Check for updates

A Journal of

RESEARCH PAPER

Evaluating the potential of full-waveform lidar for mapping pan-tropical tree species richness

Global Ecology

Suzanne M. Marselis¹ (b) | Katharine Abernethy^{2,3} | Alfonso Alonso⁴ (b) | John Armston¹ | Timothy R. Baker⁵ | Jean-Francois Bastin⁶ (b) | Jan Bogaert⁷ | Doreen S. Boyd⁸ | Pascal Boeckx⁹ | David F. R. P. Burslem¹⁰ | Robin Chazdon¹¹ | David B. Clark¹² | David Coomes¹³ | Laura Duncanson¹ | Steven Hancock¹⁴ | Ross Hill¹⁵ | Chris Hopkinson¹⁶ | Elizabeth Kearsley⁶ (b) | James R. Kellner^{17,18} | David Kenfack¹⁹ | Nicolas Labrière²⁰ | Simon L. Lewis²¹ | David Minor¹ | Hervé Memiaghe³ | Abel Monteagudo²² | Reuben Nilus²³ | Michael O'Brien²⁴ (b) | Oliver L. Phillips⁵ | John Poulsen²⁵ (b) | Hao Tang¹ (b) | Hans Verbeeck⁶ | Ralph Dubayah¹

¹Department of Geographical Sciences, University of Maryland, College Park,, Maryland, USA

Revised: 14 June 2020

²Division of Biological and Environmental Sciences Tropical Ecology and Conservation, University of Stirling, Stirling, UK

³Institut de Recherche en Ecologie Tropicale (IRET), CENAREST, Libreville, Gabon

⁴Center for Conservation and Sustainability, Smithsonian Conservation Biology Institute, Washington, District of Columbia, USA

⁵School of Geography, University of Leeds, Leeds, UK

⁶CAVELab, Computational and Applied Vegetation Ecology, Department of Environment, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium

⁷Université de Liège, Gembloux Agro-Bio Tech, Liege, Belgium

⁸School of Geography, University of Nottingham, Nottingham, UK

⁹Isotope Bioscience Laboratory - ISOFYS, Ghent University, Ghent, Belgium

¹⁰School of Biological Sciences, University of Aberdeen, Aberdeen, UK

¹¹Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut, USA

¹²Department of Biology, University of Missouri-St. Louis, St. Louis, Missouri, USA

¹³Department of Plant Sciences, University of Cambridge, Cambridge, UK

¹⁴School of Geosciences, University of Edinburgh, Edinburgh, UK

¹⁵Department of Life and Environmental Science, Bournemouth University, Poole, UK

¹⁶Department of Geography, University of Lethbridge, Lethbridge, Alberta, Canada

¹⁷Institute at Brown for Environment and Society, Brown University, Providence, Rhode Island, USA

¹⁸Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island, USA

¹⁹Center for Tropical Forest Science—Forest Global Earth Observatory, Smithsonian Tropical Research Institute, Smithsonian Institution, Washington, District of Columbia, USA

²⁰Laboratoire Evolution et Diversité Biologique, Université Toulouse III Paul Sabatier, Toulouse, France

 $^{21}\mbox{Department}$ of Geography, University College London, London, UK

²²Pasco, Jardín Botánico de Missouri, Oxapampa, Peru

²³Sabah Forestry Department, Forest Research Centre, Sandakan, Sabah, Malaysia

²⁴Área de Biodiversidad y Conservación, Universidad Rey Juan Carlos, Móstoles, Madrid, Spain

²⁵Nicolas School of the Environment, Duke University, Durham, North Carolina, USA

Correspondence

Suzanne M. Marselis, University of Maryland, 1150 Lefrak Hall, 6903 Preinkert Drive, College Park, Maryland, 20740, USA. Email: suzannemarselis@hotmail.com

Abstract

Aim: Mapping tree species richness across the tropics is of great interest for effective

Funding information

Natural Environment Research Council, Grant/Award Number: NE/B503384/1, NE/N012542/1 BIO-RED and NER/ A/S/2000/01002; Gordon and Betty Moore Foundation, Grant/Award Number: 1656; National Science Foundation, Grant/ Award Number: 0939907, DEB-0424767, DEB-0639393 and DEB-1147429; NASA Headquarters, Grant/Award Number: 80NSSC17K0321, 80NSSC18K0708 and NNL15AA03C; European Research Council, Grant/Award Number: ERC Advanced Grant 291585 - 'T-FORCES'

Global Ecology

Editor: Brian McGill

conservation and biodiversity management. 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.

A Journal o

Time period: Present.

Major taxa studied: Trees.

Methods: First, we evaluated the characteristics of vertical canopy structure across 15 study sites using (simulated) large-footprint full-waveform lidar data (22 m diameter) 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 (biogeographical 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 between canopy structure and tree species richness across different tropical forest types was found, for example across Central Africa and in South America [R^2 ranging from .44–.56, root mean squared difference as a percentage of the mean (RMSD%) ranging between 23–61%]. Modelling the relationship pan-tropically, across four continents, 39% of the variation in tree species richness could be explained with canopy structure alone (R^2 = .39 and RMSD% = 43%, 0.25-ha resolution).

