White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Leeds.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/162817/

Version: Accepted Version

Article:

Marselis, SM, Abernethy, K, Alonso, A et al. (2020) Evaluating the potential of full-waveform lidar for mapping pan-tropical tree species richness. Global Ecology and Biogeography, 29 (10). pp. 1799-1816. ISSN: 1466-822X

https://doi.org/10.1111/geb.13158

© 2020 John Wiley & Sons Ltd. This is the peer reviewed version of the following article: Marselis, SM, Abernethy, K, Alonso, A, et al. Evaluating the potential of full-waveform lidar for mapping pan-tropical tree species richness. Global Ecol Biogeogr. 2020; 00: 1– 18. https://doi.org/10.1111/geb.13158 , which has been published in final form at https://doi.org/10.1111/geb.13158. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Title

Evaluating the Potential of Full-waveform Lidar for Mapping Pan-Tropical Tree Species Richness

Short title

Lidar and Pan-Tropical Tree Species Richness

Abstract

Aim:

Mapping tree species richness across the tropics is of great interest for effective conservation and biodiversity management to help prevent species extinction. In this study, we evaluated 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.

Location:

Tropics

Time period:

Present

Major taxa studied:

Trees

Methods:

First, we evaluated the characteristics of the vertical canopy structure across 15 study sites using (simulated) full-waveform lidar data and related these findings to in-situ tree species information. Then, we developed structure-richness models at the local (within 25-50 ha plots), regional (biogeographic regions), and pan-tropical scale at three spatial resolutions (1.0, 0.25 and 0.0625 ha) using Poisson regression.

Results:

The results showed a weak structure-richness relationship at the local scale. At the regional scale (within a biogeographical region) a stronger relationship was found between canopy structure and tree species richness across different tropical forest types, for example across Central Africa and in South America (R^2 ranging from 0.44-0.56, RMSD ranging between 23-61%). A weaker relationship was found at the pan-tropical scale, including data across four continents (R^2 = 0.39 and RMSE = 43%, 0. 25 ha resolution).

Main Conclusions:

Our results may serve as a basis for future development of a set of structure-richness models to map high resolution tree species richness using vertical canopy structure information from the Global Ecosystem Dynamics Investigation (GEDI). The value of this effort would be enhanced by access to a larger set of field reference data for all tropical regions. 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.

Keywords

Biodiversity, canopy structure, GEDI, lidar, plant area index, tropical forests

1. Introduction

Tropical forests are known for their high tree species diversity. Current estimates suggest in the order of 15,000 tree species in Amazonia alone, in contrast to 124 tree species in temperate forests in Europe, and more than 40,000 different tree species across the tropical region (Slik *et al.*, 2015; Ter Steege *et al.*, 2015). High levels of tree species richness are essential for maximizing the provision of essential ecosystem services (Liang *et al.*, 2016). 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. 82% of the remaining forest is estimated to have experienced some degree of human impact (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 approximately 25% of the tree species are threatened with extinction (Ter Steege *et al.*, 2015). The Convention of Biological Diversity (CBD) and Group on Earth Observations Biodiversity Observation Network (GEO BON) 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 the identified needs is the mapping of taxonomic diversity at high spatial resolution over large scales (Pereira *et al.*, 2010). Here we focus on tree species diversity. The collection of tree species diversity data is traditionally done in the field, and such information has previously been used to create predictive maps of tree species richness across the globe at low spatial resolution (Kier *et al.*, 2005; Mutke & Barthlott, 2005). More recently, passive remote sensing data, such as optical imagery from different airborne and spaceborne platforms, has been used in combination with field reference data to predict tree species diversity in different regions (Foody & Cutler, 2006; Carlson *et al.*, 2007; Féret & Asner, 2014; Rocchini *et al.*, 2016; Schäfer *et al.*, 2016; Bongalov *et al.*, 2019). Even though such methods have been progressively developing over the last decade, they are not yet operational for mapping tree species richness across the tropics due to, among others, a lack of consistent remote sensing and training data over such scales, insufficient model accuracy and/or low spatial resolution.

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). In this study we focus on the use of active remote sensing, specifically lidar, for mapping taxonomic tree species richness in the tropics. While local tropical forest diversity is largely independent of biomass (Sullivan *et al.*, 2017), it remains unclear if substantial amounts of variation in species diversity are associated with other features of forest structure. Here, we explore for the first time whether small-scale vertical canopy structure variation is significantly associated with the spatial variation in tropical tree species richness. On a global scale it has previously been shown that canopy height explains a limited portion of the variation in tree species diversity, as such data provides information on the available niche space (Gatti *et al.*, 2017). It has since been hypothesized that including information on the vertical canopy structure, must explain more of the variation in tree species diversity than canopy height alone; as such data provide information on the occupation of the vertical niche space. Marselis *et al.*, (2019) demonstrated that information on canopy height and vertical canopy structure, expressed as the Plant Area Index (PAI) profile from full-waveform airborne lidar data, could be used to map tree species diversity in Gabon, Africa. However, it is not clear whether this relationship is of similar nature and strength across different regions, or even the entire tropics. If existent, than the use of such a structure-diversity relationship(s) could 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.

Factors influencing tree species diversity on a global scale differ from those affecting spatial patterns at regional or local scales. 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 large-grain scales historical biogeography processes are more important, whereas at the plot-scale environmental variables strongly influence diversity (Keil & Chase, 2019).

Similar to species diversity, forest structure at the global scale is influenced by interacting 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 measured in the field is mainly comprised of four variables: canopy height, biomass, basal area and tree density (Palace *et al.*, 2015). However, active remote sensing techniques have revolutionized the study of canopy structure (Newnham *et al.*, 2015). With lidar remote sensing, for example, it is now possible to obtain information on canopy height, as well as 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 possess vertical structure information which is inherently linked to ecological processes.

We hypothesize that structure-diversity relationships will vary across different biogeographical and phylogenetic regions (Corlett & Primack, 2011; Slik *et al.*, 2018) and that it may be more fruitful to develop multiple relationships rather than one pan-tropical relationship for operationalizing tree species diversity mapping with spaceborne active remote sensing data. Additionally, the strength of the relationship between a variable and tree species diversity often changes with resolution (plot size) as tree species diversity is not linearly related with area (species-area curve) (MacArthur & Wilson, 1967). This complicates the development of predictive models at specific resolutions, and also limits the extrapolation of estimates at one resolution to a larger area, which impedes the mapping of pan-tropical tree species diversity at high spatial resolution.

In sum, we know that both species diversity and canopy structure vary greatly within and across continents. Hence, our objective is to assess whether canopy structure information can explain tree species richness at the local, regional and/or global scale with the ultimate goal to evaluate the efficacy of spaceborne full-waveform lidar for mapping tree species richness across the tropics. First, we compare characteristics of the vertical canopy structure, measured with full-waveform lidar data, of tropical forests across the world. Second, we evaluate the differences in species richness and species-area curves across the different study sites using field measurements. Third, we evaluate the potential for developing local (within 25-50 ha field plots), regional (within biogeographical regions) and pan-tropical structure-richness relationships, relating canopy structure metrics from lidar to tree species richness measurements from the field at three spatial resolutions (0.0625, 0.25 and 1.0 ha). Lastly, we discuss the potential of full-waveform lidar data from GEDI for mapping tree species richness across the tropics using structure-richness relationships.

2. Materials and Methods

We address the relationship between canopy structure and tree species richness in *terra firme* forest in the tropical region between 23.5° N & S. We compiled a comprehensive field and lidar dataset covering colonizing forest, old-growth tropical forest and forests under different degrees of degradation and savanna. We included such a wide variety of forest stages as most of the Earth's tropical forests have been degraded or otherwise affected by natural and human influences (Lewis *et al.*, 2015). Hence, when developing a method that allows for estimating pan-tropical tree species richness it is important to include data from across this range of possibilities. Species diversity can be expressed with many different metrics. Generally, three levels of diversity are recognized: α, β, and γ diversity. α diversity refers to the local diversity of a community, habitat or field plot. β diversity refers to the differences in diversity between habitats and γ diversity to the total diversity of a region (Colwell, 2009). In this study we focus on α diversity. α diversity can be expressed with many different indicators. In this study we focus on species richness (S) expressed as the total number of species in a plot of a given size. Hence, from here on forward we only refer to tree species richness, used to express the local tree species diversity.

