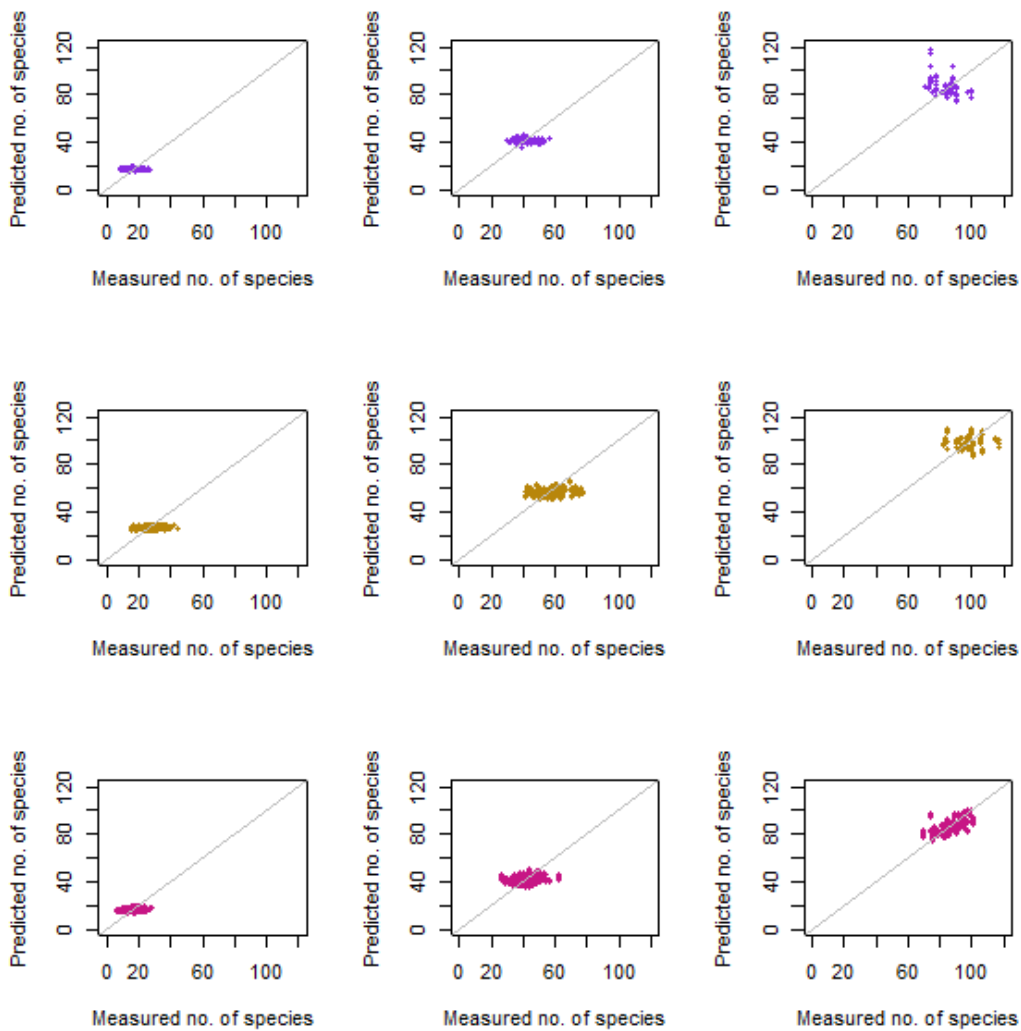


Figure 31: Predicted vs. Measured species richness from local scale predictions in *rab* (top row), *rob* (middle row) and *bci* (bottom row).

III.3 ECOLOGICAL AND STRUCTURAL DISTANCE

We computed the ecological distance between each of the plots in relation to the spatial distance between these plots. We used the method proposed by Bray-Curtis to calculate this distance (Faith et al., 1987). This method takes a table with the species in the columns, the plots in the rows and the number of each species in each plot as content and computes a measure of similarity between the compositions expressed as the semi-variance ranging between 0-1. The relationships are shown for each study site at the 0.25 ha resolution (or 0.16 ha in case the 0.25 ha plots were not available) in (Figure 32 to Figure 46).

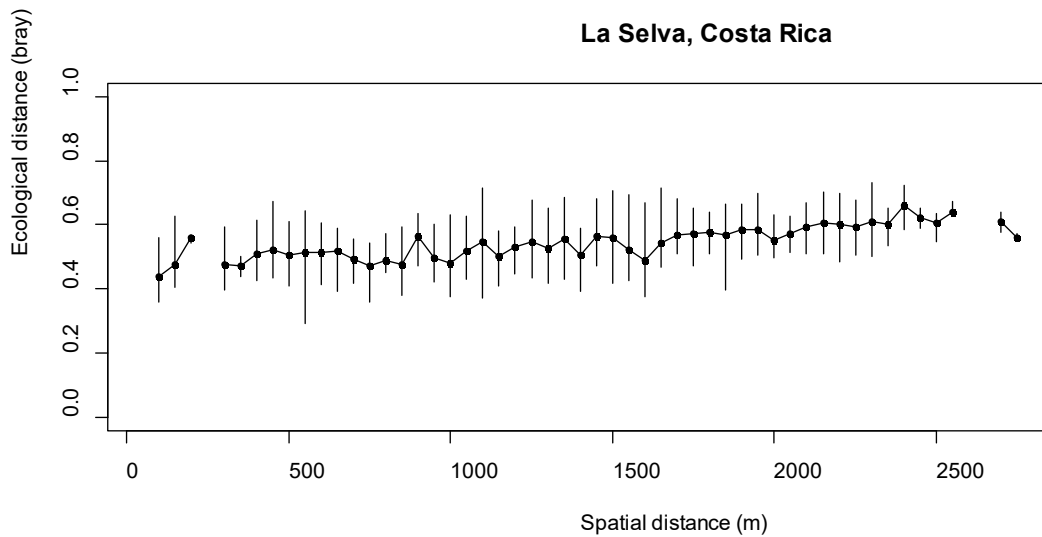


Figure 32: Ecological distance vs. spatial distance in *lsv* (Costa Rica)

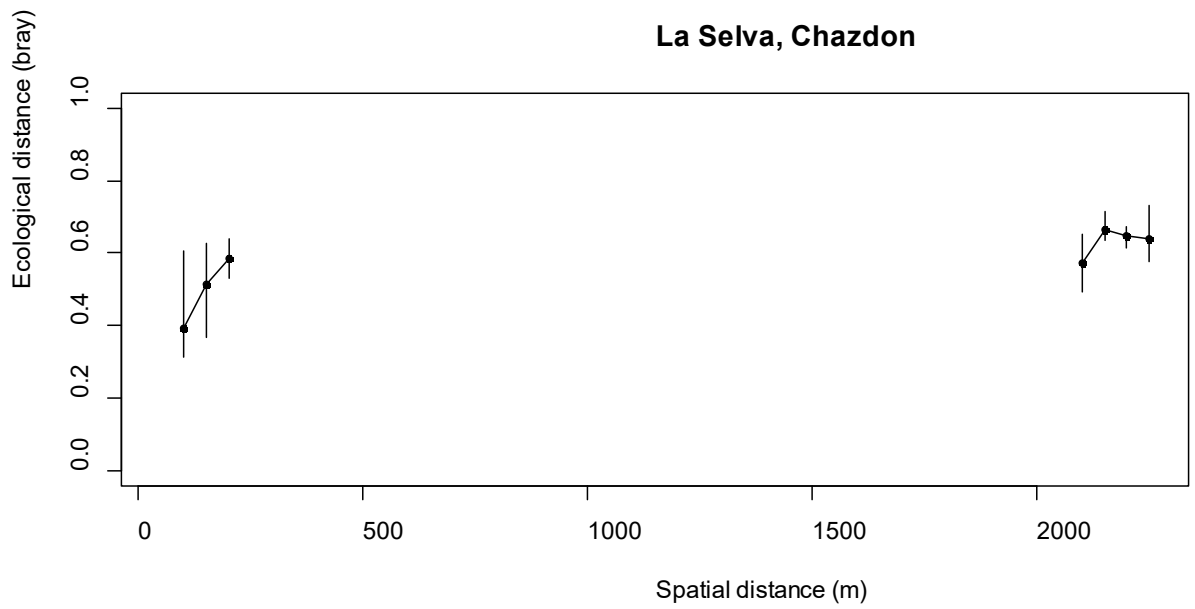


Figure 33: Ecological distance vs. spatial distance in *cha* (Costa Rica)

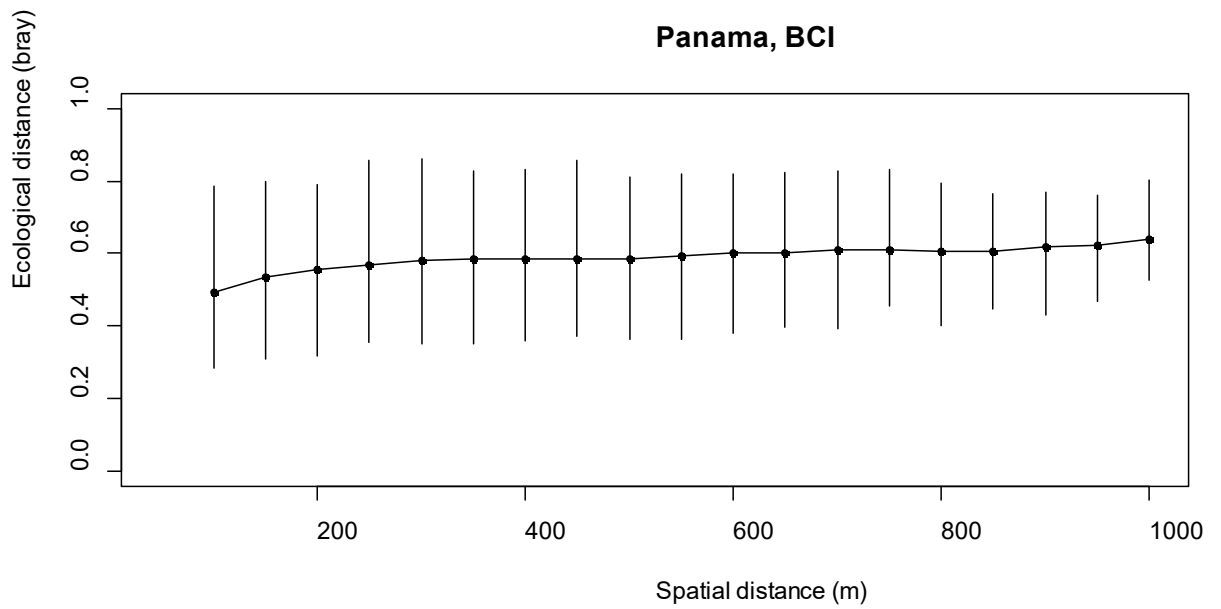


Figure 34: Ecological distance vs. spatial distance in *bci* (Panama)