Main conclusions: Our results may serve as a basis for the 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 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 may contribute to 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. Eighty-two percent of the remaining forest is estimated to have experienced some degree of human impact (Watson et al., 2018). 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 various airborne and spaceborne platforms, have been used in combination with field reference data to predict tree species diversity in different regions (Bongalov et al., 2019; Carlson, Asner, Hughes, Ostertag, & Martin, 2007; Féret & Asner, 2014; Foody & Cutler, 2006; Rocchini et al., 2016; Schäfer, Heiskanen, Heikinheimo, & Pellikka, 2016). Even though such methods have been developing progressively 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 (Lewis, Edwards, & Galbraith, 2015; 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 in intact forests (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 provide information on the available niche space (Gatti, Di Paola, Bombelli, Noce, & Valentini, 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 a similar nature and strength across different regions, or even the entire tropics. If existent, then, the use of such a structure-diversity relationship(s) could be applied 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, Blair, et al., 2020). 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 biogeographical 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 that 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 pan-tropical 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, for 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 and S. We compiled a field and lidar data set 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 covering this range of possibilities.

Global Ecology

Species diversity can be expressed with a variety of indicators. 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 metrics. In this study, we focus on one dimension of species diversity: species richness (*S*) expressed as the total number of species in a plot of a given size. From here on forward, we only refer to tree species richness, used to express the local tree species diversity. We chose species richness as it is easy to interpret, and it can probably be used most directly by ecosystem managers. This measure of species diversity is sometimes referred to as species density as it does not consider the number of trees sampled in each plot.

2.1 | Field data sets

Field data were used to calculate the reference values of tree species richness. We used 15 data sets: one from Australia, two from Southeast Asia, six from Africa, three from South America and three from Central America (Figure 1). All field data sets used in this study have been previously collected and published and have coincident airborne lidar data available. Each field data set is labelled with a three-letter code and contains information on tree location, species, and diameter at breast height (DBH). All data sets were collected by different organizations and research teams resulting in different data characteristics (Table 1, Supporting Information Appendix S1). Four data sets consisted of one large plot of 25 ha (*rob*, Australia and *rab*, Gabon) or 50 ha (*dan*, Malaysia and *bci*, Panama). The other 11 data sets consisted of multiple (3–21) smaller plots with sizes ranging from 0.16 to 4.0 ha.

In this study, we assessed the structure-richness relationship at three spatial resolutions (1.0, 0.25, 0.0625 ha) because of the nonlinear relationship between the number of tree species (*S*) and sampled area. We selected squares of 1.0 ha (100 m \times 100 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 (50 m ×50 m) because these yielded the best results describing the structure-diversity relationship in Gabon (Marselis et al., 2019), and squares of 0.0625 ha (25 m × 25 m) because they correspond to a resolution close to the GEDI footprint size. The data sets 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 (Table 1, full table in Supporting Information Appendix S1). For each of the field sites, we calculated *S* for the entire data set and for each plot at each plot size (Table 2). Only live trees with a DBH \ge 10 cm were included, to ensure consistency among data sets, and we included all plots of each resolution in which more than 80% of the trees were identified to at least the genus level.

2.2 | Lidar data sets

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 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 were 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 consistently 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 centre locations 6.25 m from the plot edge and 5 m between footprint centre locations (Figure 2). This allowed a reliable measure of canopy structure to be acquired for each plot by averaging lidar metrics from all waveforms inside



FIGURE 1 Location of field sites across the three continents; colours 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

the plot, instead of using single waveforms in the plot centre and evaluating structure-richness relationships based on such potentially 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, Dubayah, Knox, Clark, & Blair, 2002; Hancock et al., 2019; Marselis et al., 2018; Tang et al., 2012). 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 space. 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

	Oceania				
	Australia	rob	1	25	Bradford, Metcalfe, Ford, Liddell, and McKeown (2014)
	Southeast 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); Lopez- Gonzalez, Lewis, Burkitt, and Phillips (2011); Lopez-Gonzalez, Lewis, Burkitt, Baker, and Phillips (2009)
	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, Lutz, Korte, Alonso, and Kenfack (2016); Engone Obiang et al. (2019)
	Gabon	lop	11	9.5	https://www.forestplots.net/ en/; Labrière et al. (2018)
	Gabon	mon	12	12	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, Hill, Hopkinson, and Baker (2013)
	Brazil	s11	8	1.44	http://www.paisagenslidar. cnptia.embrapa.br/webgis/
	Brazil	s12	21	3.36	http://www.paisagenslidar. cnptia.embrapa.br/webgis/
	Central America				
	Costa Rica	lsv	18	9	https://tropicalstudies.org/carbo no-project/; Clark and Clark (2000)
	Costa Rica	cha	3	2	http://neoselvas.wordpress. uconn.edu/costa-rica/
	Panama	bci	1	50	https://forestgeo.si.edu/sites/ neotropics/barro-colorado- island; Lobo and Dalling (2013)

Global Ecology and Biogeography

Total

area (ha)

No. native

plots

Project

code

Country

TABLE 1Information on the originalplot size, the amount of total area sampledin the field and the source of the data,which is either a website where the dataare published and/or a publication inwhich the data are described further