2.1 Field Datasets

Field data were used to calculate the reference values of tree species richness. We used 15 datasets: one from Australia, two from South-East Asia, six from Africa, three from South America and three from Central America (Figure 1). All field datasets used in this study have been previously collected and published and have coincident airborne lidar data available. Each field dataset is labeled with a three-letter code and contained information on tree location, species and diameter at breast height (DBH). All datasets were collected by different organizations and research teams resulting in different data characteristics (Table 1, SI1). 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.

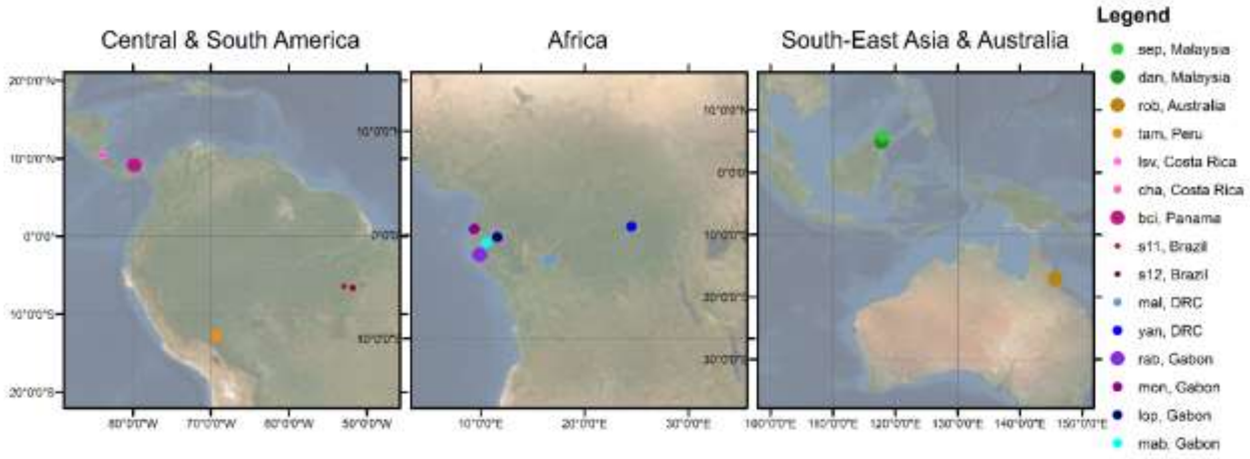


Figure 1: Location of field sites across the three continents, colors of each study site are consistent throughout the paper. Gridlines indicate 10° intervals in longitudinal and latitudinal directions. The size of the place markers represents the size of the total sampled area relative to each other.

Table 1: 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 are published and/or a publication in which the data are described further.

Country	**Project code**	**No. native plots**	**Total area (ha)**	**Source / Additional Information**
		Oceania		
Australia	*rob*	1	25	(Bradford *et al.*, 2014)
		South-East Asia		
Malaysia	*dan*	1	50	https://forestgeo.si.edu/sites/asia/danum-valley
Malaysia	*sep*	9	36	https://www.forestplots.net/en/ (Jucker *et al.*, 2018)
		Africa		
DRC	*mal*	21	21	(Bastin *et al.*, 2015)
DRC	*yan*	9	9	(Kearsley *et al.*, 2013)
Gabon	*rab*	1	25	https://forestgeo.si.edu/sites/africa/rabi (Memiaghe *et al.*, 2016)
Gabon	*lop*	11	9.5	https://www.forestplots.net/en/ (Labrière *et al.*, 2018)
Gabon	*mon*	10	10	(Fatoyinbo *et al.*, 2017)
Gabon	*mab*	10	10	(Bastin *et al.*, 2015; Labrière *et al.*, 2018)
		South America		
Peru	*tam*	6	6	https://www.forestplots.net/en/ (Boyd *et al.*, 2013)
Brazil	*s11*	9	1.44	http://www.paisagenslidar.cnptia.embrapa.br/webgis/
Brazil	*s12*	19	4.8	http://www.paisagenslidar.cnptia.embrapa.br/webgis/
		Central America		
Costa Rica	*lsv*	12	6	https://tropicalstudies.org/carbono-project/
Costa Rica	*cha*	3	2	
Panama	*bci*	1	50	https://forestgeo.si.edu/sites/neotropics/barro-colorado-island (Lobo & Dalling, 2013)

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 the number of tree species (S) and sampled area. We selected squares of 1.0 ha (100x100 m) because they are often-used in ecology and it has been shown that the spatial mismatch of plot location and remote sensing products is minimized at this resolution (Réjou-Méchain *et al.*, 2014). We used squares of 0.25 ha (50x50 m) because these yielded the best results describing the structure-diversity relationship in Gabon (Marselis *et al.*, 2019), and squares of 0.0625 ha (25x25 m) because they correspond to a resolution close to the GEDI footprint size. The datasets were used at one, two or three of the aforementioned resolutions depending on the original plot size and the availability of stem maps or subplots (**Error! Reference source not found.**, full table in SI1). For each of the field sites we calculated S for the entire dataset and for each plot at each plot size (Table 2). Only live trees with a DBH ≥ 10 cm were included, to ensure consistency among datasets and we removed all plots of each resolution in which more than 20% of the trees were not identified to the genus level.

Table 2: The total number of species identified at each study site and the average ($\bar{x}$) and standard deviation (s) of the species richness for each of the three plot sizes expressed as $\bar{x} \pm s$ (including only live trees with DBH ≥ 10 cm).

Country	**Project Name**	**Total No. species**	**Total sampled area used (ha)**	**Species richness 1.0 ha**	**Species richness 0.25 ha**	**Species richness 0.0625 ha**
			Oceania			
Australia	*rob*	205	25	98 ± 10	56 ± 8	27 ± 5
			South-East Asia			
Malaysia	*dan*	430	2	117 ± 13	51 ± 7	19 ± 4
Malaysia	*sep*	517	32	102 ± 22	53 ± 11	-
			Africa			
DRC	*mal*	116	21	37 ± 11	20 ± 7	-
DRC	*yan*	232	9	50 ± 23	24 ± 13	10 ± 6
Gabon	*rab*	234	25	84 ± 8	42 ± 6	17 ± 4
Gabon	*lop*	118	9.5	32 ± 22	17 ± 10	8 ± 4
Gabon	*mon*	146	10	32 ± 15	15 ± 9	7 ± 5
Gabon	*mab*	196	10	55 ± 8	-	-
			South America			
Peru	*tam*	517	6	171 ± 13	70 ± 9	24 ± 5
Brazil	*s11*	91	1.44	-	-	17 ± 3
Brazil	*s12*	135	4.8	-	-	16 ± 4
			Central America			
Costa Rica	*lsv*	216	6	-	48 ± 8	19 ± 5
Costa Rica	*cha*	81	2	58	28 ± 5	13 ± 4
Panama	*bci*	220	50	87 ± 8	42 ± 6	17 ± 3

2.2 Lidar Datasets

Each of the field datasets 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 field data collection. We used the GEDI simulator (Hancock *et al.*, 2019) to create lidar waveforms from the ALS data over the field plots. The ALS data was originally collected with a variety of airborne instruments, but the GEDI simulator ensures a reliable GEDI-like waveform with minimal influence of the original instrument-specific characteristics. 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 (**Error! Reference source not found.**). In this way, a reliable measure of canopy structure could be acquired for each plot by averaging lidar metrics from all waveforms inside 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; Tang *et al.*, 2012; Marselis *et al.*, 2018; Hancock *et al.*, 2019). The 1 m vertical profile was used to compare the canopy structure across the study sites. It was aggregated into a 10 m vertical profile, summing all PAI values in each 10 m vertical bin, to be used in the structure-richness analyses. We chose to use the PAI profile because it is a biophysical variable describing the amount of plant material along the vertical forest axis, thus directly indicating the occupation of vertical niche space, and Marselis *et al.*, (2019) previously showed this information relates well to tree species richness in Africa. The average of each of the resulting metrics from all waveforms within each plot was computed to represent the canopy structure for each plot at each spatial resolution.