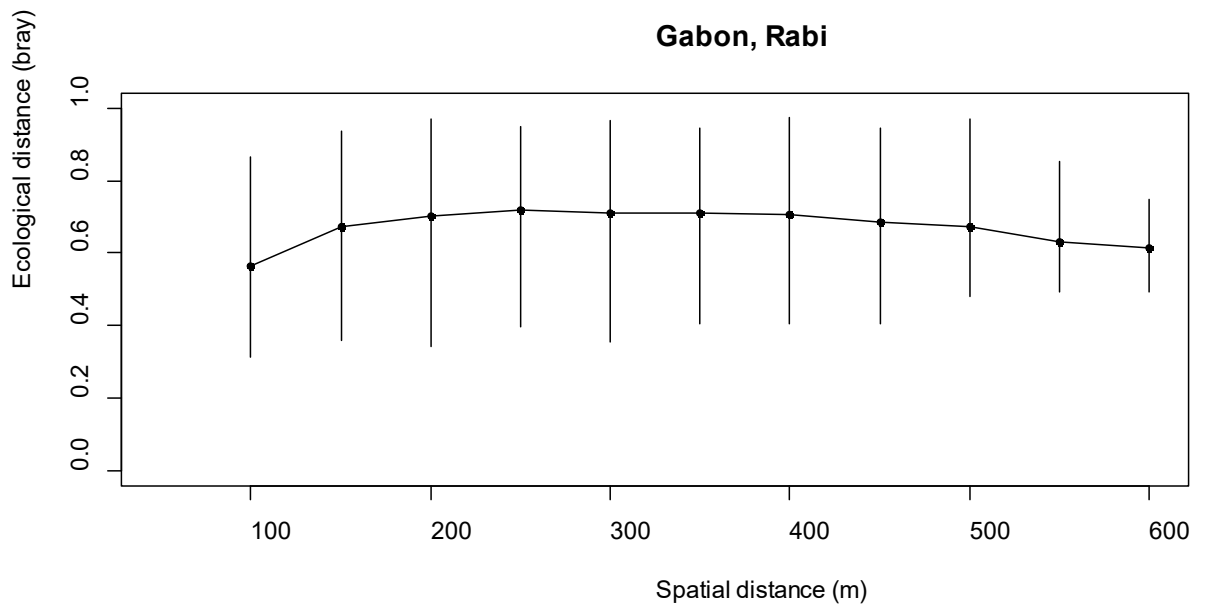


Figure 35: Ecological distance vs. spatial distance in *rab* (Gabon)

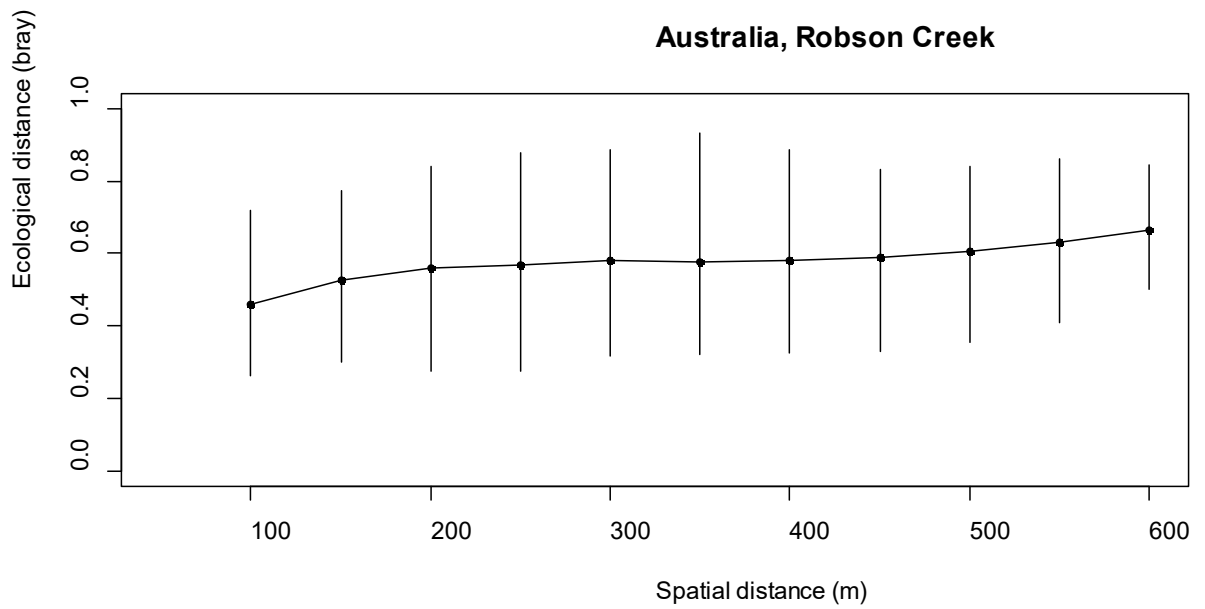


Figure 36: Ecological distance vs. spatial distance in *rob* (Australia)

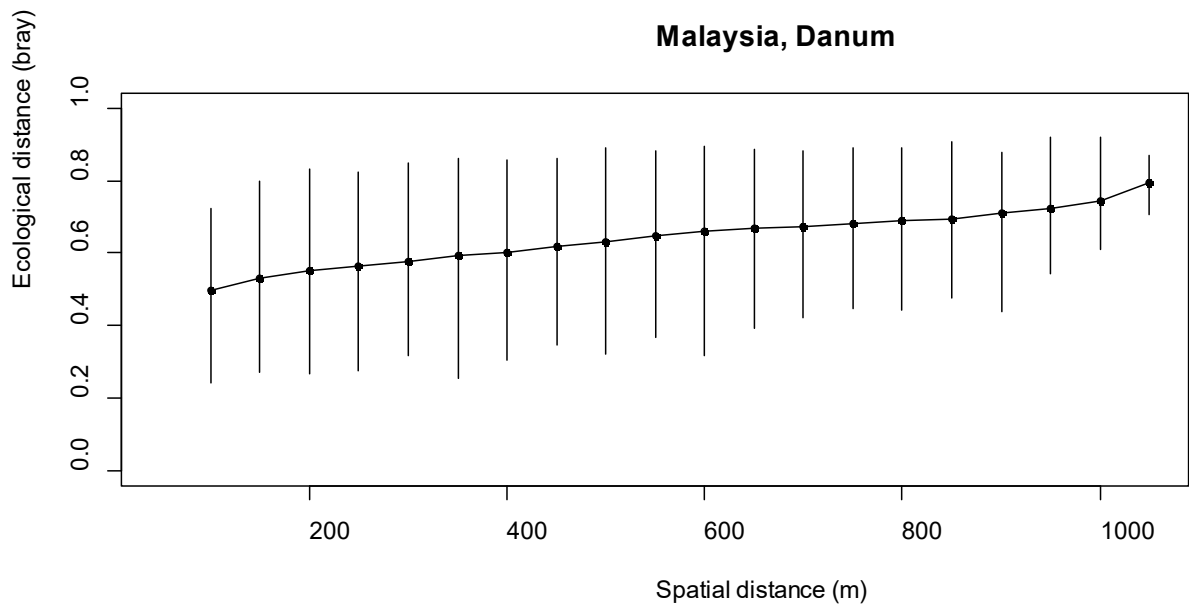


Figure 37: Ecological distance vs. spatial distance in *dan* (Malaysia)

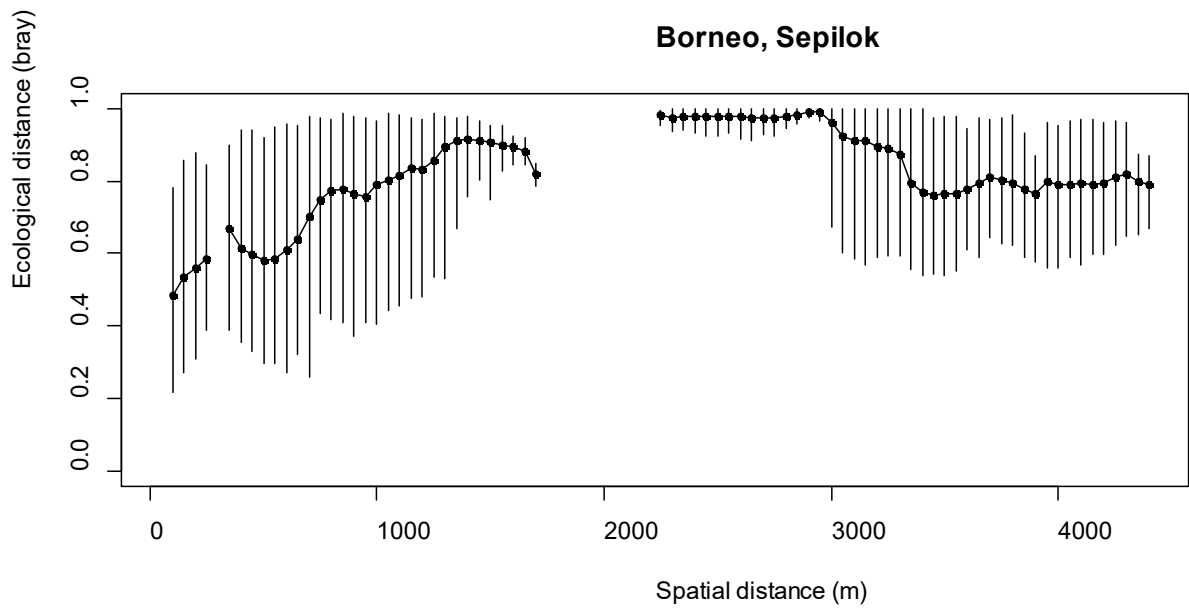


Figure 38: Ecological distance vs. spatial distance in *sep* (Borneo)