A Journal of

Source/additional information

TABLE 2 The total number of species identified at each study site and the average (\bar{x}) and standard deviation (*SD*) of the species richness for each of the three plot sizes expressed as $\bar{x} \pm SD$ [including only live trees with diameter at breast height (DBH) \ge 10 cm]

Country	Project code	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 <u>±</u> 5
Southeast Asia						
Malaysia	dan	260	6	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	12	32 ± 15	15 ± 9	7 ± 5
Gabon	mab	196	10	55 <u>+</u> 8	-	-
South America						
Peru	tam	517	6	171 ± 13	70 ± 9	24 ± 5
Brazil	s11	91	1.44	-	-	17 ± 3
Brazil	s12	135	3.36	-	-	16 ± 4
Central America						
Costa Rica	lsv	216	9	-	48 ± 8	19 ± 5
Costa Rica	cha	81	2	58	28 ± 5	13 ± 4
Panama	bci	220	50	87 <u>±</u> 8	42 ± 6	17 ± 3

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

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 10th, 30th, 70th and 90th percentiles of the PAI values in the same 1-m vertical bins, to provide a representative distribution of the canopy structure across each 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 (Supporting Information Appendix S1). 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 relationships 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



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). Note: this footprint distribution was chosen to accurately depict canopy structure within the 0.0625, 0.25 and 1.0 ha plots but it does not represent the spatial distribution of spaceborne Global Ecosystem Dynamics Investigation (GEDI) waveforms

(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 (Table 1). We related the canopy structure with *S* using a generalized linear model with a Poisson error distribution. We used fivefold 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, Perelman, Guerschman, & Paruelo, 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 (Table 3). To prevent the large plots from dominating the regional and pan-tropical analyses, we thinned their contribution to both the regional and pan-tropical data sets. From the 25-ha plots, we selected 1.0-ha plots at each corner, and from the 50-ha plots,

Global Ecology and Biogeography

-WILEY^{⊥⊥}

we selected all corner and the middle plots along the long sides of the plot (six 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 data set. 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 $n \le 25$) 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* (McLeod, Xu, & Lai, 2020) package in R Core Team (2020)), 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 were included. Thus, the same methods were applied as in the regional analysis.

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 (Figure 3). 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 Southeast Asia exceed 60 m. Many sites show a distinct understorey layer and a decrease in plant material through the canopy. Relative to the understorey, 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 c. 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 Isv (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).

3.2 | Species-area relationships

The number of species increases with plot size, but the rate of increase varies across study sites (Figure 4). For example, in *rob*

WI	LE	Y—	Glo and	bal E Biog	colo eogr	gy aphy				м	A Journal of acroecology	-				
	Total	62	56	36	I	I	35	I	27	27	11	11	I	06	104	108
	mab	10												10		
	dol	œ	11	11										8	11	11
	иош	10	11	12										10	11	12
	rab	4	4	4										4	4	4
	yan	6	6	6										6	6	6
	mal	21	21											21	21	
	s12						21									21
	s11						00									ø
	tam						9							9	9	9
	bci								9	9				9	9	9
	cha								ო	ო				1	ო	ო
	lsv								18	18					18	18
	rob													4	4	4
sct code	dan										2	2		2	2	9
Proje	dəs										6	6		6	6	
	Resolution (ha)	1	0.25	0.0625	1	0.25	0.0625	1	0.25	0.0625	1	0.25	0.0625	1	0.25	0.0625
	Region	Africa			South America			Central America			South-East Asia			Pan-tropical		



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)

(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 dissimilar 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 (Figure 4). 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).

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

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 ($R^2 = .32$ at 1.0 ha, Supporting Information Appendix S2). Even though species richness within all three large plots can be predicted with a root mean squared error between 7% and 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 for example Supporting Information Appendix 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 data sets. Regions of Africa and South America (Table 3) show the best model performance whereas regions including the Costa Rica data sets show much poorer performance (regions indicated with centralamerica). Results from an additional analysis on the compositional similarity (Bray-Curtis; Faith, Minchin, & Belbin, 1987; Supporting Information Appendix S3) of the Costa Rica data set showed that, even though species richness varies in Costa Rica (Table 2), the plots share many species, that is, the composition is similar. In the africa and southamerica data sets the variation in species richness is accompanied by a much larger variation in species composition (Supporting Information Appendix S3). The variation of the model performance for seasia is very high because of 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 for > 95% of the Monte-Carlo iterations (Table 3). The model performance does not provide clear results on the effect of the different resolutions, given the overlapping error bars for models in the same region at

multiple resolutions and the inability to create each regional model at each spatial resolution (Figure 6).

Pan-tropical structure-richness models show varying performance across the spatial resolutions with mean R^2 ranging between .25 and .39 and RMSD% between 66% and 43% for the plot sizes from 1.0 and 0.0625 ha (Figure 7). However, the error bars of the model performance at different resolutions are overlapping, indicating that no resolution has a statistically better performance. 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 pan-tropical scale. Sites with extremely high values of observed species richness are generally predicted poorly (Supporting Information Appendix S4).