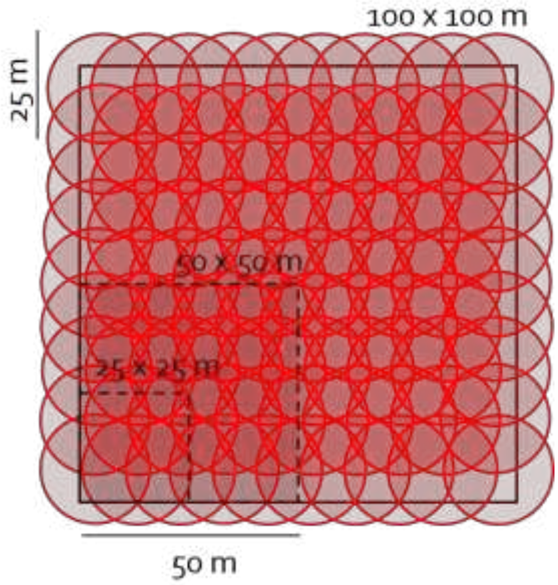


Figure 2: Illustration of simulated lidar waveform layout. The waveforms (red circles) have a Gaussian energy distribution with σ=5.5 m, resulting in a roughly 22 m diameter footprint. Example of simulated footprint distribution locations in a 1.0 (solid outline), 0.25 and 0.0625 ha field plot (dashed outline).

2.3 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 the PAI values for each 1 m vertical bin), using all simulated lidar waveforms for each study site. In addition 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 across the study site.

2.4 Species-area relationships across the tropics

We created species-area relationships, calculating the mean and standard deviation of S for plot sizes ranging between 0.01 and 50 ha, to assess how species richness changes by plot size across our study sites. Each of the original field plots was filled with as many non-overlapping subplots as possible at 17 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) with each tree assigned to a subplot at each resolution. The plot sizes used at each study site depended on the original plot size and the availability of stem maps (SI1). We visualized the mean and standard deviation of S for each plot size at each study site to evaluate the differences in species-area curves across the tropics.

2.5 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 & Chase, 2019) as do different ecosystem functions (Chisholm *et al.*, 2013). Definitions of the scales are presented in the following sections.

2.5.1 Local Analysis

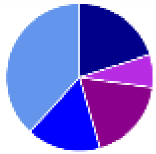
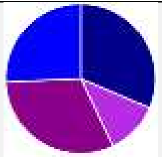
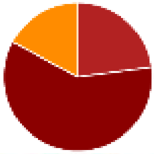
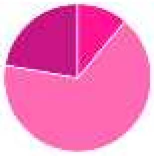
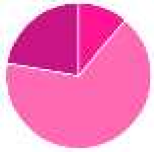
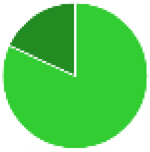
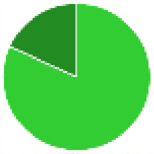
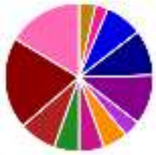
The local analysis focused on the structure-richness relationship within large (25 or 50 ha) plots. We used data from adjacent field plots to evaluate the relationship between S and the canopy structure expressed as canopy height (RH98), total PAI and vertical canopy profile (PAI at 10 m vertical intervals). 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 suitable for this analysis because the species identification was incomplete at the time of analysis (**Error! Reference source not found.**, **Error! Reference source not found.**). We related the canopy structure with S using a generalized linear model with a Poisson error distribution. 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 Bayesian Information Criterion (BIC) score, and then constructed the predictive model based on the same training data. We evaluated model performance using R^2, Root Mean Squared Difference as a percentage of the mean (RMSD%) and bias based on the predictions for the test data (Piñeiro *et al.*, 2008). The average and 95% confidence interval of these metrics were recorded for each study site at each resolution.

2.5.2 Regional and Pan-tropical Analysis

The regional analysis was focused on the structure-richness relationship based on non-adjacent plots across study sites within the same biogeographical zone. We evaluated different combinations of study sites at three spatial resolutions (**Error! Reference source not found.**). To prevent the large plots from dominating the regional and pan-tropical analyses, we thinned their contribution to both the regional and pan-tropical datasets. From the 25 ha plots we selected 1.0 ha plots at each corner, and from the 50 ha plots we selected all corner and the middle plots along the long sides of the plot (6 1.0 ha plots total). To avoid mixing 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 randomly sampled one plot at the given resolution from each original plot location (especially important at the 0.25 and 0.0625 ha resolutions at which up to 16 plots exist at the location of each original 1.0 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: first we performed variable selection on the Poisson regression model choosing the model with lowest BIC (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 according to Piñeiro *et al.* (2008).

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. Thus applying the same methods as for the regional analysis.

Table 3: 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 (values in the table indicate total number of plots for each region) and resolution 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**	
Africa (Gabon & DRC)	62		56		36	
South America (Brazil & Peru)					36	
Central America			27		27	
South-East Asia	11		11			
Pan-tropical	89		101		105	

Legend ■ rob ■ sep ■ cha ■ mal ■ yan ■ mab ■ lop ■ mon ■ rab ■ dan ■ bci ■ tam ■ s11 ■ s12 ■ lsv ■ cha

3. Results

3.1 Vertical forest structure across the tropics

The vertical canopy structure of forests, in terms of the vertical distribution of plant material varies between tropical regions (**Error! Reference source not found.**). Maximum canopy height in our study sites in the Neotropics and Central Africa is typically around 40 m, and slightly lower in Australia, while canopy heights in South-East Asia exceed 60 m. Many sites show a distinct understory layer and a decrease in plant material through the canopy. Relative to the understory, the canopy layer sharply declines in vegetation density (*sep* and *dan*, Malaysia) or steadily declines along the vertical axis (*bci*, Panama; *rab*, Gabon; *mal*, DRC; *rob*, Australia). This vertical distribution of declining vegetation is exacerbated in degraded forests: in *s11, s12* (Brazil) and *mon* (Gabon), where the bulk of the vegetation exists close to the forest floor at ~5 m height, but remnant trees in some plots may reach 40 m. Other sites, especially undisturbed ones, have distinct canopy layers. In *tam* (Peru) and in the old-growth forest in *lsv* (Costa Rica) there are multiple peaks of high-density vegetation across the vertical strata of the forest. The profiles of *yan* (DRC) and *lop* (Gabon) are characterized by a multiple-peak pattern, with one peak 20-30 m in the canopy and another within 5 m of the ground, reflecting the inherent structure of the forest-savanna mosaic. The less disturbed *mab* (Gabon) forest shows high variability in canopy structure between plots (e.g. the wide shaded area in Figure 3) and higher canopies.