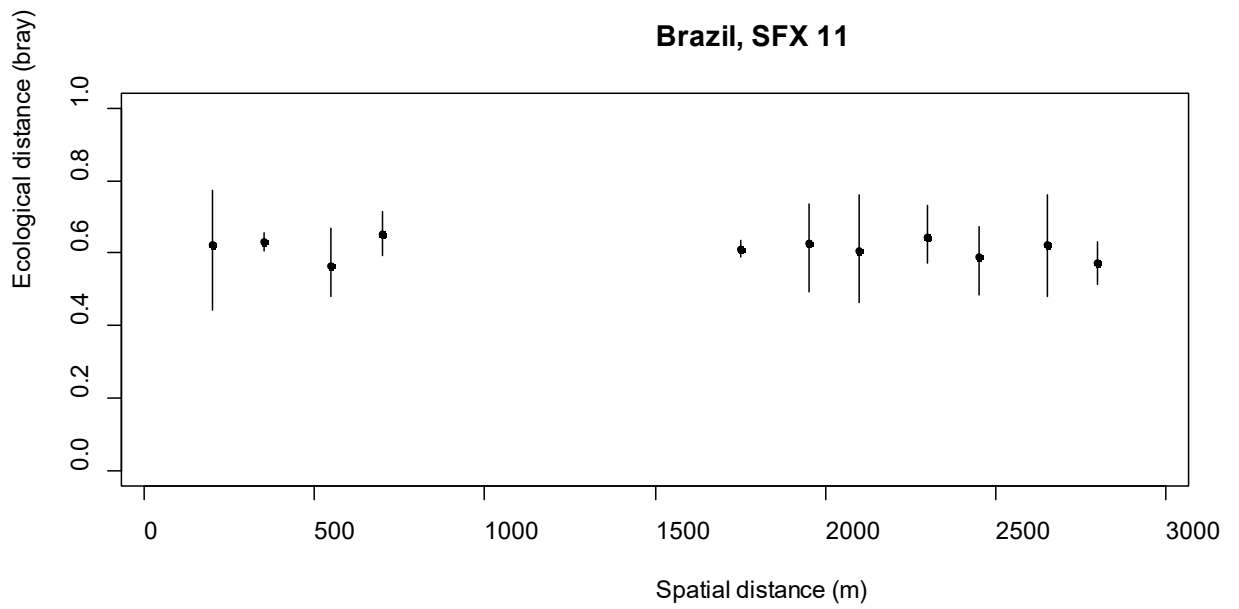


Figure 39: Ecological distance vs. spatial distance in *s11* (Brazil)

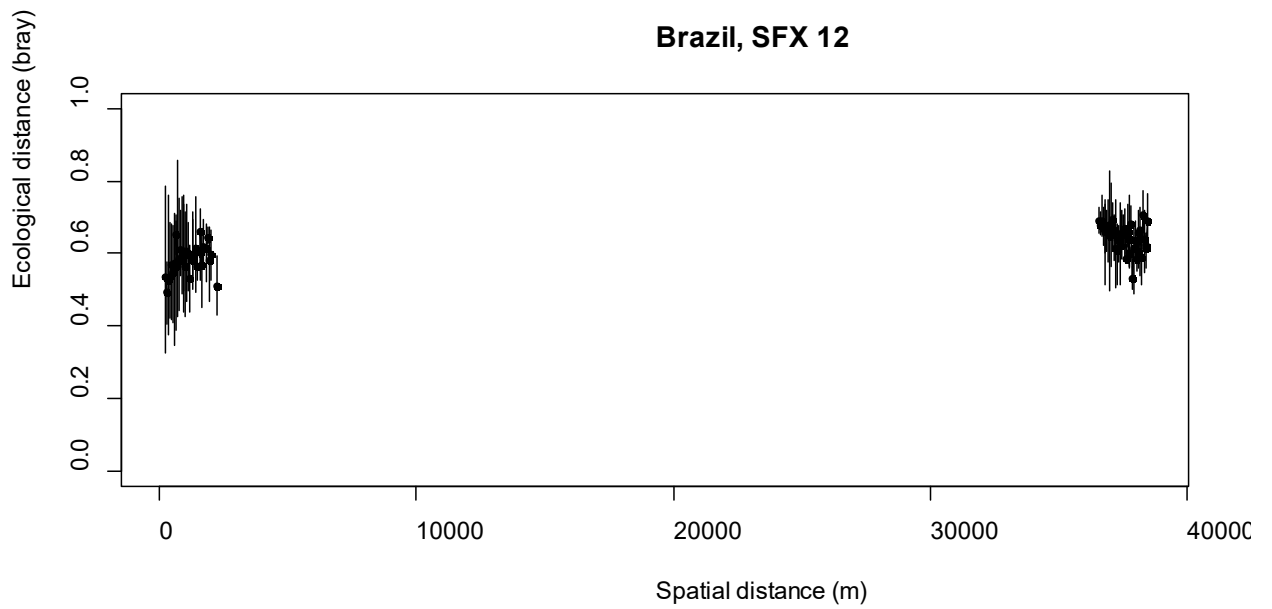


Figure 40: Ecological distance vs. spatial distance in *s12* (Brazil)

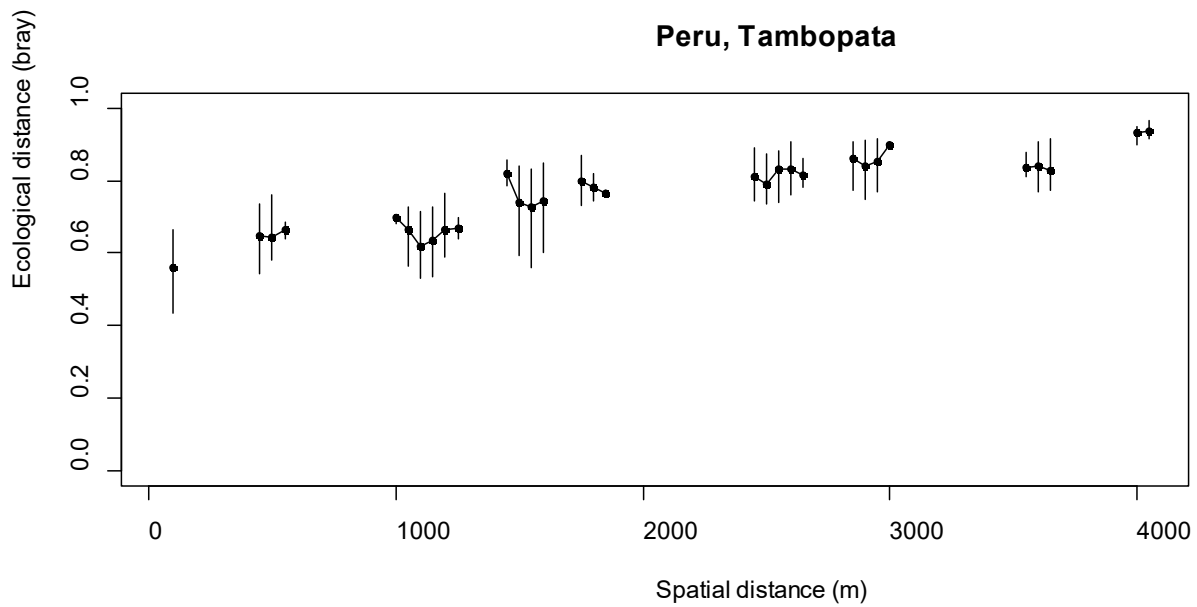


Figure 41: Ecological distance vs. spatial distance in *tam* (Peru)

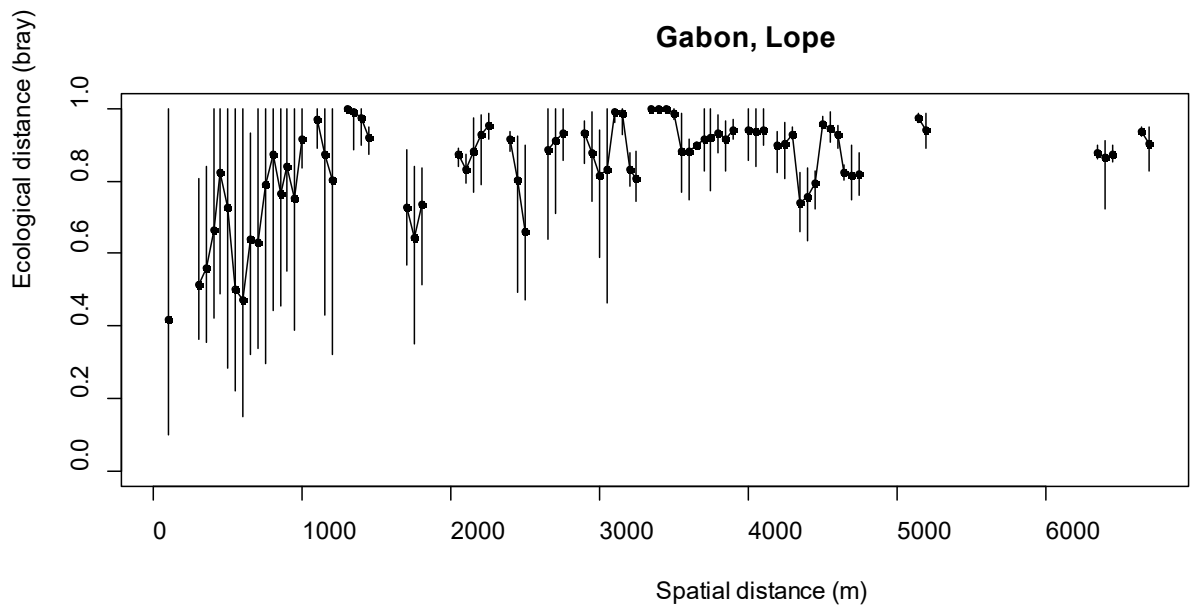


Figure 42: Ecological distance vs. spatial distance in *lop* (Gabon)

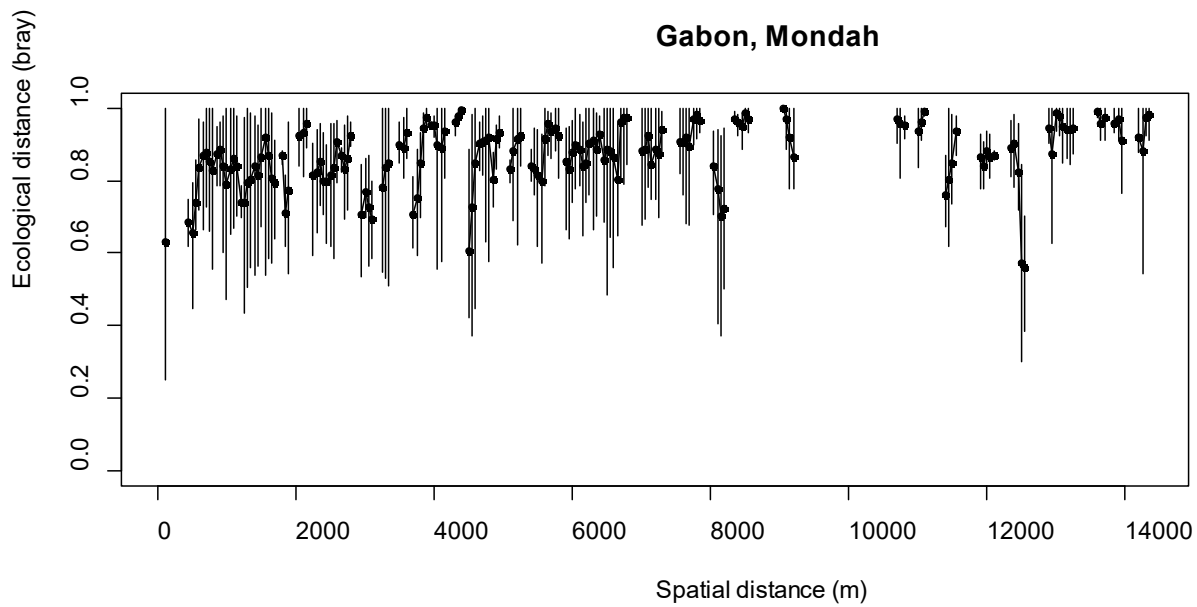


Figure 43: Ecological distance vs. spatial distance in *mon* (Gabon)