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, using a total of 15 study sites with coincident lidar and field data across the tropics. 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 large plot inside one forest type, the variation in the canopy structure is determined mostly by variability in growth structure within the same species (the 25 and 50 ha plots have a similar composition throughout the plot, Supporting Information Appendix S1 and S3). For example, an adult tree of species X may range in height from 20–40 m, so even though the canopy structure may differ between two plots of similar composition, the difference is not attributed to a difference in species composition. **FIGURE 5** Relationship between canopy height (left) and total plant area index (PAI; right) across three spatial scales for all study sites across the tropics. Each point represents one plot at the specific resolution. Dots are coloured by study site corresponding according to legend in Figure 1





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 contrary, as area increases it is more likely that the difference in structure is caused by a difference in composition. Do keep in mind that structure can also change due to other variables such as topography, soil and microclimate. Individuals of most tropical forest species are spatially aggregated (Condit et al., 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 with a species richness gradient was sampled (Fricker, Wolf, Saatchi, & Gillespie, 2015) 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 (Supporting Information Appendix S3).

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 structurerichness relationship. However, we note that model performance



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. RMSD% = root mean squared difference as a percentage of the mean

differed quite drastically across regions. The forest in *lsv*, Costa Rica, consists of largely similar species composition, whereas species composition is much more varied 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 (Table 2), whereas at the 0.0625-ha resolution the species richness ranges between 11 and 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 1.0ha pan-tropical model predictions for *tam* contain highly erroneous predictions that are not present in 0.0625-ha models (Supporting Information Appendix 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



FIGURE 7 Cross-validated model performance at the pan-tropical scale in terms of R^2 and root mean squared difference as a percentage of the mean (RMSD%). Error bars indicate the range between which 95% of the performance values of the cross-validated models fall

time-consuming. Here, we were able to exploit 15 independently collected data sets (Supporting Information Appendix S1), but data gaps exist, especially in the Amazon basin, high biomass forests of Central Africa, the mainland of Southeast Asia, New Guinea and Australia as well as the dry tropics and montane ecosystems. Apart from the spatial representation problem, the low number of plots for certain regions likely influences the observed variability in model performance. The pan-tropical models (with $n \ge 90$) show more stable performance than models of regions with low numbers of plots (e.g. seasia). A training data set that does not fully represent the range of structure in the full data set can lead to biased 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 (Colwell, 2009; Gotelli & Colwell, 2001) 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 abundance of the different species. Such information can be considered, for example, by using the Shannon diversity index or rarefaction curves. Moreover, depending on the type of metric, a different model may need to be selected to describe the structure-richness relationship as different metrics are related differently to canopy structure information. For example, a generalized linear regression with a Poisson error distribution, as used here, is more suitable for estimated tree species richness values as these are count data, whereas a linear model with a Gaussian error distribution will be better suited 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 data set differed widely because all data were collected by different researchers and institutions. We accounted for this as much as possible by using data sets 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.

-WILEY

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 data sets used here, but is not yet feasible for all data sets. Additionally, including information on species for trees with DBH \geq = 10 cm omits the (large) diversity found in the understorey. Fricker et al. (2015) showed that especially this diversity variation in small trees related well to the canopy structure. Future research should examine if these findings are consistent across the tropics.

The availability of stem maps and subplots in each study site determined the spatial resolutions at which data sets could be used. This resulted in the inclusion of different data sets 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 data sets 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 (Fricker et al., 2015). However, geolocation of field plots in tropical forests 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 built 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 oldgrowth 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). We acknowledge that climatic, edaphic and topographic variables could also impact tree species richness across the tropics, such as mean annual temperature and precipitation (Keil & Chase, 2019) and slope and elevation (Robinson et al., 2018). However, in this study, we specifically focused on the relationship between canopy structure and tree species diversity, in light of the recently launched GEDI mission. We recognize that including such information on topographic and environmental variables may further improve the mapping of tree species richness across the tropics.

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 WILEY-

Global Ecology

may be of more immediate interest to further develop pan-tropical models, as these have been shown here to 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 eight tracks with 600-m between-track spacing, 60-m along-track spacing (Figure 8). By the end of its nominal 2-year mission, GEDI will have sampled roughly 4% of total land area.

The footprint-level GEDI information on vertical canopy structure is stored in the Level-2 data products that are publicly available from the NASA Land Processes Distributed Active Archive Center (LPDAAC; https://lpdaac.usgs.gov; Dubayah, Hofton, et al., 2020; Dubayah, Luthcke, et al., 2020; Dubayah, Tang, et al., 2020). GEDI gridded data products will have a 1-km² or finer resolution (Dubayah, Blair, et al., 2020). 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 gridded GEDI data products by estimating 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 km². 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 (Bae et al., 2019; Bongalov et al., 2019; Foody & Cutler, 2006; Marselis et al., 2019; Schäfer et al., 2016; Wolf et al., 2012) 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 imager suite (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), the Ice, Cloud and land Elevation Satellite 2 mission (Duncanson et al., 2020), the TerraSAR-X add-on for Digital Elevation Measurement mission (Oi, Saarela, Armston, Stahl, & Dubayah, 2019) and Landsat (Saarela et al., 2018), may be highly relevant for such applications. Alternatively, we believe that the inclusion of structural data within previously developed environmental and biogeographical frameworks will help to predict tree species diversity (Keil & Chase, 2019) as such frameworks already display intrinsic differences in tree species diversity. 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, which could then be compared to existing predictions such as from Slik et al. (2015). Moreover, it has previously been shown that lidar data can provide interesting information about the diversity of other taxa as well (Huang, Swatantran, Dubayah, & Goetz, 2014; Rappaport, Royle, & Morton, 2020) and future avenues for using lidar data to provide information on a holistic measure of species diversity, including many taxa, could be of incredible value.