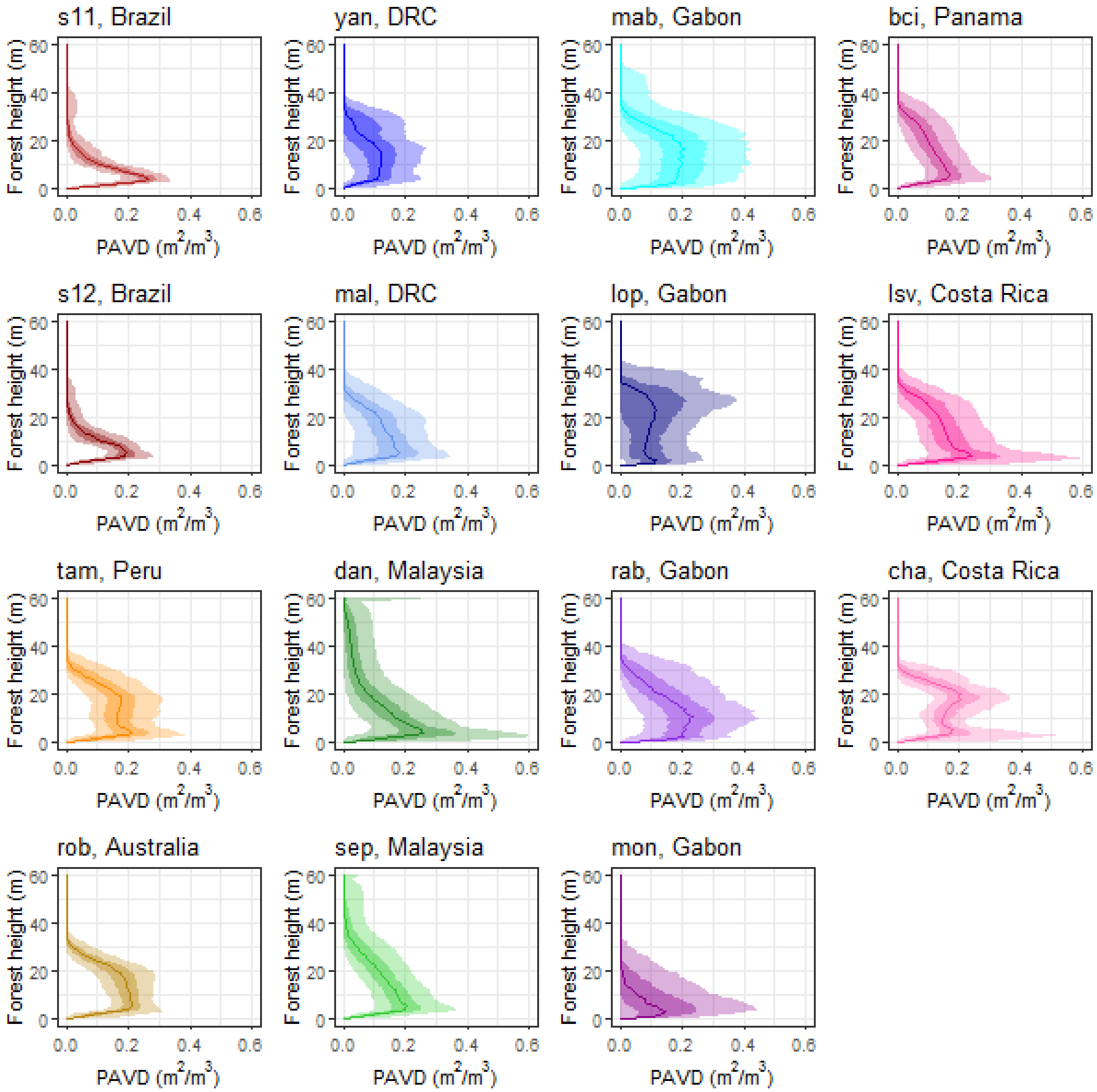


Figure 3: Canopy structure expressed as the Plant Area Volume Density profile (PAVD), expressing the Plant Area Index for each 1 m vertical bin, displayed as the median of all plots within each study site (solid line), the 30th-70th percentile (darker shaded area) and 10th-90th percentile (lighter shaded area).

3.2 Species-area relationships

The number of species increases with plot size, but the rate of increase varies across study sites (**Error! Reference source not found.**). For example, in *rob* (Australia) 82-117 species occur in a 1.0 ha plot compared to 16-44 species in 0.0625 ha plots. By contrast, *tam* (Peru) contains 154-185 species/ha, but only 11-35 species in a 0.0625 ha plot, similar to *rob*. Thus, species' composition of adjacent 0.0625 ha plots in *tam* must be more different from each other than adjacent 0.0625 ha plots in *rob* (Australia), in other words, the β diversity of the plots in *tam* is higher than in *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 (**Error! Reference source not found.**). Curves that are initially steep and decrease in slope at larger plot sizes indicate a high α diversity but a lower β diversity (e.g. when the area is increased, the same species are encountered).

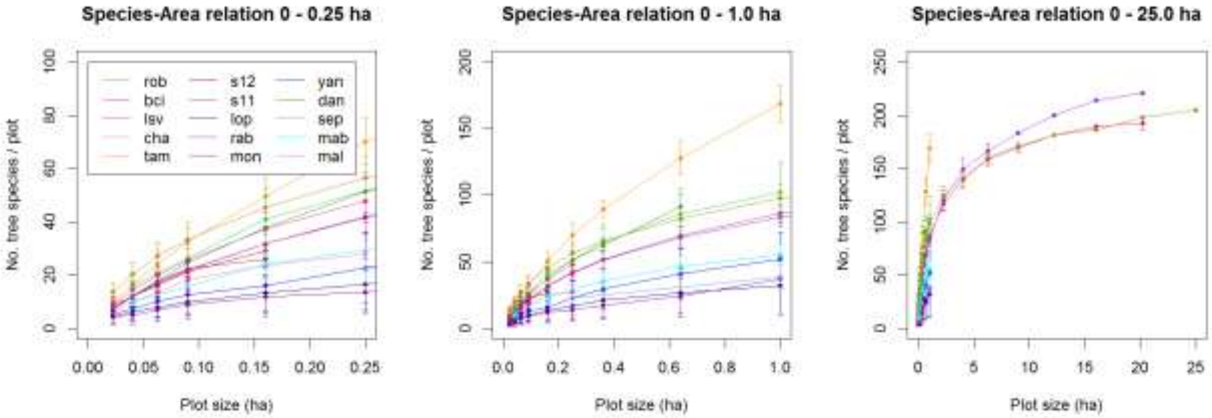


Figure 4: Relationships between tree species richness and area for each study site (note the change in y-axis across panels from left to right).

3.3 Structure-richness relationships

Pulling together the information on tree species richness and canopy structure (RH98 and Total PAI), species richness generally increases with increasing canopy height and increasing total Plant Area Index across the tropics (Figure 5).

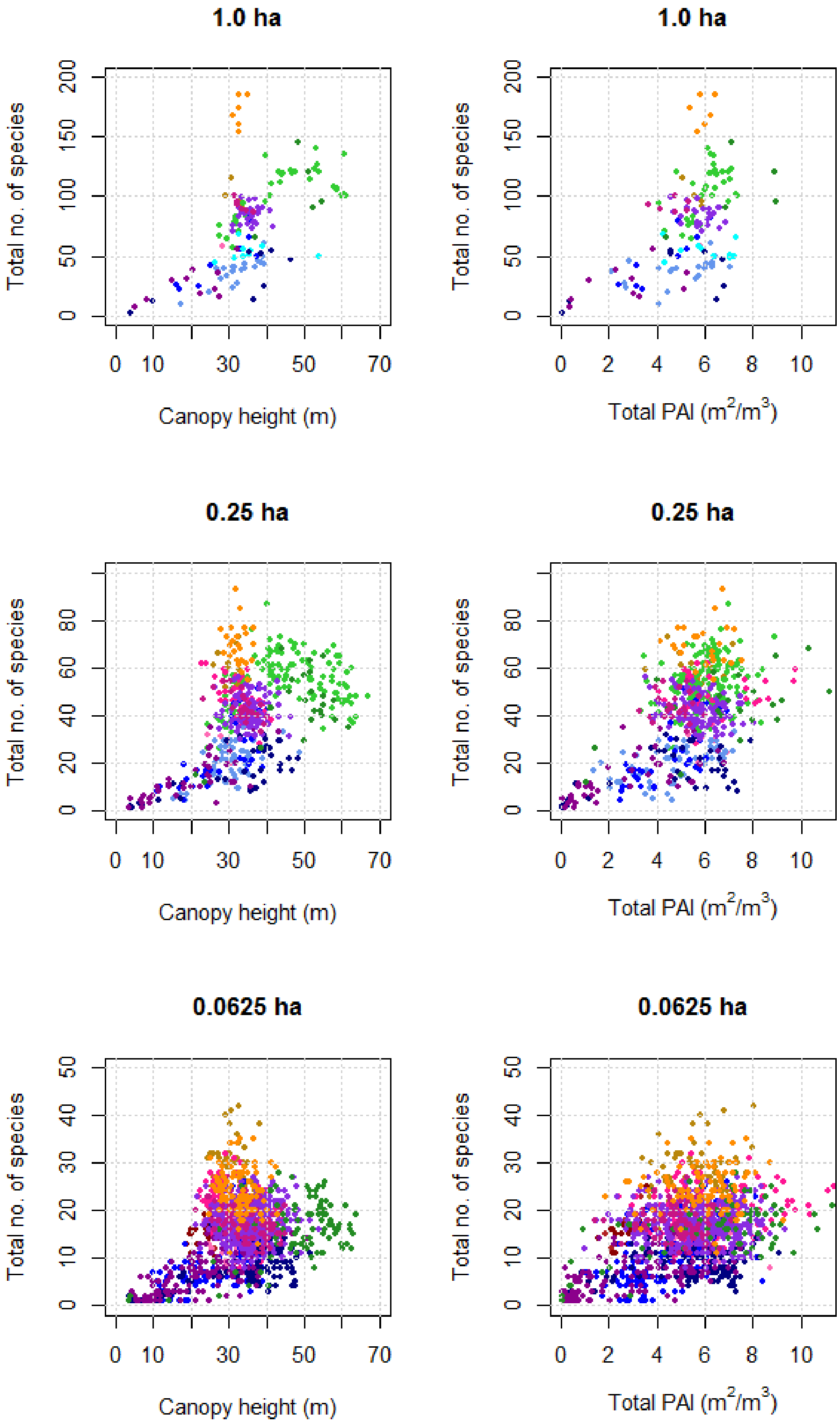


Figure 5: Relation between canopy height (left) and total PAI (right) across three spatial scales for all study sites across the tropics. Each point represents one plot at the specific resolution. Dots are colored by study site corresponding to legend in Figure 1.