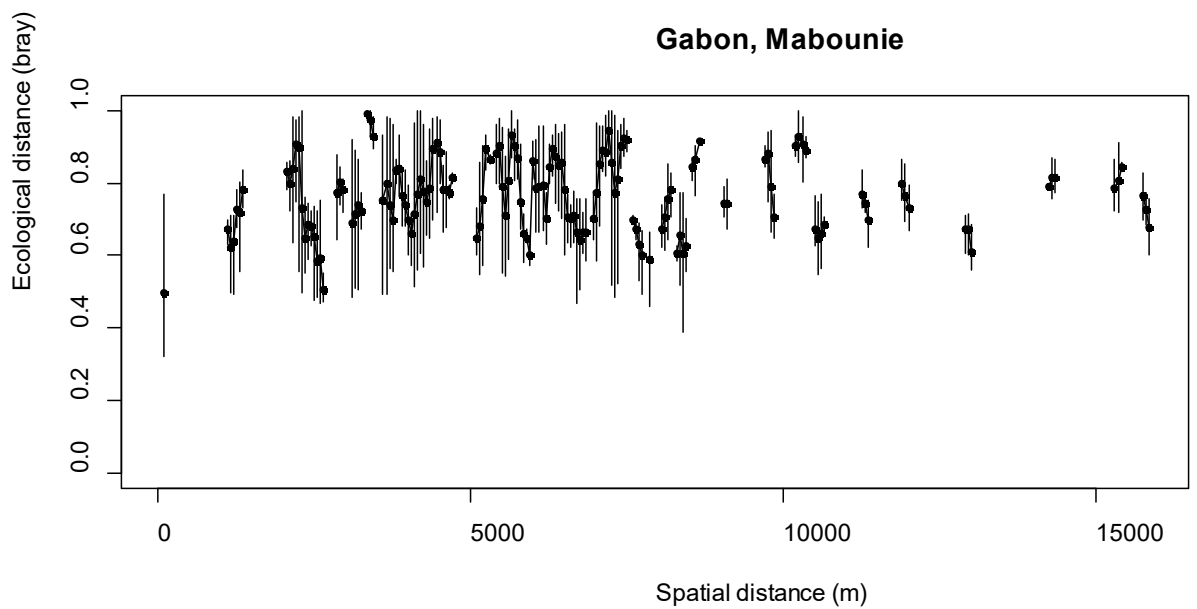


Figure 44: Ecological distance vs. spatial distance in *mab* (Gabon)

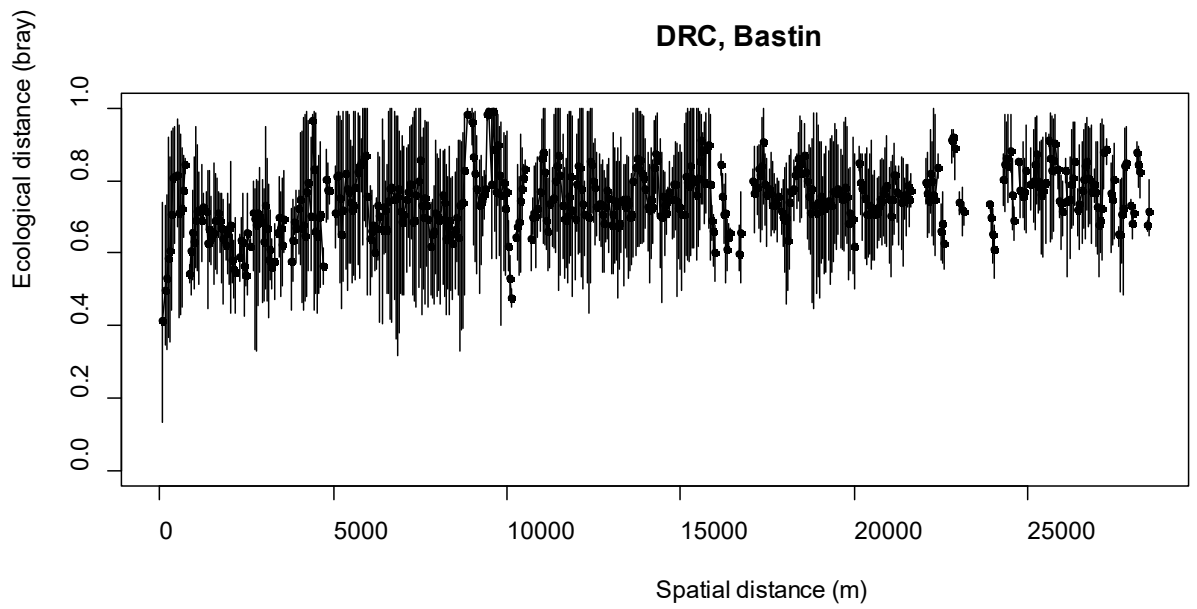


Figure 45: Ecological distance vs. spatial distance in *mal* (DRC)

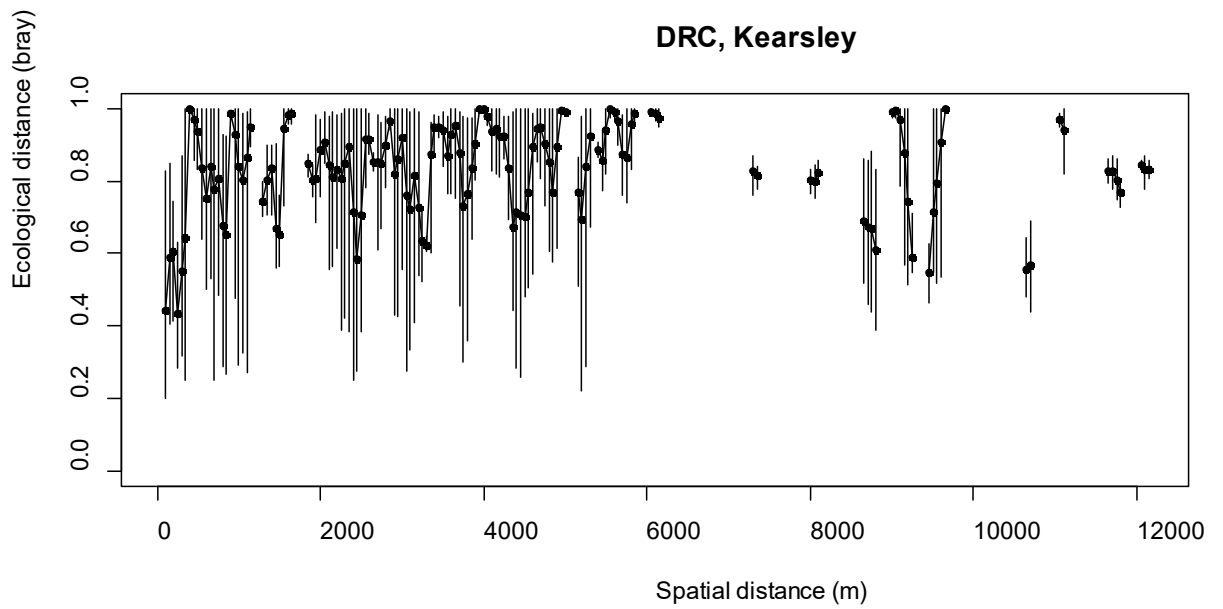


Figure 46: Ecological distance vs. spatial distance in *kea* (DRC)

III.4 DETAILED GLOBAL MODELING RESULTS

Examples of the observed vs. predicted plots from the global analysis are shown in (Figure 47). The plots illustrate that the *tam* plot in Peru (orange dot) is a major outlier in the global predictions at 1 ha resolution. At the 0.0625 ha resolution this study site is no longer an outlier as observed values of species richness are now less extreme and within the range of values of the other datasets.

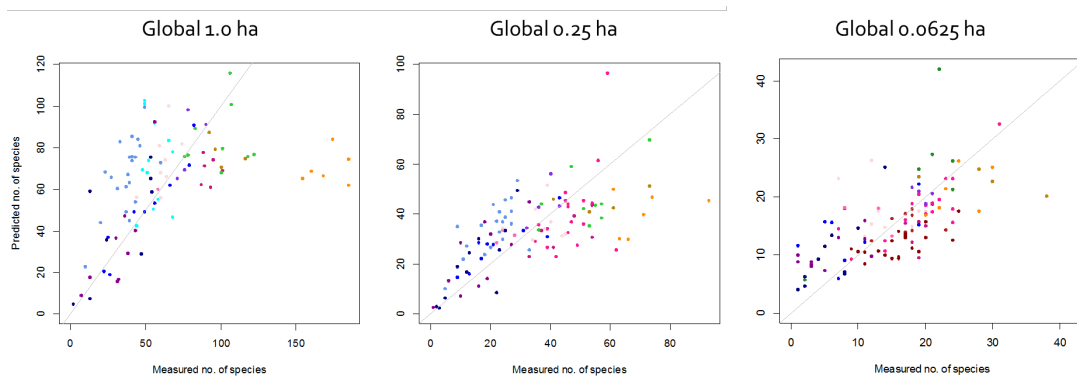


Figure 47: Examples of observed vs. predicted tree species richness using cross-validated global models from one random Monte-Carlo simulation at each of the three spatial resolutions. Colors of points coincide with colors in Figure 15.