FIGURE 8 (a) Example of Global Ecosystem Dynamics Investigation (GEDI) data captured over the east of Mondah forest, north-west of Libreville, in Gabon, Africa. The lidar waveforms are collected along-track with eight tracks, a between-track spacing of 600 m and an along-track spacing of 60 m. (b) Example GEDI waveform with relative height metrics (shot number = 31151116800411054; orbit = 03115; track = 05633) at the location indicated with the blue box on (a). (c) The accompanying plant area index (PAI) profile at 5-m vertical intervals from the Level-2 data product. RHx = relative height at which x percentage of the energy has been returned.

5 | CONCLUSIONS

In this study, we evaluated the existence of local, regional and pantropical relationships between vertical canopy structure and tree species richness in the tropics at three spatial resolutions: 1.0, 0.25 and 0.0625 ha. Full-waveform lidar data provide detailed information on the differences in vertical canopy structure between forests across the tropics. Our results show that canopy structure can explain a significant percentage of variation in tree species richness across different biogeographical regions. 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 and lidar data across the tropics. Using one single predictive model at a pantropical scale, 39% of the variation in tree species richness could be explained using the vertical canopy structure. Given this canopy structure is measured directly from GEDI waveforms at the footprint level, this provides an interesting avenue for mapping tree species richness at high spatial resolution. Alternatively, canopy structure information from GEDI could be included in existing modelling frameworks, combining structural with spectral, environmental and topographic information to create more accurate tree species richness predictions.

ACKNOWLEDGMENTS

This work is supported by NASA Headquarters under the NASA Earth and Space Science Fellowship 414 Program - Grant 80NSSC17K0321; NASA contract #NNL 15AA03C to the University of Maryland for the development and execution of the GEDI mission (Principal Investigator, R. Dubayah); and the NASA New Investigator grant 80NSSC18K0708. We express our sincere gratitude to the following people and institutions for collecting field and lidar data and permitting us to use their data in this research: NASA's LVIS team, specifically Bryan Blair, Michelle Hofton and David Rabine for collecting airborne lidar data in lsv, cha, lop, mon, mab and rab, Gabon; Agence Nationale des Parcs Nationaux (ANPN) and Agence Gabonaise d'Etudes et d'Observation for logistical support that facilitated both fieldwork and lidar data collection in Gabon, specifically Kathryn Jeffery, Lee White, Flore Koumba Pambo, Josue Edzang Ndong and David Lehmann from ANPN; European Space Agency for funding field data collection in lop through the AfriSAR campaign, ANPN and the University of Stirling at the Station d'Etudes des Gorilles et Chimpanzes field station for hosting, and specifically Carl Ditougou, Pacôme Dimbonda, Arthur Dibambou, Edmond Dimoto, and Napo Milamizokou; NASA for funding field data collection in mon through the AfriSAR campaign and ANPN for hosting it; Nicolas Barbier, Missouri Botanical Garden (Tariq Stevart), Golder Associates, P. Ploton, V. Droissart and Y. Issembe, for field data collection in mab. Shell Gabon and the Smithsonian Tropical Research Institute for funding, and Pulcherie Bissiengou for guiding, field data collection in rab. This is contribution no. 196 of the Gabon Biodiversity Program. We thank Deborah Clark for her efforts in collecting field data in Isv. s11 and s12 field and lidar data sets were

-WILEY

acquired by the Sustainable Landscapes Brazil project supported by the Brazilian Agricultural Research Corporation (EMBRAPA), the US Forest Service, and United States Agency for International Development, and the US Department of State. Smithsonian Tropical Research Institute and the Smithsonian the Forest Global Earth Observatory (ForestGEO) Global Earth Observatory Network for funding and publishing field data collection in bci and J. W. Dalling for providing the lidar data in *bci*, which were funded through National Science Foundation (NSF) grant 0939907. 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. tam plot measurements have been supported by several grants including from Gordon and Betty Moore Foundation #1656 ('RAINFOR') to O. L. Phillips and National Geographic. We also thank the Jardín Botánico de Missouri (Peru) for their field data assistance. We kindly thank Bryan Mark and Horizons Peru for collecting and providing the lidar data over tam. sep plot measurements have been supported by several grants, including the European Research Council (ERC Advanced Grant 291585 - 'T-FORCES') and the Natural Environment Research Council (NER/A/S/2000/01002) grants to O. L. Phillips and special thanks go to Lan Qie. Data from RAINFOR, African Tropical Rainforest Observatory Network and tropical forests in the changing earth system (T-FORCES) are curated by ForestPlots.net, a cyber-infrastructure initiative hosted at the University of Leeds that unites permanent plot records and their contributing scientists from the world's tropical forests. This paper is an outcome of the ForestPlots.net approved research project #60 'Towards mapping pan-tropical tree species diversity using GEDI lidar data'. The development of ForestPlots.net was funded by several grants, including NE/B503384/1, NE/N012542/1 BIO-RED, ERC AdG 291585 'T-FORCES', and Gordon and Betty Moore Foundation #1656 ('RAINFOR'). The collection of field data in yan was done in the framework of the COBIMFO project (Congo Basin integrated monitoring for forest carbon mitigation and biodiversity; contract no. SD/AR/01A) and was funded by the Belgian Science Policy Office (Belspo). The 'Institut National pour l'Étude et la Recherche Agronomiques' (INERA) assisted in plot establishment and provided logistical support (Belspo). We thank the World Wildlife Fund for funding and facilitating lidar data collection over yan and mal. Data collection on tree diversity in the Costa Rican sites (cha) was supported by grants from the Andrew W. Mellon Foundation, NSF DEB-0424767, NSF DEB-0639393, NSF DEB-1147429, NASA Terrestrial Ecology Program, and the University of Connecticut Research Foundation.