The cross-validation results of the local models reveal weak structure-richness relationships. Of the three large plots (25 and 50 ha), only the models for *bci* (50 ha) show evidence of a significant relationship between the predicted and observed values (R^2=0.32 at 1.0 ha, SI2). Even though species richness within all three large plots can be predicted with a root mean squared error between 7-20% of the mean species richness, the low RMSD% found only indicates that the predictions at the local scale are close to the mean species richness, however in *rab* and *rob* the canopy structure is insensitive to the local variation in tree species richness (see example figure in SI2).

Regional structure-richness models generally show much better performance (Figure 6) than the local models in terms of the variance in species richness that can be explained with the canopy structure information (mostly significant models and 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 Africa and South America (Table 3) show the best model performance whereas regions including the Costa Rica datasets show much poorer performance (regions indicated with *centralamerica*). Results from an additional analysis on the compositional similarity (Bray-Curtis; Faith *et al.*, 1987, SI3) of the Costa Rica dataset showed that, even though species richness varies in Costa Rica (**Error! Reference source not found.**), the plots share many species, i.e. the composition is similar. In the *africa* and *southamerica* datasets the variation in species richness is accompanied by a much larger variation in species composition (SI3). The variation of the model performance for *seasia* is very high to the low number of plots available for this region and at the 0.25 ha resolution it was not possible to create a significant model >95% of the monte-carlo iterations (Table 3).

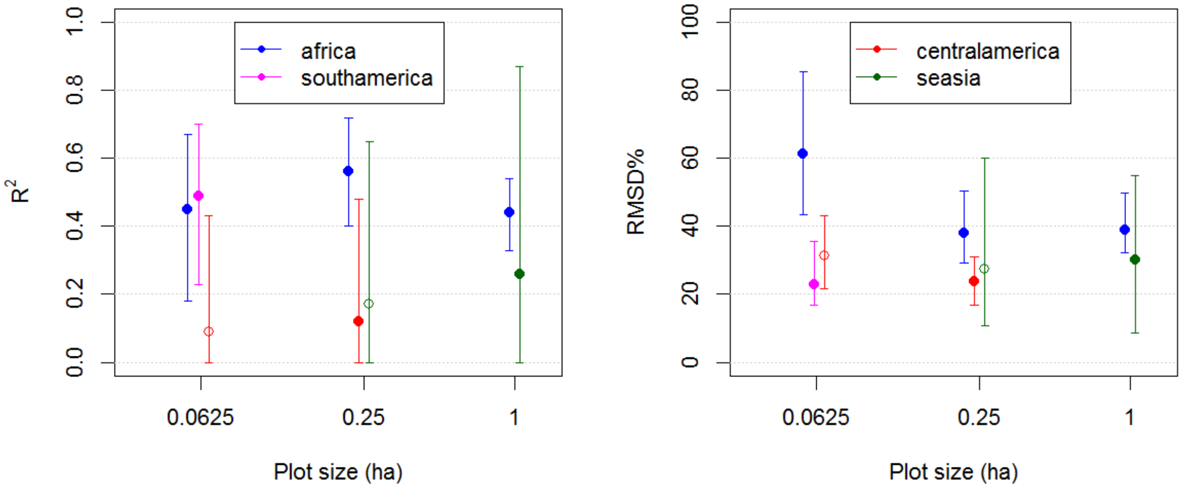


Figure 6: Cross-validated model performance of regional structure-richness models. Error bars indicate the 95% range of values for each performance metric. Solid dots indicate >95% of the generated models was statistically significant, open circles indicate a lower percentage was significant.

Pan-tropical structure-richness models show more similar performance across all spatial resolutions with mean R^2 ranging between 0.25 and 0.39 and RMSD% between 66 and 43%, for the plot sizes from 1.0 and 0.0625 ha (Figure 8), indicating that around 39% of the variation in tree species richness can be explained using canopy structure metrics alone at the 0.25 ha resolution at the global scale. Sites with extremely high values of observed species richness are generally predicted poorly (SI4).

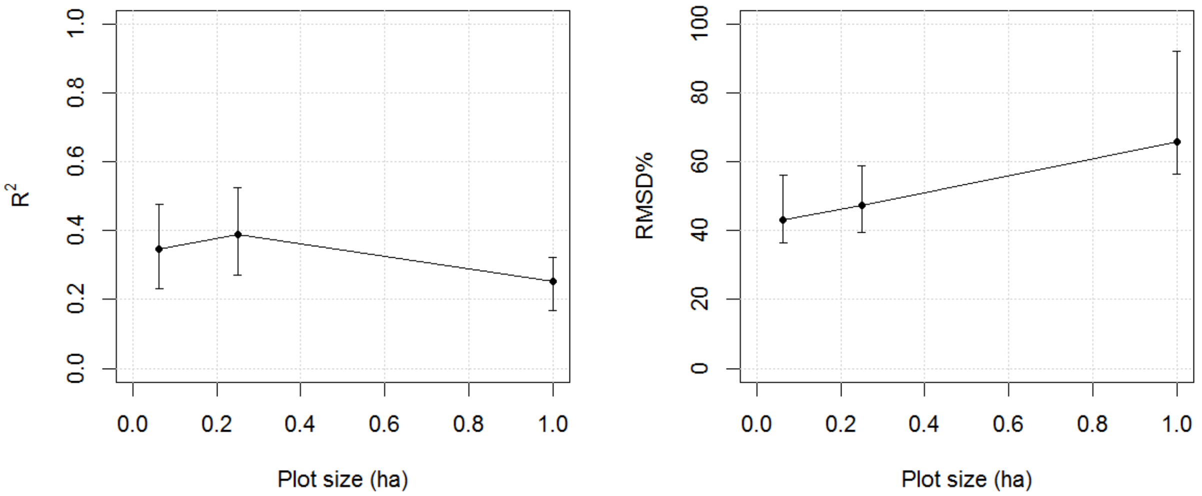


Figure 7: Cross-validated model performance at the pan-tropical scale in terms of R^2 *and RMSD%. Error bars indicate the range between which 95% of the performance values of the cross-validated models fall.*

4. Discussion

4.1 Structure-richness relationships across scales

In this study we explored the relationships between vertical canopy structure and tree species richness at different resolutions across local, regional and pan-tropical scales. We found weak relationships between canopy structure and tree species richness at the local scale and the strongest relationship at the regional scales in Africa and South America. We also found significant relationships between canopy structure and tree species richness combining the data from all study sites across the tropics.

At the local scale, within one big plot inside one forest type, the variation in the canopy structure is determined largely by variability in growth structure within the same species (the 25 and 50 ha plots have a similar composition throughout the plot, SI1). For example, an adult tree of species X may range in height from 20-40 m, so even though structure may differ between two plots of similar composition, the difference 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, due to the difference in canopy structure the model may predict a species richness of 2 based on variation in structure. On the other hand, as area increases 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 of two adjacent plots is more similar than the composition of two more distant plots. This is the case for *bci*, where a 50 ha area was sampled and included in the local analysis, which led to more successful prediction of species richness based on structure. Within the 25 ha plots sampled at *rab* and *rob*, the variation in composition is smaller and no significant structure-richness relationships were found (SI3).

Increasing the scale, we found that regions consisting of sites exhibiting a large variation in species composition among plots, but with a similar biogeographical history, show a much stronger structure-richness relationship. However, we note that model performance differed quite drastically across regions. The forest in *lsv,* Costa Rica, consists of largely similar species composition, whereas species composition is much more different in regions where the structure-richness models perform better (South-America, Africa), supporting the result from local scale models that species richness can be better predicted from canopy structure in areas with greater β diversity.

At the pan-tropical scale we find a significant relationship between canopy structure and tree species richness across all spatial resolutions. At the intermediate resolution (0. 25 ha) this relationship appears to be slightly stronger than at the higher and lower resolutions, 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 1.0 ha resolution (**Error! Reference source not found.**), whereas at the 0.0625 ha resolution the species richness ranges between 11-35 species, which is still higher than most other sites but much less 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 pan-tropical model predictions for *tam* contain highly erroneous predictions that are not present in 0.0625 ha models (SI4). Rare species do not contribute much to the canopy structure, thereby complicating the relationship between structure and richness at a scale at which they contribute largely to species richness numbers.

4.2 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, the collection of such data is costly and time-consuming. Here, we were able to exploit 15 independently collected datasets (SI1). However, there is quite a data gap, especially in the Amazon basin, the mainland of South-East Asia, New Guinea and Australia. Apart from the spatial representation problem, the low number of plots for certain regions attributes largely to the observed variability in model performance. The pan-tropical models (with n ≥ 89) show much more stable performance than models of regions with low numbers of plots (e.g. *seasia*). A training dataset that does not fully represent the range of structure in the full dataset can lead to highly erroneous predictions for some of the test plots. Such errors are exacerbated by the logarithmic link model in Poisson regression because errors can increase exponentially. Even so, negative predictions are possible with linear regression and the risk of underestimating tree species richness is higher for diverse areas. Hence, we chose to use Poisson regression, knowing that it may lead to extreme predictions in some cases that should be accounted for when operationalizing this method.