REFERENCES

- Anderson-Teixeira, K.J., Davies, S.J., Bennett, A.C., Gonzalez-Akre, E.B., Muller-Landau, H.C., Joseph Wright, S., Abu Salim, K., Almeyda Zambrano, A.M., Alonso, A., Baltzer, J.L., Basset, Y., Bourg, N.A., Broadbent, E.N., Brockelman, W.Y., Bunyavejchewin, S., Burslem, D.F.R.P., Butt, N., Cao, M., Cardenas, D., Chuyong, G.B., Clay, K., Cordell, S., Dattaraja, H.S., Deng, X., Detto, M., Du, X., Duque, A., Erikson, D.L., Ewango, C.E.N., Fischer, G.A., Fletcher, C., Foster, R.B., Giardina, C.P., Gilbert, G.S., Gunatilleke, N., Gunatilleke, S., Hao, Z., Hargrove, W.W., Hart, T.B., Hau, B.C.H., He, F., Hoffman, F.M., Howe, R.W., Hubbell, S.P., Inman-Narahari, F.M., Jansen, P.A., Jiang, M., Johnson, D.J., Kanzaki, M., Kassim, A.R., Kenfack, D., Kibet, S., Kinnaird, M.F., Korte, L., Kral, K., Kumar, J., Larson, A.J., Li, Y., Li, X., Liu, S., Lum, S.K.Y., Lutz, J.A., Ma, K., Maddalena, D.M., Makana, J.R., Malhi, Y., Marthews, T., Mat Serudin, R., McMahon, S.M., McShea, W.J., Memiaghe, H.R., Mi, X., Mizuno, T., Morecroft, M., Myers, J.A., Novotny, V., de Oliveira, A.A., Ong, P.S., Orwig, D.A., Ostertag, R., den Ouden, J., Parker, G.G., Phillips, R.P., Sack, L., Sainge, M.N., Sang, W., Sri-ngernyuang, K., Sukumar, R., Sun, I.F., Sungpalee, W., Suresh, H.S., Tan, S., Thomas, S.C., Thomas, D.W., Thompson, J., Turner, B.L., Uriarte, M., Valencia, R., Vallejo, M.I., Vicentini, A., Vrška, T., Wang, Xihua, Wang, Xugao, Weiblen, G., Wolf, A., Xu, H., Yap, S., Zimmerman, J., 2015. CTFS-ForestGEO: A worldwide network monitoring forests in an era of global change. *Glob. Chang. Biol.* 21, 528–549. <https://doi.org/10.1111/gcb.12712>
- Anderson, C.B., 2018. Biodiversity monitoring, earth observations and the ecology of scale. *Ecol. Lett.* 21. <https://doi.org/10.1111/ele.13106>
- Asner, G.P., Rudel, T.K., Aide, T.M., DeFries, R., Emerson, R., 2009. A contemporary assessment of change in humid tropical forests. *Conserv. Biol.* 23, 1386–1395.
- Barthlott, W., Mutke, J., Rafiqpoor, D., Kier, G., Kreft, H., 2005. Global Centers of Vascular Plant Diversity. *Nov. Acta Leopoldina* 92, 61–83.
- Bastin, J.F., Barbier, N., Réjou-Méchain, M., Fayolle, A., Gourlet-Fleury, S., Maniatis, D., De Haulleville, T., Baya, F., Beeckman, H., Beina, D., Coueron, P., Chuyong, G., Dauby, G., Doucet, J.L., Droissart, V., Dufrêne, M., Ewango, C., Gillet, J.F., Gonmadje, C.H., Hart, T., Kavali, T., Kenfack, D., Libalah, M., Malhi, Y., Makana, J.R., Pélissier, R., Ploton, P., Serckx, A., Sonké, B., Stevart, T., Thomas, D.W., De Cannière, C., Bogaert, J., 2015. Seeing Central African forests through their largest trees. *Sci. Rep.* 5. <https://doi.org/10.1038/srep13156>
- Bergen, K.M., Goetz, S.J., Dubayah, R.O., Henebry, G.M., Hunsaker, C.T., Imhoff,

- M.L., Nelson, R.F., Parker, G.G., Radeloff, V.C., 2009. Remote sensing of vegetation 3-D structure for biodiversity and habitat: Review and implications for lidar and radar spaceborne missions. *J. Geophys. Res.* 114, 13. <https://doi.org/10.1029/2008jg000883>
- Blair, J.B., Rabine, D.L., Hofton, M.A., 1999. The Laser Vegetation Imaging Sensor: a medium-altitude, digitisation-only, airborne laser altimeter for mapping vegetation and topography. *Isprs J. Photogramm. Remote Sens.* 54, 115–122. [https://doi.org/10.1016/s0924-2716\(99\)00002-7](https://doi.org/10.1016/s0924-2716(99)00002-7)
- Bonan, G.B., 2008. Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. *Science* (80-.). 320, 1444–1449. <https://doi.org/10.1126/science.1155121>
- Boyd, D.S., Hill, R.A., Hopkinson, C., Baker, T.R., 2013. Landscape-scale forest disturbance regimes in southern Peruvian Amazonia. *Ecol. Appl.* 23, 1588–1602. <https://doi.org/10.1890/12-0371.1>
- Bradford, M.G., Metcalfe, D.J., Ford, A., Liddell, M.J., McKeown, A., 2014. Floristics, stand structure and aboveground biomass of a 25-ha rainforest plot in the wet tropics of Australia. *J. Trop. For. Sci.* 26, 543–553.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Calders, K., Armston, J., Newnham, G., Herold, M., Goodwin, N., 2014. Implications of sensor configuration and topography on vertical plant profiles derived from terrestrial LiDAR. *Agric. For. Meteorol.* <https://doi.org/10.1016/j.agrformet.2014.03.022>
- Condit, R., 2000. Spatial patterns in the distribution of tropical tree species. *Science* (80-.). 288, 1414–1418. <https://doi.org/10.1126/science.288.5470.1414>
- Coomes, D.A., Dalponte, M., Jucker, T., Asner, G.P., Banin, L.F., Burslem, D.F.R.P., Lewis, S.L., Nilus, R., Phillips, O.L., Phua, M.H., Qie, L., 2017. Area-based vs tree-centric approaches to mapping forest carbon in Southeast Asian forests from airborne laser scanning data. *Remote Sens. Environ.* 194, 77–88.
- Corlett, R.T., Primack, R.B., 2011. *Tropical Rain Forests: An Ecological and Biogeographical Comparison: Second Edition*, 2nd ed, Tropical Rain Forests: An Ecological and Biogeographical Comparison: Second Edition. Blackwell Publishing. <https://doi.org/10.1002/9781444392296>
- Cuni-Sanchez, A., White, L.J.T., Calders, K., Jeffery, K.J., Abernethy, K., Burt, A., Disney, M., Gilpin, M., Gomez-Dans, J.L., Lewis, S.L., 2016. African Savanna-Forest Boundary Dynamics: A 20-Year Study. *PLoS One* 11, 23. <https://doi.org/10.1371/journal.pone.0156934>
- de Groot, R.S., Wilson, M.A., Boumans, R.M.J., 2002. A typology for the

- classification, description and valuation of ecosystem functions, goods and services. *Ecol. Econ.* 41, 393–408. [https://doi.org/10.1016/s0921-8009\(02\)00089-7](https://doi.org/10.1016/s0921-8009(02)00089-7)
- Decuyper, M., Mulatu, K.A., Brede, B., Calders, K., Armston, J., Rozendaal, D.M.A., Mora, B., Clevers, J.G.P.W., Kooistra, L., Herold, M., others, 2018. Assessing the structural differences between tropical forest types using Terrestrial Laser Scanning. *For. Ecol. Manage.* 429, 327–335.
- Denslow, J.S., 1980. Patterns of plant species diversity during succession under different disturbance regimes. *Oecologia* 46, 18–21.
- DeWalt, S.J., Ickes, K., Nilus, R., Harms, K.E., Burslem, D.F.R.P., 2006. Liana habitat associations and community structure in a Bornean lowland tropical forest. *Plant Ecol.* 186, 203–216. <https://doi.org/10.1007/s11258-006-9123-6>
- Drake, J.B., Dubayah, R.O., Knox, R.G., Clark, D.B., Blair, J.B., 2002. Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. *Remote Sens. Environ.* 81, 378–392. [https://doi.org/10.1016/s0034-4257\(02\)00013-5](https://doi.org/10.1016/s0034-4257(02)00013-5)
- Dubayah, R.O., Blair, J.B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P., Qi, W., Silva, C., n.d. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Sci. Remote Sens.*
- Faith, D.P., Minchin, P.R., Belbin, L., 1987. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* 69, 57–68. <https://doi.org/10.1007/BF00038687>
- Falkowski, M.J., Evans, J.S., Martinuzzi, S., Gessler, P.E., Hudak, A.T., 2009. Characterizing forest succession with lidar data: An evaluation for the Inland Northwest, USA. *Remote Sens. Environ.* 113, 946–956.
- Fatoyinbo, T.E., Pinto, N., Simard, M., Armston, J., Duncanson, L., Hofton, M., Saatchi, S., Laval, M., Lou, Y., Denbina, M., Dubayah, R., Marselis, S.M., Tang, H., Hancock, S., Hensley, S., 2017. The 2016 NASA AfriSAR campaign: airborne SAR and Lidar measurements of tropical forest structure and biomass in support of future satellite missions. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 4286–4287.
- Fatoyinbo, T.E., Saatchi, S.S., Armston, J.D., Poulsen, J., Marselis, S.M., Pinto, N., White, L.J.T., Jeffery, K.J., 2018. AfriSAR: Mondah Forest Tree Species, Biophysical, and Biomass Data, Gabon, 2016. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/https://doi.org/10.3334/ORNLDAAAC/1580>
- Fedrigo, M., Newnham, G.J., Coops, N.C., Culvenor, D.S., Bolton, D.K., Nitschke,

- C.R., 2018. Predicting temperate forest stand types using only structural profiles from discrete return airborne lidar. *ISPRS J. Photogramm. Remote Sens.* 136, 106–119.
- Foody, G.M., Cutler, M.E.J., 2006. Mapping the species richness and composition of tropical forests from remotely sensed data with neural networks. *Ecol. Modell.* 195, 37–42. <https://doi.org/10.1016/j.ecolmodel.2005.11.007>
- Fricker, G.A., Wolf, J. a., Saatchi, S.S., Gillespie, T.W., 2015. Predicting spatial variations of tree species richness in tropical forests from high resolution remote sensing. *Ecol. Appl.* 25, 150218095111005. <https://doi.org/10.1890/14-1593.1>
- Gaston, K.J., 2000. Global patters in biodiversity. *Nature* 405, 220–227.
- Gatti, R.C., Di Paola, A., Bombelli, A., Noce, S., Valentini, R., 2017. Exploring the relationship between canopy height and terrestrial plant diversity. *Plant Ecol.* 218, 899–908. <https://doi.org/10.1007/s11258-017-0738-6>
- Givnish, T.J., 2017. Tree diversity in relation to tree height: alternative perspectives. *Ecol. Lett.* 20, 395–397.
- Givnish, T.J., 1999. On the causes of gradients in tropical tree diversity. *J. Ecol.* 87, 193–210. <https://doi.org/10.1046/j.1365-2745.1999.00333.x>
- Goetz, S., Steinberg, D., Dubayah, R., Blair, B., 2007. Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA. *Remote Sens. Environ.* 108, 254–263. <https://doi.org/10.1016/j.rse.2006.11.016>
- Hancock, S., Armston, J., Hofton, M., Sun, X., Tang, H., Duncanson, L.I., Kellner, J.R., Dubayah, R., 2019. The GEDI Simulator: A Large-Footprint Waveform Lidar Simulator for Calibration and Validation of Spaceborne Missions. *Earth Sp. Sci.* 6, 294–310. <https://doi.org/10.1029/2018EA000506>
- Hansen, M.C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-resolution global maps of 21st-century forest cover change. *Science* (80-.). <https://doi.org/10.1126/science.1244693>
- Hilker, T., van Leeuwen, M., Coops, N.C., Wulder, M.A., Newnham, G.J., Jupp, D.L.B., Culvenor, D.S., 2010. Comparing canopy metrics derived from terrestrial and airborne laser scanning in a Douglas-fir dominated forest stand. *Trees* 24, 819–832.
- Hill, M.O., 1973. Diversity and evenness: a unifying notation and its consequences. *Ecology* 54, 427–432.