DATA AVAILABILITY STATEMENT

Most of the field and lidar data used in this study are available and can be downloaded directly from the internet. Otherwise the data sets can be requested as described below. We have grouped the data in four groups: (a) LVIS lidar data, (b) ALS lidar data, (c) field data and (d) GEDI lidar data.

(a) 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 https://doi.org/10.3334/ORNLDAAC/1591

The LVIS data for the *cha* and *lsv* study sites are available on the following website: https://lvis.gsfc.nasa.gov/Data/Maps/CR200 5Map.html

(b) ALS lidar data

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

The ALS data over *s*11 and *s*12 can be downloaded from the sustainable landscapes data portal http://www.paisagenslidar.cnptia. embrapa.br/webgis/

The ALS data over *yan* and *mal* are available through ARcGIS online at https://www.arcgis.com/home/item.html?id=a6095e7754 1d4ad88dc6f0945639d089

The ALS data over *bci* can be downloaded directly using the following download link: http://www.life.illinois.edu/dalling/lidar_data. tgz

The ALS data over *tam* are not publicly available online as they are actively supporting external research projects. However, anyone interested in working with this data can contact Chris Hopkinson (c.hopkinson@uleth.ca) or Ross Hill (rhill@bournemouth.ac.uk) to request access.

The ALS data over *dan* and *sep* are currently in the process of being made available through the Centre for Environmental Data Analysis (CEDA) https://www.ceda.ac.uk/

(c) Field data

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

The *dan*, *rab* and *bci* field data are all available on request through the Forestgeo website at https://forestgeo.si.edu/explo re-data: https://forestgeo.si.edu/explore-data/rabi-termscondi tionsrequest-form, https://forestgeo.si.edu/explore-data/barro -colorado-island-termsconditionsrequest-forms, https://fores tgeo.si.edu/explore-data/danum-valley-termsconditionsreque st-forms

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

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

The *s*11 and *s*12 data were available through 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/ The *cha* field data set can be requested here http://neoselvas. wordpress.uconn.edu/data/

The *lsv* data can be accessed through the following website: https://tropicalstudies.org/carbono-project/#1554994367217-6bb19222-75b7

The *mab* field data are available through the following website: https://github.com/umr-amap/centrafriplots

The *mal* data are available upon request through https://www. gfbinitiative.org/datarequest

(d) GEDI lidar data

The different lidar data products from GEDI used to create Figure 8 can be download through https://doi.org/10.5067/GEDI/ GEDI01_B.001, https://doi.org/10.5067/GEDI/GEDI02_A.001 and https://doi.org/10.5067/GEDI/GEDI02_B.001

ORCID

Suzanne M. Marselis https://orcid.org/0000-0001-9187-8799 Alfonso Alonso https://orcid.org/0000-0001-6860-8432 Jean-Francois Bastin https://orcid.org/0000-0003-2602-7247 Elizabeth Kearsley https://orcid.org/0000-0003-0046-3606 Michael O'Brien https://orcid.org/0000-0003-0943-8423 John Poulsen https://orcid.org/0000-0002-1532-9808 Hao Tang https://orcid.org/0000-0001-7935-5848

REFERENCES

- Bae, S., Levick, S. R., Heidrich, L., Magdon, P., Leutner, B. F., Wollauer, S., ... Muller, J. (2019). Radar vision in the mapping of forest biodiversity from space. *Nature Communications*, 10, 4757. https://doi. org/10.1038/s41467-019-12737-x
- Bastin, J. F., Barbier, N., Réjou-Méchain, M., Fayolle, A., Gourlet-Fleury, S., Maniatis, D., ... Bogaert, J. (2015). Seeing Central African forests through their largest trees. *Scientific Reports*, *5*, 13156.
- Bongalov, B., Burslem, D. F. R. P., Jucker, T., Thompson, S. E. D., Rosindell, J., Swinfield, T., ... 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. https://doi.org/10.1111/ele.13357

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. 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. *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. https://doi. org/10.1007/s10021-007-9041-z

Chisholm, R. A., Muller-Landau, H. C., Abdul Rahman, K., Bebber, D. P., Bin, Y., Bohlman, S. A., ... Zimmerman, J. K. (2013). Scale-dependent relationships between tree species richness and ecosystem function in forests. *Journal of Ecology*, 101, 1214–1224. https://doi. org/10.1111/1365-2745.12132

Clark, D. B., & Clark, D. A. (2000). Landscape-scale variation in forest structure and biomass in a tropical rain forest. *Forest Ecology* and Management, 137, 185–198. https://doi.org/10.1016/S0378 -1127(99)00327-8

Colwell, R. K. (2009). Biodiversity: Concepts, patterns and measurement. In S. A. Levin, S. R. Carpenter, H. C. J. Godfray, A. P. Kinzig, M. Loreau, J. B. Losos, ... D. S. Wilcove (Eds.), *The Princeton Guide to Ecology* (pp. 257–263). Princeton and Oxford: Princeton University Press.