Species diversity can be identified in many different ways (Gotelli & Colwell, 2001; Colwell, 2009) and there are risks and pitfalls using just one metric. In this study we only used ‘species richness’ (S), defined by the number of different tree species in a defined area (the plot, with different sizes), as this metric is easy to interpret and a prediction of the number of species/area can probably be used most directly by ecosystem managers. Hereby we did not control for the number of stems in the plot, nor for the abundancy of the different species. Such things can be taken into account for example by using the Shannon diversity index or rarefaction curves. Moreover, depending on the type of metric, a different model will need to be selected. For example, a generalized linear regression with a Poisson error distribution, as used here, is more suitable for estimated tree species richness as this is count data, whereas a linear model with a Gaussian error distribution will be better for estimating Shannon diversity. Hence, we chose to focus on one metric of diversity to test the structure-richness relationships, while acknowledging other metrics may provide better, worse, or more useful predictions of tree species diversity and these should be considered in the future.

This study serves as a first attempt to study the pan-tropical structure-richness relationship and should be improved and further developed when more data become available. Additionally, the characteristics of each dataset differed widely because all data were 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 less than 20% of unidentified trees at the genus level. Nonetheless, we acknowledge that the quality of the species identification varied and may have affected our models as species identification in the tropics can be challenging due to the vast variety of tree species and the fact that new species are still encountered. Species identification of new and existing data could be improved using more botanists or genetic tests in the lab, which has been done for some of the datasets used here, but is not yet feasible for all datasets.

The availability of stem maps and subplots in each study site determined the spatial resolutions at which datasets could be used. This resulted in the inclusion of different datasets for each region (Table 3). This makes the 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 sizes of the datasets used to build the structure-richness models. Hence, no conclusion can be drawn about the optimal resolution for the structure-richness relationships.

Accurate geolocation of field plots is key for the development of reliable species-richness models. However, geolocation of field plots in the tropical forest can be challenging due to difficulties receiving a reliable GPS signal under dense canopy. This should be taken into account, especially when evaluating the performance of models build with small field plots, where the effects of such geolocation errors will be larger (Réjou-Méchain *et al.*, 2014).

We included data from a range of forest stages, including old-growth forest, successional stages, disturbed forest and even low tree density savanna sites. The relationships we found are partially driven by this gradient (Figure 5). However, we deemed it essential to include data from across this range of forest types, because if this method is to be operationalized using canopy structure information from across the tropics, we will encounter all these different stages of forest (Lewis *et al.*, 2015).

4.3 Future research & Applications

Our results provide confidence regarding the existence of regional and pan-tropical structure-richness relationships that may be used to map pan-tropical tree species richness. The most accurate predictions seem to be achieved at the regional scale when adequate data are available and when forested areas are grouped by regions of similar biogeographical history. However, in the absence of such data it may be of more immediate interest to further develop pan-tropical models that can explain up to 39% of variation in tree species richness. At the time of writing, GEDI is collecting canopy structure information close to the finest resolution tested here (0.0625 ha) and thus these 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 km^2 resolution in which the GEDI data samples are averaged to 1 km^2 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 detect local nuances in species richness within forests of uniform composition. We suggest that the species richness predictions could potentially be used in a similar way as for gridded GEDI data products and estimate the average number of species/0.0625 ha within a 1 km^2 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 species richness predictions at 0.0625 ha resolution to the expected number of tree species in 1 km^2. Also, the amount of variance in species richness explained is limited. Therefore, we propose two future research avenues of interest: fusion with spectral and/or radar data and using an environmental framework. Both spectral data and radar data have previously been shown to predict some of the variance in tree species richness (Foody & Cutler, 2006; Wolf *et al.*, 2012; Schäfer *et al.*, 2016; Bae *et al.*, 2019; Bongalov *et al.*, 2019; Marselis *et al.*, 2019) and may improve our models and allow for more accurate predictions of tree species richness across the tropics and the creation of wall-to-wall data products at higher spatial resolution. Especially data from the hyperspectral HISUI (Matsunaga *et al.*, 2013) instrument, that is soon to be launched to the International Space Station, the radar BIOMASS mission (Le Toan *et al.*, 2011), or the TanDEM-X mission (Qi *et al.*, 2019), may be highly relevant for such applications. Alternatively, we believe that the inclusion of structural data within previously developed environment and biogeographical frameworks will help to predict tree species diversity (Keil & Chase, 2019). Such frameworks could benefit from GEDI lidar data providing information on the occupation of the vertical niche space and likely improve predictions of tree species richness across the tropics.

5. Conclusions

In this study we evaluated the existence of local, regional and pan-tropical 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, 39% of the variation in tree species richness can be explained with the vertical canopy structure using one single predictive model at a 0.25 ha plot size. A full set of regional structure-richness models will most likely aid accurate pan-tropical species richness mapping, but the development of such a set of models is contingent on the availability of sufficient coincident field & lidar 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.

References

Bae, S., Levick, S.R., Heidrich, L., Magdon, P., Leutner, B.F., Wollauer, S., Serebryanyk, A., Nauss, T., Krzystek, P., Gossner, M.M., Schall, P., Heibl, C., Bassler, C., Doerfler, I., Schulze, E.., Krah, F.., Culmsee, H., Jung, K., Heurich, M., Fischer, M., Seibold, S., Thorn, S., Gerlach, T., Hothorn, T., Weisser, W.W. & Muller, J. (2019) Radar vision in the mapping of forest biodiversity from space. *Nature Communications*, **10**, 4757.

Bastin, J.F., Barbier, N., Réjou-Méchain, M., Fayolle, A., Gourlet-Fleury, S., Maniatis, D., De Haulleville, T., Baya, F., Beeckman, H., Beina, D., Couteron, 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. *Scientific Reports*, **5**.

Bongalov, B., Burslem, D.F.R.P., Jucker, T., Thompson, S.E.D., Rosindell, J., Swinfield, T., Nilus, R., Clewley, D., Phillips, O.L. & Coomes, D.A. (2019) Reconciling the contribution of environmental and stochastic structuring of tropical forest diversity through the lens of imaging spectroscopy. *Ecology Letters*, **22**, 1608–1619.

Boyd, D.S., Hill, R.A., Hopkinson, C. & Baker, T.R. (2013) Landscape-scale forest disturbance regimes in southern Peruvian Amazonia. *Ecological Applications*, **23**, 1588–1602.

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. *Journal of Tropical Forest Science*, **26**, 543–553.

Carlson, K.M., Asner, G.P., Hughes, R.F., Ostertag, R. & Martin, R.E. (2007) Hyperspectral remote sensing of canopy biodiversity in Hawaiian lowland rainforests. *Ecosystems*, **10**, 536–549.

Chisholm, R.A., Muller-Landau, H.C., Abdul Rahman, K., Bebber, D.P., Bin, Y., Bohlman, S.A., Bourg, N.A., Brinks, J., Bunyavejchewin, S., Butt, N., Cao, H., Cao, M., Cárdenas, D., Chang, L.W., Chiang, J.M., Chuyong, G., Condit, R., Dattaraja, H.S., Davies, S., Duque, A., Fletcher, C., Gunatilleke, N., Gunatilleke, S., Hao, Z., Harrison, R.D., Howe, R., Hsieh, C.F., Hubbell, S.P., Itoh, A., Kenfack, D., Kiratiprayoon, S., Larson, A.J., Lian, J., Lin, D., Liu, H., Lutz, J.A., Ma, K., Malhi, Y., Mcmahon, S., Mcshea, W., Meegaskumbura, M., Mohd. Razman, S., Morecroft, M.D., Nytch, C.J., Oliveira, A., Parker, G.G., Pulla, S., Punchi-Manage, R., Romero-Saltos, H., Sang, W., Schurman, J., Su, S.H., Sukumar, R., Sun, I.F., Suresh, H.S., Tan, S., Thomas, D., Thomas, S., Thompson, J., Valencia, R., Wolf, A., Yap, S., Ye, W., Yuan, Z. & Zimmerman, J.K. (2013) Scale-dependent relationships between tree species richness and ecosystem function in forests. *Journal of Ecology*, **101**, 1214–1224.