- Hopkinson, C., Chasmer, L., Young-Pow, C., Treitz, P., 2004. Assessing forest metrics with a ground-based scanning lidar. *Can. J. For. Res.* 34, 573–583.
- Huang, Q.Y., Swatantran, A., Dubayah, R., Goetz, S.J., 2014. The Influence of Vegetation Height Heterogeneity on Forest and Woodland Bird Species Richness across the United States. *PLoS One* 9, 10. <https://doi.org/10.1371/journal.pone.0103236>
- Hubbell, S.P., 2001. *The unified neutral theory of biodiversity and biogeography*. Princeton University Press.
- Huston, M., 1979. A General Hypothesis of Species Diversity. *Am. Nat.* 113, 81–101. <https://doi.org/10.1086/283366>
- Jantz, S.M., Barker, B., Brooks, T.M., Chini, L.P., Huang, Q.Y., Moore, R.M., Noel, J., Hurtt, G.C., 2015. Future habitat loss and extinctions driven by land-use change in biodiversity hotspots under four scenarios of climate-change mitigation. *Conserv. Biol.* 29, 1122–1131. <https://doi.org/10.1111/cobi.12549>
- Jenson, S.K., Domingue, J.O., 1988. Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogramm. Eng. Remote Sensing* 54, 1593–1600. [https://doi.org/0099-1112/88/5411-1593\\$02.25/0](https://doi.org/0099-1112/88/5411-1593$02.25/0)
- Jetz, W., Cavender-Bares, J., Pavlick, R., Schimel, D., Davis, F.W., Asner, G.P., Guralnick, R., Kattge, J., Latimer, M.L., Moorcroft, P., Schaepman, M.E., Schildhauer, M.P., Schneider, F.D., Schrodt, F., Stahl, U., Ustin, S.L., 2016. Monitoring plant functional diversity from space 2, 16024.
- John, R., Dalling, J.W., Harms, K.E., Yavitt, J.B., Stallard, R.F., Mirabello, M., Hubbell, S.P., Valencia, R., Navarrete, H., Vallejo, M., Foster, R.B., 2007. Soil nutrients influence spatial distributions of tropical trees species. *Proc. Natl. Acad. Sci. U. S. A.* 104, 864–869. <https://doi.org/10.1073/pnas.0604666104>
- Jucker, T., Asner, G.P., Dalponte, M., Brodrick, P.G., Philipson, C.D., Vaughn, N.R., Arn Teh, Y., Brelsford, C., Burslem, D.F.R.P., Deere, N.J., Ewers, R.M., Kvasnica, J., Lewis, S.L., Malhi, Y., Milne, S., Nilus, R., Pfeifer, M., Phillips, O.L., Qie, L., Renneboog, N., Reynolds, G., Riutta, T., Struebig, M.J., Svátek, M., Turner, E.C., Coomes, D.A., 2018. Estimating aboveground carbon density and its uncertainty in Borneo’s structurally complex tropical forests using airborne laser scanning. *Biogeosciences* 15, 3811–3830. <https://doi.org/10.5194/bg-15-3811-2018>
- Jupp, D.L.B., Culvenor, D.S., Lovell, J.L., Newnham, G.J., Strahler, A.H., Woodcock, C.E., 2009. Estimating forest LAI profiles and structural parameters using a ground-based laser called 'Echidna®'. *Tree Physiol.* 29, 171–181. <https://doi.org/10.1093/treephys/tpn022>

- Kearsley, E., De Haulleville, T., Hufkens, K., Kidimbu, A., Toirambe, B., Baert, G., Huygens, D., Kebede, Y., Defourny, P., Bogaert, J., Beeckman, H., Steppe, K., Boeckx, P., Verbeeck, H., 2013. Conventional tree height-diameter relationships significantly overestimate aboveground carbon stocks in the Central Congo Basin. *Nat. Commun.* 4. <https://doi.org/10.1038/ncomms3269>
- Keil, P., Chase, J.M., 2019. Global patterns and drivers of tree diversity integrated across a continuum of spatial grains. *Nat. Ecol. Evol.* 3, 390. <https://doi.org/10.1038/s41559-019-0799-0>
- Kent, R., Lindsell, J., Laurin, G., Valentini, R., Coomes, D., 2015. Airborne LiDAR detects selectively logged tropical forest even in an advanced stage of recovery. *Remote Sens.* 7, 8348–8367.
- Kier, G., Mutke, J., Dinerstein, E., Ricketts, T.H., Küper, W., Kreft, H., Barthlott, W., 2005. Global patterns of plant diversity and floristic knowledge. *J. Biogeogr.* 32, 1107–1116. <https://doi.org/10.1111/j.1365-2699.2005.01272.x>
- Kohyama, T., 1996. the role of architecture in enhancing plant species diversity, in: Abe, T., Levin, S. A. & Higashi, M. (eds). (Ed.), *Biodiversity: An Ecological Perspective*. Springer-Verlag, New York, pp. 21–33.
- Kohyama, T., 1993. Size-structured tree populations in gap-dynamic forest--the forest architecture hypothesis for the stable coexistence of species. *J. Ecol.* 131–143.
- Kreft, H., Jetz, W., 2007. Global patterns and determinants of vascular plant diversity. *Proc. Natl. Acad. Sci.* 104, 5925–5930. <https://doi.org/10.1073/pnas.0608361104>
- Labrière, N., Tao, S., Chave, J., Scipal, K., Le Toan, T., Abernethy, K., Alonso, A., Barbier, N., Bissiengou, P., Casal, T., others, 2018. In Situ Reference Datasets From the TropiSAR and AfriSAR Campaigns in Support of Upcoming Spaceborne Biomass Missions. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 1–11.
- Laurin, G. V, Chen, Q., Lindsell, J.A., Coomes, D.A., Del Frate, F., Guerriero, L., Pirotti, F., Valentini, R., 2014. Above ground biomass estimation in an African tropical forest with lidar and hyperspectral data. *Isprs J. Photogramm. Remote Sens.* 89, 49–58. <https://doi.org/10.1016/j.isprsjprs.2014.01.001>
- Lawrence, I., Lin, K., 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 255–268.
- Lee, M.E., Alonso, A., Dallmeier, F., Campbell, P., Pauwels, O.S., 2006. The Gamba complex of protected areas: an illustration of Gabon's biodiversity. *Bull. Biol. Soc. Washingt.* 12, 229–241.
- Lee, S.-K., Fatoyinbo, T., Qi, W., Hancock, S., Armston, J., Dubayah, R., 2018. Gedi and Tandem-X Fusion for 3D Forest Structure Parameter Retrieval, in: *IGARSS*

- 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. pp. 380–382.
- Lefsky, M.A., Cohen, W.B., Parker, G.G., Harding, D.J., 2002. Lidar remote sensing for ecosystem studies. *Bioscience* 52.
- Ligon, F.K., Dietrich, W.E., Trush, W.J., 1995. Downstream ecological effects of dams. *Bioscience* 45, 183–192. <https://doi.org/10.2307/1312557>
- Longo, M., Keller, M., Dos-Santos, M.N., Leitold, V., Pinagé, E.R., Baccini, A., Saatchi, S., Nogueira, E.M., Batistella, M., Morton, D.C., 2016. Aboveground biomass variability across intact and degraded forests in the Brazilian Amazon. *Global Biogeochem. Cycles* 30, 1639–1660. <https://doi.org/10.1002/2016GB005465>
- Lopez-Gonzalez, G., Lewis, S.L., Burkit, M., Phillips, O.L., 2011. ForestPlots.net: a web application and research tool to manage and analyse tropical forest plot data. *J. Veg. Sci.* 22, 610–613. <https://doi.org/10.1111/j.1654-1103.2011.01312.x>
- Lopez-Gonzalez, G., Lewis, S.L., Burkitt, M., Baker, T.R., Phillips, O.L., 2009. ForestPlots.net database [WWW Document]. www.forestplots.net.
- Luo, W., Liang, J., Cazzolla Gatti, R., Zhao, X., Zhang, C., 2019. Parameterization of biodiversity--productivity relationship and its scale dependency using georeferenced tree-level data. *J. Ecol.* 107, 1106–1119.
- MacArthur, R.H., MacArthur, T.W., 1961. On bird species diversity. *Ecology* 42, 594–598.
- MacArthur, R.H., Wilson, E.O., 1967. *The theory of Island Biogeography*. Princeton University Press, Princeton and Oxford.
- Mao, L., Dennett, J., Bater, C.W., Tompalski, P., Coops, N.C., Farr, D., Kohler, M., White, B., Stadt, J.J., Nielsen, S.E., 2018. Using airborne laser scanning to predict plant species richness and assess conservation threats in the oil sands region of Alberta's boreal forest. *For. Ecol. Manage.* 409, 29–37.
- Marks, C.O., Muller-Landau, H.C., Tilman, D., 2016. Tree diversity, tree height and environmental harshness in eastern and western North America. *Ecol. Lett.* 19, 743–751. <https://doi.org/10.1111/ele.12608>
- Marselis, S.M., Tang, H., Armston, J., Abernethy, K., Alonso, A., Barbier, N., Bissengou, P., Jeffery, K., Kenfack, D., Labrière, N., others, 2019. Exploring the relation between remotely sensed vertical canopy structure and tree species diversity in Gabon. *Environ. Res. Lett.* 14. <https://doi.org/https://doi.org/10.1088/1748-9326/ab2dcd>

- Marselis, S.M., Tang, H., Armston, J.D., Calders, K., Labrière, N., Dubayah, R., 2018. Distinguishing vegetation types with airborne waveform lidar data in a tropical forest-savanna mosaic: A case study in Lopé National Park, Gabon. *Remote Sens. Environ.* 216, 626–634.
- Marselis, S.M., Yebra, M., Jovanovic, T., van Dijk, A., 2016. Deriving comprehensive forest structure information from mobile laser scanning observations using automated point cloud classification. *Environ. Model. Softw.* 82, 142–151. <https://doi.org/10.1016/j.envsoft.2016.04.025>
- Marzeion, B., Cogley, J.G., Richter, K., Parkes, D., 2014. Attribution of global glacier mass loss to anthropogenic and natural causes. *Science (80-.)*. 345, 919–921. <https://doi.org/10.1126/science.1254702>
- Matsunaga, T., Iwasaki, A., Tsuchida, S., Tanii, J., Kashimura, O., Nakamura, R., Yamamoto, H., Tachikawa, T., Rokugawa, S., 2013. Current status of Hyperspectral Imager Suite (HISUI), in: *International Geoscience and Remote Sensing Symposium (IGARSS)*. <https://doi.org/10.1109/IGARSS.2013.6723586>
- McElhinny, C., Gibbons, P., Brack, C., Bauhus, J., 2005. Forest and woodland stand structural complexity: Its definition and measurement. *For. Ecol. Manage.* 218, 1–24. <https://doi.org/10.1016/j.foreco.2005.08.034>
- Memighe, H.R., Lutz, J.A., Korte, L., Alonso, A., Kenfack, D., 2016. Ecological Importance of Small-Diameter Trees to the Structure, Diversity and Biomass of a Tropical Evergreen Forest at Rabi, Gabon. *PLoS One* 11. <https://doi.org/10.1371/journal.pone.0154988>
- Moles, A.T., Warton, D.I., Warman, L., Swenson, N.G., Laffan, S.W., Zanne, A.E., Pitman, A., Hemmings, F.A., Leishman, M.R., 2009. Global patterns in plant height. *J. Ecol.* 97, 923–932. <https://doi.org/10.1111/j.1365-2745.2009.01526.x>
- Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B., Worm, B., 2011. How Many Species Are There on Earth and in the Ocean? *Plos Biol.* 9, 8. <https://doi.org/10.1371/journal.pbio.1001127>
- Moran, C.J., Rowell, E.M., Seielstad, C.A., 2018. A data-driven framework to identify and compare forest structure classes using LiDAR. *Remote Sens. Environ.* 211, 154–166. <https://doi.org/10.1016/j.rse.2018.04.005>
- Morris, E.K., Caruso, T., Buscot, F., Fischer, M., Hancock, C., Maier, T.S., Meiners, T., Müller, C., Obermaier, E., Prati, D., others, 2014. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecol. Evol.* 4, 3514–3524.
- Mutke, J., Barthlott, W., 2005. Patterns of vascular plant diversity at continental to global scales. *Biol. Skr.* 55, 521–531.

- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. *Nature* 403, 853–858. <https://doi.org/10.1038/35002501>
- Naeem, S., Chapin, C.F.S., Costanza, R., Ehrlich, P.R., Golley, F.B., Hooper, D.U., Lawton, J.H., O'Neill, R. V, Mooney, H.A., Sala, O.E., Symstad, A.J., Tilman, D., 1999. Biodiversity and ecosystem functioning: maintaining natural life support processes. *Issues Ecol.* 4.
- Newnham, G.J., Armston, J.D., Calders, K., Disney, M.I., Lovell, J.L., Schaaf, C.B., Strahler, A.H., Mark Danson, F., 2015. Terrestrial laser scanning for plot-scale forest measurement. *Curr. For. Reports.* <https://doi.org/10.1007/s40725-015-0025-5>
- Noss, R.F., 1990. Indicators for monitoring biodiversity - a hierarchical approach. *Conserv. Biol.* 4, 355–364. <https://doi.org/10.1111/j.1523-1739.1990.tb00309.x>
- Noss, R.F., Dobson, A.P., Baldwin, R., Beier, P., Davis, C.R., Dellasala, D.A., Francis, J., Locke, H., Nowak, K., Lopez, R., Reining, C., Trombulak, S.C., Tabor, G., 2012. Bolder Thinking for Conservation. *Conserv. Biol.* 26, 1–4. <https://doi.org/10.1111/j.1523-1739.2011.01738.x>
- Palace, M.W., Sullivan, F.B., Ducey, M.J., Treuhart, R.N., Herrick, C., Shimbo, J.Z., Mota-E-Silva, J., 2015. Estimating forest structure in a tropical forest using field measurements, a synthetic model and discrete return lidar data. *Remote Sens. Environ.* 161, 1–11. <https://doi.org/10.1016/j.rse.2015.01.020>
- Patterson, P.L., Healey, S.P., Ståhl, G., Saarela, S., Holm, S., Andersen, H.-E., Dubayah, R., Duncanson, L. i, Hancock, S., Armston, J., Kellner, J.R., Cohen, W.B., Yang, Z., 2019. Statistical properties of hybrid estimators proposed for GEDI – NASA’s Global Ecosystem Dynamics Investigation. *Environ. Res. Lett.* 14, 065007. <https://doi.org/10.1088/1748-9326/ab18df>
- Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith, D.P., Freyhof, J., Gregory, R.D., Heip, C., Höft, R., Hurtt, G., Jetz, W., Karp, D.S., McGeoch, M.A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M., Wegmann, M., 2013. Essential biodiversity variables. *Science* (80-). 339, 277–278. <https://doi.org/10.1126/science.1229931>
- Pimm, S.L., Jenkins, C.N., Abell, R., Brooks, T.M., Gittleman, J.L., Joppa, L.N., Raven, P.H., Roberts, C.M., Sexton, J.O., 2014. The biodiversity of species and their rates of extinction, distribution, and protection. *Science* (80-). 344. <https://doi.org/10.1126/science.1246752>
- Piñeiro, G., Perelman, S., Guerschman, J.P., Paruelo, J.M., 2008. How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecol. Modell.* 216,

316–322.

- Qi, W., Dubayah, R.O., 2016. Combining Tandem-X InSAR and simulated GEDI lidar observations for forest structure mapping. *Remote Sens. Environ.* 187, 253–266.
- Qi, W., Saarela, S., Armston, J., Stahl, G., Dubayah, R., 2019. Forest biomass estimation over three distinct forest types using TanDEM-X InSAR data and simulated GEDI lidar data. *Remote Sens. Environ.* 232. <https://doi.org/https://doi.org/10.1016/j.rse.2019.111283>
- Rios-Saldaña, C.A., Delibes-Mateos, M., Ferreira, C.C., 2018. Are fieldwork studies being relegated to second place in conservation science? *Glob. Ecol. Conserv.* 14.
- Robinson, C., Saatchi, S., Clark, D., Hurtado Astaiza, J., Hubel, A.F., Gillespie, T.W., 2018. Topography and Three-Dimensional Structure Can Estimate Tree Diversity along a Tropical Elevational Gradient in Costa Rica. *Remote Sens.* 10, 629.
- Rocchini, D., Boyd, D.S., Féret, J.B., Foody, G.M., He, K.S., Lausch, A., Nagendra, H., Wegmann, M., Pettorelli, N., 2016. Satellite remote sensing to monitor species diversity: potential and pitfalls. *Remote Sens. Ecol. Conserv.* 2, 25–36. <https://doi.org/10.1002/rse2.9>
- Rockstrom, J., Steffen, W., Noone, K., Persson, A., Chapin, F.S., Lambin, E.F., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H.J., Nykvist, B., de Wit, C.A., Hughes, T., van der Leeuw, S., Rodhe, H., Sorlin, S., Snyder, P.K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R.W., Fabry, V.J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., Foley, J.A., 2009. A safe operating space for humanity. *Nature* 461, 472–475. <https://doi.org/10.1038/461472a>
- Rose, R.A., Byler, D., Eastman, J.R., Fleishman, E., Geller, G., Goetz, S., Guild, L., Hamilton, H., Hansen, M., Headley, R., Hewson, J., Horning, N., Kaplin, B.A., Laporte, N., Leidner, A., Leinagruber, P., Morisette, J., Musinsky, J., Pintea, L., Prados, A., Radeloff, V.C., Rowen, M., Saatchi, S., Schil, S., Tabor, K., Turner, W., Vodacek, A., Vogelnaann, J., Wegmann, M., Wilkie, D., 2015. Ten ways remote sensing can contribute to conservation. *Conserv. Biol.* 29, 350–359. <https://doi.org/10.1111/cobi.12397>
- Saarela, S., Holm, S., Healey, S.P., Andersen, H.E., Petersson, H., Prentius, W., Patterson, P.L., Næsset, E., Gregoire, T.G., Ståhl, G., 2018. Generalized hierarchical model-based estimation for aboveground biomass assessment using GEDI and landsat data. *Remote Sens.* <https://doi.org/10.3390/rs10111832>
- Schäfer, E., Heiskanen, J., Heikinheimo, V., Pellikka, P., 2016. Mapping tree species diversity of a tropical montane forest by unsupervised clustering of airborne

imaging spectroscopy data. *Ecol. Indic.* 64, 49–58.
<https://doi.org/10.1016/j.ecolind.2015.12.026>

- Schimel, D.S., Asner, G.P., Moorcroft, P., 2013. Observing changing ecological diversity in the Anthropocene. *Front. Ecol. Environ.* 11, 129–137.
<https://doi.org/10.1890/120111>
- Sheil, D., Burslem, D.F.R.P., 2003. Disturbing hypotheses in tropical forests. *Trends Ecol. Evol.* 18, 18–26.
- Simard, M., Fatoyinbo, L., Smetanka, C., Rivera-Monroy, V.H., Castañeda-Moya, E., Thomas, N., Van der Stocken, T., 2019. Mangrove canopy height globally related to precipitation, temperature and cyclone frequency. *Nat. Geosci.* 12, 40–45. <https://doi.org/10.1038/s41561-018-0279-1>
- Skidmore, A.K., Pettorelli, N., Coops, N.C., Geller, G.N., Hansen, M., Lucas, R., Mucher, C.A., O'Connor, B., Paganini, M., Pereira, H.M., Schaepman, M.E., Turner, W., Wang, T.J., Wegmann, M., 2015. Agree on biodiversity metrics to track from space. *Nature* 523, 403–405.
- Slik, J.W.F., Arroyo-Rodríguez, V., Aiba, S.-I., Alvarez-Loayza, P., Alves, L.F., Ashton, P., Balvanera, P., Bastian, M.L., Bellingham, P.J., van den Berg, E., Bernacci, L., da Conceição Bispo, P., Blanc, L., Böhning-Gaese, K., Boeckx, P., Bongers, F., Boyle, B., Bradford, M., Brearley, F.Q., Breuer-Ndoundou Hockemba, M., Bunyavejchewin, S., Calderado Leal Matos, D., Castillo-Santiago, M., Catharino, E.L.M., Chai, S.-L., Chen, Y., Colwell, R.K., Chazdon, R.L., Clark, C., Clark, D.B., Clark, D.A., Culmsee, H., Damas, K., Dattaraja, H.S., Dauby, G., Davidar, P., DeWalt, S.J., Doucet, J.-L., Duque, A., Durigan, G., Eichhorn, K.A.O., Eisenlohr, P. V., Eler, E., Ewango, C., Farwig, N., Feeley, K.J., Ferreira, L., Field, R., de Oliveira Filho, A.T., Fletcher, C., Forshed, O., Franco, G., Fredriksson, G., Gillespie, T., Gillet, J.-F., Amarnath, G., Griffith, D.M., Grogan, J., Gunatilleke, N., Harris, D., Harrison, R., Hector, A., Homeier, J., Imai, N., Itoh, A., Jansen, P.A., Joly, C.A., de Jong, B.H.J., Kartawinata, K., Kearsley, E., Kelly, D.L., Kenfack, D., Kessler, M., Kitayama, K., Kooyman, R., Larney, E., Laumonier, Y., Laurance, S., Laurance, W.F., Lawes, M.J., Amaral, I.L. do, Letcher, S.G., Lindsell, J., Lu, X., Mansor, A., Marjokorpi, A., Martin, E.H., Meilby, H., Melo, F.P.L., Metcalfe, D.J., Medjibe, V.P., Metzger, J.P., Millet, J., Mohandass, D., Montero, J.C., de Morisson Valeriano, M., Mugerwa, B., Nagamasu, H., Nilus, R., Ochoa-Gaona, S., Onrizal, Page, N., Parolin, P., Parren, M., Parthasarathy, N., Paudel, E., Permana, A., Piedade, M.T.F., Pitman, N.C.A., Poorter, L., Poulsen, A.D., Poulsen, J., Powers, J., Prasad, R.C., Puyravaud, J.-P., Razafimahaimodison, J.-C., Reitsma, J., dos Santos, J.R., Roberto Spironello, W., Romero-Saltos, H., Rovero, F., Rozak, A.H., Ruokolainen, K., Rutishauser, E., Saiter, F., Saner, P., Santos, B.A., Santos, F., Sarker, S.K., Satdichanh, M., Schmitt, C.B., Schöngart, J., Schulze, M., Suganuma, M.S., Sheil, D., da Silva Pinheiro, E., Sist, P., Stevart, T., Sukumar, R., Sun, I.-F., Sunderland, T., Suresh, H.S., Suzuki, E., Tabarelli, M., Tang, J.,