- Condit, R., Ashton, P. S., Baker, P., Bunyavejchewin, S., Gunatilleke, S., Gunatilleke, N., ... Yamakura, T. (2000). Spatial patterns in the distribution of tropical tree species. *Science*, 288, 1414–1418. https://doi. org/10.1126/science.288.5470.1414
- Corlett, R. T., & Primack, R. B. (2011). Tropical rain forests: An ecological and biogeographical comparison (2nd ed.). West-Sussex: 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. https://doi.org/10.1016/S0034-4257(02)00013-5
- Dubayah, R. O., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., ... Silva, C. (2020). The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing*, 1, 100002. https://doi.org/10.1016/j. srs.2020.100002
- Dubayah, R., Hofton, M., Blair, J. B., Armston, J., Tang, H., & Luthcke, S. (2020). GEDI L2A Elevation and Height Metrics Data Global Footprint Level V001 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2020-04-21 from https://doi.org/10.5067/GEDI/ GEDI01_B.001
- Dubayah, R., Luthcke, S., Blair, J. B., Hofton, M., Armston, J., & Tang, H. (2020). GEDI L1B Geolocated Waveform Data Global Footprint Level V001 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2020-04-21 from https://doi.org/10.5067/GEDI/GEDI02_A.001
- Dubayah, R., Tang, H., Armston, J., Luthcke, S., Hofton, M., & Blair, J. B. (2020). GEDI L2B Canopy Cover and Vertical Profile Metrics Data Global Footprint Level V001 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2020-04-21 from https://doi. org/10.5067/GEDI/GEDI02_B.001
- Duncanson, L., Neuenschwander, A., Hancock, S., Thomas, N., Fatoyinbo, T., Simard, M., ... Dubayah, R. (2020). Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California. *Remote Sensing of Environment, 242*, 111779. https://doi.org/10.1016/j.rse.2020.111779
- Engone Obiang, N. L., Kenfack, D., Picard, N., Lutz, J. A., Bissiengou, P., Memiaghe, H. R., & Alonso, A. (2019) Determinants of spatial patterns of canopy tree species in a tropical evergreen forest in Gabon. *Journal of Vegetation Science*, 30(5), 929–939.
- 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
- Fatoyinbo, T. E., Pinto, N., Simard, M., Armston, J., Duncanson, L., Hofton, M., ... 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. https://doi.org/10.1890/13-1824.1
- 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. 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. *Ecological Applications*, 25, 1776–1789. 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 Ecology*, 218, 899–908. https://doi.org/10.1007/s11258-017-0738-6

- Givnish, T. J. (1999). On the causes of gradients in tropical tree diversity. *Journal of Ecology*, 87, 193–210. https://doi. org/10.1046/j.1365-2745.1999.00333.x
- 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. https://doi. org/10.1046/j.1461-0248.2001.00230.x
- Hancock, S., Armston, J., Hofton, M., Sun, X., Tang, H., Duncanson, L. I., ... 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. https://doi.org/10.1029/2018E A000506
- 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, e103236. https://doi.org/10.1371/journal.pone.0103236
- Jucker, T., Asner, G. P., Dalponte, M., Brodrick, P. G., Philipson, C. D., Vaughn, N. R., ... 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
- Kearsley, E., De Haulleville, T., Hufkens, K., Kidimbu, A., Toirambe, B., Baert, G., ... Verbeeck, H. (2013). Conventional tree height-diameter relationships significantly overestimate aboveground carbon stocks in the Central Congo Basin. *Nature Communications*, 4(1), 1–8.
- 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(3), 390–399. https://doi.org/10.1038/s41559-019-0799-0
- 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. https://doi. org/10.1111/j.1365-2699.2005.01272.x
- Labrière, N., Tao, S., Chave, J., Scipal, K., Le Toan, T., Abernethy, K., ... Saatchi, S.. (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, 11(10), 1–11.
- Le Toan, T., Quegan, S., Davidson, M. W. J., Balzter, H., Paillou, P., Papathanassiou, K., ... Ulander, L. (2011). The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment*.115 (11), 2850–2860.
- Lewis, S. L., Edwards, D. P., & Galbraith, D. (2015). Increasing human dominance of tropical forests. *Science*, 349(6250), 827–832.
- Liang, J., Crowther, T. W., Picard, N., Wiser, S., Zhou, M., Alberti, G., ... Reich, P. B. (2016). Positive biodiversity-productivity relationship predominant in global forests. *Science*, 354, (6309).
- 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. https:// doi.org/10.5194/bg-10-6769-2013
- Lopez-Gonzalez, G., Lewis, S. L., Burkitt, M., Baker, T. R., & Phillips, O. L. (2009). ForestPlots.net database. www.forestplots.net. Date of extraction 1 December 2017.
- Lopez-Gonzalez, G., Lewis, S. L., Burkitt, M., & Phillips, O. L. (2011). ForestPlots.net: A web application and research tool to manage and analyse tropical forest plot data. *Journal of Vegetation Science*, 22, 610–613. https://doi.org/10.1111/j.1654-1103.2011.01312.x
- MacArthur, R. H., & Wilson, E. O. (1967). The theory of island biogeography. Princeton, NJ: Princeton University Press.
- Marselis, S. M., Tang, H., Armston, J., Abernethy, K., Alonso, A., Barbier, N., ... White, L. (2019). Exploring the relation between remotely sensed vertical canopy structure and tree species diversity in Gabon. *Environmental Research Letters*, 14.