Colwell, R.K. (2009) *Biodiversity: concepts, patterns and measurement*. *The Princeton guide to ecology*, pp. 257–263.

Condit, R. (2000) Spatial patterns in the distribution of tropical tree species. *Science*, **288**, 1414–1418.

Corlett, R.T. & Primack, R.B. (2011) *Tropical Rain Forests: An Ecological and Biogeographical Comparison: Second Edition*, 2nd edn. Blackwell Publishing.

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 Sensing of Environment*, **81**, 378–392.

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. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing*.

Faith, D.P., Minchin, P.R. & Belbin, L. (1987) Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, **69**, 57–68.

Fatoyinbo, T.E., Pinto, N., Simard, M., Armston, J., Duncanson, L., Hofton, M., Saatchi, S., Lavalle, 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4286–4287.

Féret, J.B. & Asner, G.P. (2014) Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy. *Ecological Applications*, **24**, 1289–1296.

Foody, G.M. & Cutler, M.E.J. (2006) Mapping the species richness and composition of tropical forests from remotely sensed data with neural networks. *Ecological Modelling*, **195**, 37–42.

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 Ecology*, **218**, 899–908.

Givnish, T.J. (1999) On the causes of gradients in tropical tree diversity. *Journal of Ecology*, **87**, 193–210.

Gotelli, N.J. & Colwell, R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.

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 and Space Science*, **6**, 294–310.

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.

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. *Nature Communications*, **4**.

Keil, P. & Chase, J.M. (2019) Global patterns and drivers of tree diversity integrated across a continuum of spatial grains. *Nature Ecology and Evolution*, **3**, 390.

Kier, G., Mutke, J., Dinerstein, E., Ricketts, T.H., Küper, W., Kreft, H. & Barthlott, W. (2005) Global patterns of plant diversity and floristic knowledge. *Journal of Biogeography*, **32**, 1107–1116.

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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1–11.

Lewis, S.L., Edwards, D.P. & Galbraith, D. (2015) Increasing human dominance of tropical forests. *Science*, **349**.

Liang, J., Crowther, T.W., Picard, N., Wiser, S., Zhou, M., Alberti, G., Schulze, E.D., McGuire, A.D., Bozzato, F., Pretzsch, H., De-Miguel, S., Paquette, A., Hérault, B., Scherer-Lorenzen, M., Barrett, C.B., Glick, H.B., Hengeveld, G.M., Nabuurs, G.J., Pfautsch, S., Viana, H., Vibrans, A.C., Ammer, C., Schall, P., Verbyla, D., Tchebakova, N., Fischer, M., Watson, J. V., Chen, H.Y.H., Lei, X., Schelhaas, M.J., Lu, H., Gianelle, D., Parfenova, E.I., Salas, C., Lee, E., Lee, B., Kim, H.S., Bruelheide, H., Coomes, D.A., Piotto, D., Sunderland, T., Schmid, B., Gourlet-Fleury, S., Sonké, B., Tavani, R., Zhu, J., Brandl, S., Vayreda, J., Kitahara, F., Searle, E.B., Neldner, V.J., Ngugi, M.R., Baraloto, C., Frizzera, L., Bałazy, R., Oleksyn, J., Zawiła-Niedźwiecki, T., Bouriaud, O., Bussotti, F., Finér, L., Jaroszewicz, B., Jucker, T., Valladares, F., Jagodzinski, A.M., Peri, P.L., Gonmadje, C., Marthy, W., O'Brien, T., Martin, E.H., Marshall, A.R., Rovero, F., Bitariho, R., Niklaus, P.A., Alvarez-Loayza, P., Chamuya, N., Valencia, R., Mortier, F., Wortel, V., Engone-Obiang, N.L., Ferreira, L. V., Odeke, D.E., Vasquez, R.M., Lewis, S.L. & Reich, P.B. (2016) Positive biodiversity-productivity relationship predominant in global forests. *Science*.

Lobo, E. & Dalling, J.W. (2013) Effects of topography, soil type and forest age on the frequency and size distribution of canopy gap disturbances in a tropical forest. *Biogeosciences*, **10**, 6769–6781.

MacArthur, R.H. & Wilson, E.O. (1967) *The theory of Island Biogeography*, Princeton University Press, Princeton and Oxford.

Marselis, S.M., Tang, H., Armston, J., Abernethy, K., Alonso, A., Barbier, N., Bissiengou, P., Jeffery, K., Kenfack, D., Labrière, N. & others (2019) Exploring the relation between remotely sensed vertical canopy structure and tree species diversity in Gabon. *Environmental Research Letters*, **14**.

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.

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)*. *International Geoscience and Remote Sensing Symposium (IGARSS)*,.

Memiaghe, 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**.

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. *Journal of Ecology*, **97**, 923–932.

Mutke, J. & Barthlott, W. (2005) Patterns of vascular plant diversity at continental to global scales. *Biologiske skrifter*, **55**, 521–531.

Newnham, G.J., Armston, J.D., Calders, K., Disney, M.I., Lovell, J.L., Schaaf, C.B., Strahler, A.H. & Danson, F.M. (2015) Terrestrial laser scanning for plot-scale forest measurement. *Current Forestry Reports*, **1**, 239–251.

Palace, M.W., Sullivan, F.B., Ducey, M.J., Treuhaft, 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 Sensing of Environment*, **161**, 1–11.

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*, **339**, 277–278.

Pereira, H.M., Leadley, P.W., Proenca, V., Alkemade, R., Scharlemann, J.P.W., Fernandez-Manjarres, J.F., Araujo, M.B., Balvanera, P., Biggs, R., Cheung, W.W.L., Chini, L., Cooper, H.D., Gilman, E.L., Guenette, S., Hurtt, G.C., Huntington, H.P., Mace, G.M., Oberdorff, T., Revenga, C., Rodrigues, P., Scholes, R.J., Sumaila, U.R. & Walpole, M. (2010) Scenarios for Global Biodiversity in the 21st Century. *Science*, **330**, 1496–1501.

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*, **344**.

Piñeiro, G., Perelman, S., Guerschman, J.P. & Paruelo, J.M. (2008) How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecological Modelling*, **216**, 316–322.

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 Sensing of Environment*, **232**.

Réjou-Méchain, M., Muller-Landau, H.C., Detto, M., Thomas, S.C., Le Toan, T., Saatchi, S.S., Barreto-Silva, J.S., Bourg, N.A., Bunyavejchewin, S., Butt, N., Brockelman, W.Y., Cao, M., Cárdenas, D., Chiang, J.M., Chuyong, G.B., Clay, K., Condit, R., Dattaraja, H.S., Davies, S.J., Duque, A., Esufali, S., Ewango, C., Fernando, R.H.S., Fletcher, C.D., N. Gunatilleke, I.A.U., Hao, Z., Harms, K.E., Hart, T.B., Hérault, B., Howe, R.W., Hubbell, S.P., Johnson, D.J., Kenfack, D., Larson, A.J., Lin, L., Lin, Y., Lutz, J.A., Makana, J.R., Malhi, Y., Marthews, T.R., Mcewan, R.W., Mcmahon, S.M., Mcshea, W.J., Muscarella, R., Nathalang, A., Noor, N.S.M., Nytch, C.J., Oliveira, A.A., Phillips, R.P., Pongpattananurak, N., Punchi-Manage, R., Salim, R., Schurman, J., Sukumar, R., Suresh, H.S., Suwanvecho, U., Thomas, D.W., Thompson, J., Uríarte, M., Valencia, R., Vicentini, A., Wolf, A.T., Yap, S., Yuan, Z., Zartman, C.E., Zimmerman, J.K. & Chave, J. (2014) Local spatial structure of forest biomass and its consequences for remote sensing of carbon stocks. *Biogeosciences*, **11**, 6827–6840.

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 Sensing in Ecology and Conservation*, **2**, 25–36.

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. *Ecological Indicators*, **64**, 49–58.

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. *Proceedings of the National Academy of Sciences*, **112**, 7472–7477.