- Targhetta, N., Theilade, I., Thomas, D.W., Tchouto, P., Hurtado, J., Valencia, R., van Valkenburg, J.L.C.H., Van Do, T., Vasquez, R., Verbeeck, H., Adekunle, V., Vieira, S.A., Webb, C.O., Whitfeld, T., Wich, S.A., Williams, J., Wittmann, F., Wöll, H., Yang, X., Adou Yao, C.Y., Yap, S.L., Yoneda, T., Zahawi, R.A., Zakaria, R., Zang, R., de Assis, R.L., Garcia Luize, B., Venticinque, E.M., 2015. An estimate of the number of tropical tree species. *Proc. Natl. Acad. Sci.* 112, 7472–7477. <https://doi.org/10.1073/pnas.1423147112>
- Stavros, E.N., Schimel, D., Pavlick, R., Serbin, S., Swann, A., Duncanson, L., Fisher, J.B., Fassnacht, F., Ustin, S., Dubayah, R., others, 2017. ISS observations offer insights into plant function. *Nat. Ecol. Evol.* 1, s41559--017.
- Stysley, P.R., Coyle, D.B., Clarke, G.B., Frese, C.E., Blalock, G., Morey, P., Kay, R.B., Poulos, D., Hersh, M., 2016. Laser pulse production for NASA's global ecosystem dynamics investigation (GEDI) Lidar.
- Swatantran, A., Dubayah, R., Goetz, S., Hofton, M., Betts, M.G., Sun, M., Simard, M., Holmes, R., 2012. Mapping Migratory Bird Prevalence Using Remote Sensing Data Fusion. *PLoS One* 7, 11. <https://doi.org/10.1371/journal.pone.0028922>
- Tang, H., Armston, J.D., Hancock, S., Hofton, M., Blair, J.B., Dubayah, R.O., 2018. AfriSAR: Canopy Cover and Vertical Profile Metrics Derived from LVIS, Gabon, 2016. ORNL DAAC, Oak Ridge, Tennessee, USA.
- Tang, H., Dubayah, R., 2017. Light-driven growth in Amazon evergreen forests explained by seasonal variations of vertical canopy structure. *Proc. Natl. Acad. Sci.* 114, 2640–2644.
- Tang, H., Dubayah, R., Brolly, M., Ganguly, S., Zhang, G., 2014. Large-scale retrieval of leaf area index and vertical foliage profile from the spaceborne waveform lidar (GLAS/ICESat). *Remote Sens. Environ.* 154, 8–18. <https://doi.org/10.1016/j.rse.2014.08.007>
- Tang, H., Dubayah, R., Swatantran, A., Hofton, M., Sheldon, S., Clark, D.B., Blair, B., 2012. Retrieval of vertical LAI profiles over tropical rain forests using waveform lidar at La Selva, Costa Rica. *Remote Sens. Environ.* 124, 242–250. <https://doi.org/10.1016/j.rse.2012.05.005>
- Tang, H., Ganguly, S., Zhang, G., Hofton, M.A., Nelson, R.F., Dubayah, R., 2016. Characterizing leaf area index (LAI) and vertical foliage profile (VFP) over the United States. *Biogeosciences* 13, 239–252.
- Tarboton, D., Bras, R., Rodriguez-Iturbe, I., 1991. On the extraction of channel networks from digital elevation data. *Hydrol. Process.* 5, 81–100. <https://doi.org/10.1002/hyp.3360050107>
- Tattoni, C., Rizzolli, F., Pedrini, P., 2012. Can LiDAR data improve bird habitat

- suitability models? *Ecol. Modell.* 245, 103–110.
<https://doi.org/10.1016/j.ecolmodel.2012.03.020>
- Ter Steege, H., Pitman, N., Sabatier, D., Castellanos, H., Van Der Hout, P., Daly, D.C., Silveira, M., Phillips, O., Vasquez, R., Van Andel, T., others, 2003. A spatial model of tree α -diversity and tree density for the Amazon. *Biodivers. Conserv.* 12, 2255–2277.
- Ter Steege, H., Pitman, N.C.A., Killeen, T.J., Laurance, W.F., Peres, C.A., Guevara, J.E., Salomão, R.P., Castilho, C. V, Amaral, I.L., de Almeida Matos, F.D., others, 2015. Estimating the global conservation status of more than 15,000 Amazonian tree species. *Sci. Adv.* 1, e1500936.
- Tews, J., Brose, U., Grimm, V., Tielborger, K., Wichmann, M.C., Schwager, M., Jeltsch, F., 2004. Animal species diversity driven by habitat heterogeneity/diversity: the importance of keystone structures. *J. Biogeogr.* 31, 79–92.
- Tuck, S.L., O'Brien, M.J., Philipson, C.D., Saner, P., Tanadini, M., Dzulkifli, D., Godfray, H.C.J., Godoong, E., Nilus, R., Ong, R.C., Schmid, B., 2016. The value of biodiversity for the functioning of tropical forests: insurance effects during the first decade of the Sabah biodiversity experiment. *Proc. R. Soc. B Biol. Sci.* 283.
- Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steininger, M., 2003. Remote sensing for biodiversity science and conservation. *Trends Ecol. Evol.* 18, 306–314. [https://doi.org/10.1016/s0169-5347\(03\)00070-3](https://doi.org/10.1016/s0169-5347(03)00070-3)
- Ukizintambara, T., White, L., Abernethy, K., Thébaud, C., 2007. Gallery forests versus bosquets: conservation of natural fragments at Lope National Park in central Gabon. *Afr. J. Ecol.* 45, 476–482.
- Vierling, K.T., Bassler, C., Brandl, R., Vierling, L.A., Weiss, I., Muller, J., 2011. Spinning a laser web: predicting spider distributions using LiDAR. *Ecol. Appl.* 21, 577–588. <https://doi.org/10.1890/09-2155.1>
- Vitousek, P.M., Mooney, H.A., Lubchenco, J., Melillo, J.M., 1997. Human domination of Earth's ecosystems. *Science* (80-.). 277, 494–499.
- Vorosmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., Davies, P.M., 2010. Global threats to human water security and river biodiversity. *Nature* 467, 555–561. <https://doi.org/10.1038/nature09440>
- Walters, G., Ngagnia Ndjabounda, E., Ikabanga, D., Biteau, J.P., Hymas, O., White, L.J.T., Ndong Obiang, A.-M., Ndong Ondo, P., Jeffery, K.J., Lachenaud, O., Stévant, T., 2016. Peri-urban conservation in the Mondah forest of Libreville, Gabon: Red List assessments of endemic plant species, and avoiding protected

- area downsizing. *Oryx* 50, 419–430.
<https://doi.org/10.1017/S0030605315000204>
- Watson, J.E.M., Darling, E.S., Venter, O., Maron, M., Walston, J., Possingham, H.P., Dudley, N., Hockings, M., Barnes, M., Brooks, T.M., 2016. Bolder science needed now for protected areas 30, 243–248.
- Watson, J.E.M., Evans, T., Venter, O., Williams, B., Tulloch, A., Stewart, C., Thompson, I., Ray, J.C., Murray, K., Salazar, A., McAlpine, C., Potapov, P., Walston, J., Robinson, J.G., Painter, M., Wilkie, D., Filardi, C., Laurance, W.F., Houghton, R.A., Maxwell, S., Grantham, H., Samper, C., Wang, S., Laestadius, L., Runting, R.K., Silva-Chávez, G.A., Ervin, J., Lindenmayer, D., 2018. The exceptional value of intact forest ecosystems. *Nat. Ecol. Evol.* 2, 599.
<https://doi.org/10.1038/s41559-018-0490-x>
- White, L., 2001. Chapter 11: Forest-savanna dynamics and the origins of Marantaceae Forest in the Lope Reserve, Gabon, in: *African Rain Forest Ecology and Conservation: An Interdisciplinary Perspective*. pp. 165–182.
- White, L., Abernethy, K., 1997. *A Guide to the Vegetation of the Lopé Reserve*. New York Wildl. Conserv. Soc. viii, 224p.-illus., col. illus.. ISBN 963206427.
- Whitehurst, A.S., Swatantran, A., Blair, J.B., Hofton, M.A., Dubayah, R., 2013. Characterization of Canopy Layering in Forested Ecosystems Using Full Waveform Lidar. *Remote Sens.* 5, 2014–2036.
- Whittaker, R.J., Willis, K.J., Field, R., 2001. Scale and species richness: Towards a general, hierarchical theory of species diversity. *J. Biogeogr.*
<https://doi.org/10.1046/j.1365-2699.2001.00563.x>
- Wilsey, B.J., Potvin, C., 2000. Biodiversity and ecosystem functioning: importance of species evenness in an old field. *Ecol. Ecol. Soc. Am.* 81, 887–892.
- Wolf, J.A., Fricker, G.A., Meyer, V., Hubbell, S.P., Gillespie, T.W., Saatchi, S.S., 2012. Plant species richness is associated with canopy height and topography in a neotropical forest. *Remote Sens.* 4, 4010–4021.
<https://doi.org/10.3390/rs4124010>
- Wright, S.J., Muller-Landau, H.C., 2006. The future of tropical forest species. *Biotropica* 38, 287–301. <https://doi.org/10.1111/j.1744-7429.2006.00154.x>
- Zhao, K.G., Valle, D., Popescu, S., Zhang, X.S., Mallick, B., 2013. Hyperspectral remote sensing of plant biochemistry using Bayesian model averaging with variable and band selection. *Remote Sens. Environ.* 132, 102–119.
<https://doi.org/10.1016/j.rse.2012.12.026>
- Zimble, D.A., Evans, D.L., Carlson, G.C., Parker, R.C., Grado, S.C., Gerard, P.D., 2003. Characterizing vertical forest structure using small-footprint airborne

LiDAR. *Remote Sens. Environ.* 87, 171–182.