WILEN

ILEY Global Ecology

- 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
- Matsunaga, T., Iwasaki, A., Tsuchida, S., Tanii, J., Kashimura, O., Nakamura, R., ... Rokugawa, S. (2013). Current status of Hyperspectral Imager Suite (HISUI). International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 3510–3513). Melbourne, Australia: IEEE.
- McLeod, A. I., Xu, C., & Lai, Y. (2020). bestglm: Best subset GLM and regression utilities. R package version 0.37.3. Retrieved from https:// CRAN.R-project.org/package=bestglm
- 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(5), e0154988. https://doi.org/10.1371/journ al.pone.0154988
- Moles, A. T., Warton, D. I., Warman, L., Swenson, N. G., Laffan, S. W., Zanne, A. E., ... Leishman, M. R. (2009). Global patterns in plant height. *Journal of Ecology*, 97, 923–932. https://doi. org/10.1111/j.1365-2745.2009.01526.x
- 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., ... Danson, F. M. (2015). Terrestrial laser scanning for plot-scale forest measurement. *Current Forestry Reports*, 1, 239–251. https://doi.org/10.1007/s40725-015-0025-5
- 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. https://doi.org/10.1016/j.rse.2015.01.020
- Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., ... Wegmann, M. (2013). Essential biodiversity variables. *Science*, 339, 277–278. https://doi.org/10.1126/science.1229931
- Pereira, H. M., Leadley, P. W., Proenca, V., Alkemade, R., Scharlemann, J. P. W., Fernandez-Manjarres, J. F., ... Walpole, M. (2010). Scenarios for global biodiversity in the 21st century. *Science*, 330, 1496–1501. https://doi.org/10.1126/science.1196624
- 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, 111283.
- Rappaport, D. I., Royle, J. A., & Morton, D. C. (2020). Acoustic space occupancy: Combining ecoacoustics and lidar to model biodiversity variation and detection bias across heterogeneous landscapes. *Ecological Indicators*, 113, 106172.
- R Core Team. (2020). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/
- Réjou-Méchain, M., Muller-Landau, H. C., Detto, M., Thomas, S. C., Le Toan, T., Saatchi, S. S., ... Chave, J. (2014). Local spatial structure of forest biomass and its consequences for remote sensing of carbon stocks. *Biogeosciences*, 11, 6827–6840.
- 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 Sensing*, 10, 629.
- Rocchini, D., Boyd, D. S., Féret, J. B., Foody, G. M., He, K. S., Lausch, A., ... Pettorelli, N. (2016). Satellite remote sensing to monitor species diversity: Potential and pitfalls. *Remote Sensing in Ecology and Conservation*, 2, 25–36.

- Saarela, S., Holm, S., Healey, S. P., Andersen, H. E., Petersson, H., Prentius, W., ... Ståhl, G. (2018). Generalized hierarchical model-based estimation for aboveground biomass assessment using GEDI and landsat data. *Remote Sensing*, 10(11), 1832.
- 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., ... 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., ... Venticinque, E. M. (2015). An estimate of the number of tropical tree species. *Proceedings of the National Academy of Sciences USA*, 112, 7472–7477.
- Slik, J. W. F., Franklin, J., Arroyo-Rodríguez, V., Field, R., Aguilar, S., Aguirre, N., ... Zang, R. (2018). Phylogenetic classification of the world's tropical forests. *Proceedings of the National Academy of Sciences USA*, 115, 1837–1842.
- Sullivan, M. J. P., Talbot, J., Lewis, S. L., Phillips, O. L., Qie, L., Begne, S. K., ... 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.
- Ter Steege, H., Pitman, N. C. A., Killeen, T. J., Laurance, W. F., Peres, C. A., Guevara, J. E., ... Valenzuela Gamarra, L. (2015). Estimating the global conservation status of more than 15,000 Amazonian tree species. *Science Advances*, 1(10), e1500936.
- Watson, J. E. M., Darling, E. S., Venter, O., Maron, M., Walston, J., Possingham, H. P., ... 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., ... Lindenmayer, D. (2018). The exceptional value of intact forest ecosystems. *Nature Ecology and Evolution*, 2(4), 599–610.
- 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.

BIOSKETCH

Suzanne Marselis received her PhD degree in Geographical Sciences at the University of Maryland. She has a broad interest in the application of remote sensing data for mapping different aspects of biodiversity across the tropics and has an extensive background in the use of lidar data for mapping three-dimensional canopy structure.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this 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