Slik, J.W.F., Franklin, J., Arroyo-Rodríguez, V., Field, R., Aguilar, S., Aguirre, N., Ahumada, J., Aiba, S.I., Alves, L.F., Anitha, K., Avella, A., Mora, F., Aymard, G.A.C., Báez, S., Balvanera, P., Bastian, M.L., Bastin, J.F., Bellingham, P.J., Van Den Berg, E., Da Conceição Bispo, P., Boeckx, P., Boehning-Gaese, K., Bongers, F., Boyle, B., Brambach, F., Brearley, F.Q., Brown, S., Chai, S.L., Chazdon, R.L., Chen, S., Chhang, P., Chuyong, G., Ewango, C., Coronado, I.M., Cristóbal-Azkarate, J., Culmsee, H., Damas, K., Dattaraja, H.S., Davidar, P., DeWalt, S.J., DIn, H., Drake, D.R., Duque, A., Durigan, G., Eichhorn, K., Eler, E.S., Enoki, T., Ensslin, A., Fandohan, A.B., Farwig, N., Feeley, K.J., Fischer, M., Forshed, O., Garcia, Q.S., Garkoti, S.C., Gillespie, T.W., Gillet, J.F., Gonmadje, C., Granzow-De La Cerda, I., Griffith, D.M., Grogan, J., Hakeem, K.R., Harris, D.J., Harrison, R.D., Hector, A., Hemp, A., Homeier, J., Hussain, M.S., Ibarra-Manríquez, G., Hanum, I.F., Imai, N., Jansen, P.A., Joly, C.A., Joseph, S., Kartawinata, K., Kearsley, E., Kelly, D.L., Kessler, M., Killeen, T.J., Kooyman, R.M., Laumonier, Y., Laurance, S.G., Laurance, W.F., Lawes, M.J., Letcher, S.G., Lindsell, J., Lovett, J., Lozada, J., Lu, X., Lykke, A.M., Bin Mahmud, K., Mahayani, N.P.Di., Mansor, A., Marshall, A.R., Martin, E.H., Matos, D.C.L., Meave, J.A., Melo, F.P.L., Mendoza, Z.H.A., Metali, F., Medjibe, V.P., Metzger, J.P., Metzker, T., Mohandass, D., Munguía-Rosas, M.A., Muñoz, R., Nurtjahya, E., De Oliveira, E.L., Onrizal, Parolin, P., Parren, M., Parthasarathy, N., Paudel, E., Perez, R., Pérez-García, E.A., Pommer, U., Poorter, L., Qi, L., Piedade, M.T.F., Pinto, J.R.R., Poulsen, A.D., Poulsen, J.R., Powers, J.S., Prasad, R.C., Puyravaud, J.P., Rangel, O., Reitsma, J., Rocha, Di.S.B., Rolim, S., Rovero, F., Rozak, A., Ruokolainen, K., Rutishauser, E., Rutten, G., Mohd Said, M.N., Saiter, F.Z., Saner, P., Santos, B., Dos Santos, J.R., Sarker, S.K., Schmitt, C.B., Schoengart, J., Schulze, M., Sheil, D., Sist, P., Souza, A.F., Spironello, W.R., Sposito, T., Steinmetz, R., Stevart, T., Suganuma, M.S., Sukri, R., Sultana, A., Sukumar, R., Sunderland, T., Supriyadi, Suresh, H.S., Suzuki, E., Tabarelli, M., Tang, J., Tanner, E.V.J., Targhetta, N., Theilade, I., Thomas, D., Timberlake, J., De Morisson Valeriano, M., Van Valkenburg, J., Van Do, T., Van Sam, H., Vandermeer, J.H., Verbeeck, H., Vetaas, O.R., Adekunle, V., Vieira, S.A., Webb, C.O., Webb, E.L., Whitfeld, T., Wich, S., Williams, J., Wiser, S., Wittmann, F., Yang, X., Yao, C.Y.A., Yap, S.L., Zahawi, R.A., Zakaria, R. & Zang, R. (2018) Phylogenetic classification of the world's tropical forests. *Proceedings of the National Academy of Sciences of the United States of America*, **115**, 1837–1842.

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. *Science advances*, **1**, e1500936.

Sullivan, M.J.P., Talbot, J., Lewis, S.L., Phillips, O.L., Qie, L., Begne, S.K., Chave, J., Cuni-Sanchez, A., Hubau, W., Lopez-Gonzalez, G., Miles, L., Monteagudo-Mendoza, A., Sonké, B., Sunderland, T., Ter Steege, H., White, L.J.T., Affum-Baffoe, K., Aiba, S.I., De Almeida, E.C., De Oliveira, E.A., Alvarez-Loayza, P., Dávila, E.Á., Andrade, A., Aragão, L.E.O.C., Ashton, P., Aymard, G.A., Baker, T.R., Balinga, M., Banin, L.F., Baraloto, C., Bastin, J.F., Berry, N., Bogaert, J., Bonal, D., Bongers, F., Brienen, R., Camargo, J.L.C., Cerón, C., Moscoso, V.C., Chezeaux, E., Clark, C.J., Pacheco, Á.C., Comiskey, J.A., Valverde, F.C., Coronado, E.N.H., Dargie, G., Davies, S.J., De Canniere, C., Djuikouo, M.N., Doucet, J.L., Erwin, T.L., Espejo, J.S., Ewango, C.E.N., Fauset, S., Feldpausch, T.R., Herrera, R., Gilpin, M., Gloor, E., Hall, J.S., Harris, D.J., Hart, T.B., Kartawinata, K., Kho, L.K., Kitayama, K., Laurance, S.G.W., Laurance, W.F., Leal, M.E., Lovejoy, T., Lovett, J.C., Lukasu, F.M., Makana, J.R., Malhi, Y., Maracahipes, L., Marimon, B.S., Junior, B.H.M., Marshall, A.R., Morandi, P.S., Mukendi, J.T., Mukinzi, J., Nilus, R., Vargas, P.N., Camacho, N.C.P., Pardo, G., Peña-Claros, M., Pétronelli, P., Pickavance, G.C., Poulsen, A.D., Poulsen, J.R., Primack, R.B., Priyadi, H., Quesada, C.A., Reitsma, J., Réjou-Méchain, M., Restrepo, Z., Rutishauser, E., Salim, K.A., Salomão, R.P., Samsoedin, I., Sheil, D., Sierra, R., Silveira, M., Slik, J.W.F., Steel, L., Taedoumg, H., Tan, S., Terborgh, J.W., Thomas, S.C., Toledo, M., Umunay, P.M., Gamarra, L.V., Vieira, I.C.G., Vos, V.A., Wang, O., Willcock, S. & Zemagho, L. (2017) Diversity and carbon storage across the tropical forest biome. *Scientific Reports*, **7**, 39102.

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 Sensing of Environment*, **124**, 242–250.

Le Toan, T., Quegan, S., Davidson, M.W.J., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S., Shugart, H. & Ulander, L. (2011) The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment*.

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. *Conservation Biology*, **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. *Nature Ecology and Evolution*, **2**, 599.

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 Sensing*, **4**, 4010–4021.

Data Availability Statement

Some of the field and lidar data used in this study can be downloaded directly from the internet. We have grouped the data in three groups here: (i) LVIS lidar data, (ii) ALS lidar data and (iii) field data. All datasets not mentioned in this statement were previously collected but have not been made publicly available and were accessed through personal collaboration with the data providers.

(i) LVIS lidar data

The LVIS data for the *rab, lop, mon* and *mab* study sites can be downloaded from the NASA data archive at the following DOI: https://doi.org/10.3334/ORNLDAAC/1591.

The LVIS data for the *cha* and *lsv* study sites is available on the following website: https://lvis.gsfc.nasa.gov/Data/Maps/CR2005Map.html.

(ii) ALS lidar data

The ALS data over *rob* is available through the auscover data portal ftp://qld.auscover.org.au/airborne_validation/lidar/robsons_creek/.

The ALS data over *s11* and *s12* can be downloaded from the sustainable landscapes data portal http://www.paisagenslidar.cnptia.embrapa.br/webgis/.

(iii) Field data

Field data from *rob* has been published through the Terrestrial Ecosystem Research Network (TERN) data portal linked from https://supersites.tern.org.au/supersites/fnqr-robson.

The *dan* and *rab* field data are all available through the Forestgeo website at

https://forestgeo.si.edu/sites/asia/danum-valley, https://forestgeo.si.edu/sites/africa/rabi and https://forestgeo.si.edu/sites/neotropics/barro-colorado-island.

The *sep, lop* and *tam* field data are all available through forestplots.net and can be found under the project names ‘sepilok’, ‘lope’ and ‘tambopata’ at https://www.forestplots.net/en/.

The *mon* field data is archived through the NASA data archiving center and available at DOI: https://doi.org/10.3334/ORNLDAAC/1580.

The *s11* and *s12* were available throught the data portals of the sustainable landscapes projects and can be found under the field data from the São Félix do Xingu region collected in 2011 and 2012 in the following data portal: http://www.paisagenslidar.cnptia.embrapa.br/webgis/.