1 **Title**

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

3

4 **Short title**

5 Lidar and Pan-Tropical Tree Species Richness

6

7 **Abstract**

8 **Aim:**

9 Mapping tree species richness across the tropics is of great interest for effective conservation and

10 biodiversity management. In this study, we evaluated the potential of full-waveform lidar data for

11 mapping tree species richness across the tropics by relating measurements of vertical canopy structure,

12 as a proxy for the occupation of vertical niche space, to tree species richness.

13 **Location:**

14 Tropics

15 **Time period:**

16 Present

17 **Major taxa studied:**

18 Trees

19   **Methods:**

20   First, we evaluated the characteristics of vertical canopy structure across 15 study sites using (simulated)

21   large-footprint full-waveform lidar data (22 m diameter) and related these findings to in-situ tree

22   species information. Then, we developed structure-richness models at the local (within 25-50 ha plots),

23   regional (biogeographic regions), and pan-tropical scale at three spatial resolutions (1.0, 0.25 and 0.0625

24   ha) using Poisson regression.

25   **Results:**

26   The results showed a weak structure-richness relationship at the local scale. At the regional scale (within

27   a biogeographical region) a stronger relationship between canopy structure and tree species richness

28   across different tropical forest types was found, for example across Central Africa and in South America

29   (R^2 ranging from 0.44-0.56, RMSD ranging between 23-61%). Modelling the relationship pan-tropically,

30   across four continents, 39% of the variation in tree species richness could be explained with canopy

31   structure alone (R^2 = 0.39 and RMSD = 43%, 0.25 ha resolution).

32   **Main Conclusions:**

33   Our results may serve as a basis for the future development of a set of structure-richness models to map

34   high resolution tree species richness using vertical canopy structure information from the Global

35   Ecosystem Dynamics Investigation (GEDI). The value of this effort would be enhanced by access to a

36   larger set of field reference data for all tropical regions. Future research could also support the use of

37   GEDI data in frameworks using environmental and spectral information for modelling tree species

38   richness across the tropics.

39   # Keywords

40   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, thirty-five percent 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, 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 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.

64 The scientific community has called for bolder science in conservation strategies to enable effective

65 management of the Earth's forests and allow for better conservation of our natural ecosystems (Lewis *et*

66 *al.*, 2015; Watson *et al.*, 2016). In this study we focus on the use of active remote sensing, specifically

67 lidar, for mapping taxonomic tree species richness in the tropics. While local tropical forest diversity is

68 largely independent of biomass in intact forests (Sullivan *et al.*, 2017), it remains unclear if substantial

69 amounts of variation in species diversity are associated with other features of forest structure. Here, we

70 explore for the first time whether small-scale vertical canopy structure variation is significantly

71 associated with the spatial variation in tropical tree species richness. On a global scale it has previously

72 been shown that canopy height explains a limited portion of the variation in tree species diversity, as

73 such data provide information on the available niche space (Gatti *et al.*, 2017). It has since been

74 hypothesized that including information on the vertical canopy structure, must explain more of the

75 variation in tree species diversity than canopy height alone, as such data provide information on the

76 occupation of the vertical niche space. Marselis *et al.* (2019) demonstrated that information on canopy

77 height and vertical canopy structure, expressed as the Plant Area Index (PAI) profile from full-waveform

78 airborne lidar data, could be used to map tree species diversity in Gabon, Africa. However, it is not clear

79 whether this relationship is of a similar nature and strength across different regions, or even the entire

80 tropics. If existent, then the use of such a structure-diversity relationship(s) could be applied at a pan-

81 tropical scale with the rapidly increasing availability of spaceborne canopy structure information derived

82 from the Global Ecosystem Dynamics Investigation (GEDI), a full-waveform spaceborne lidar system

83 (Dubayah *et al.*, 2020d). GEDI is expected to provide over 10 billion measurements of vertical canopy

84 structure across the temperate and tropical forests between 2019 and 2021.

85 Factors influencing tree species diversity on a global scale differ from those affecting spatial patterns at

86 regional or local scales. In general, tropical tree species diversity increases with increasing precipitation,

87 forest stature, soil fertility, time since catastrophic disturbance, and rate of canopy turnover; and

88   decreases with seasonality, latitude, and altitude (Givnish, 1999). At large-grain scales historical

89   biogeographical processes are more important, whereas at the plot-scale environmental variables

90   strongly influence diversity (Keil & Chase, 2019).

91   Similar to species diversity, forest structure at the global scale is influenced by interacting historic,

92   environmental, and human related variables; precipitation in the wettest month being the most

93   important single predictor of plant height (Moles *et al.*, 2009). Forest structure measured in the field is

94   mainly comprised of four variables: canopy height, biomass, basal area, and tree density (Palace *et al.*,

95   2015). However, active remote sensing techniques have revolutionized the study of canopy structure

96   (Newnham *et al.*, 2015). With lidar remote sensing, for example, it is now possible to obtain information

97   on canopy height, as well as the position and amount of plant material along the vertical axis of the

98   canopy (Tang *et al.*, 2012). Palace *et al.* (2015) stressed that high resolution lidar data possess vertical

99   structure information which is inherently linked to ecological processes.

100   We hypothesize that structure-diversity relationships will vary across different biogeographical and

101   phylogenetic regions (Corlett & Primack, 2011; Slik *et al.*, 2018) and that it may be more fruitful to

102   develop multiple relationships rather than one pan-tropical relationship for operationalizing tree species

103   diversity mapping with spaceborne active remote sensing data. Additionally, the strength of the

104   relationship between a variable and tree species diversity often changes with resolution (plot size) as

105   tree species diversity is not linearly related with area (species-area curve) (MacArthur & Wilson, 1967).

106   This complicates the development of predictive models at specific resolutions, and also limits the

107   extrapolation of estimates at one resolution to a larger area, which impedes the mapping of pan-tropical

108   tree species diversity at high spatial resolution.

109   In sum, we know that both species diversity and canopy structure vary greatly within and across

110   continents. Hence, our objective is to assess whether canopy structure information can explain tree

111   species richness at the local, regional and/or pan-tropical scale with the ultimate goal to evaluate the

112   efficacy of spaceborne full-waveform lidar for mapping tree species richness across the tropics. First, we

113   compare characteristics of the vertical canopy structure, measured with full-waveform lidar data, for

114   tropical forests across the world. Second, we evaluate the differences in species richness and species-

115   area curves across the different study sites using field measurements. Third, we evaluate the potential

116   for developing local (within 25-50 ha field plots), regional (within biogeographical regions) and pan-

117   tropical structure-richness relationships, relating canopy structure metrics from lidar to tree species

118   richness measurements from the field at three spatial resolutions (0.0625, 0.25 and 1.0 ha). Lastly, we

119   discuss the potential of full-waveform lidar data from GEDI for mapping tree species richness across the

120   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 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 covering this range of possibilities.

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 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-

144   letter code and contained information on tree location, species, and diameter at breast height (DBH). All

145   datasets were collected by different organizations and research teams resulting in different data

146   characteristics (Table 1, SI1). Four datasets consisted of one large plot of 25 ha (*rob*, Australia and *rab*,

147   Gabon) or 50 ha (*dan*, Malaysia and *bci*, Panama). The other eleven datasets consisted of multiple (3-21)

148   smaller plots with sizes ranging from 0.16 ha to 4.0 ha.
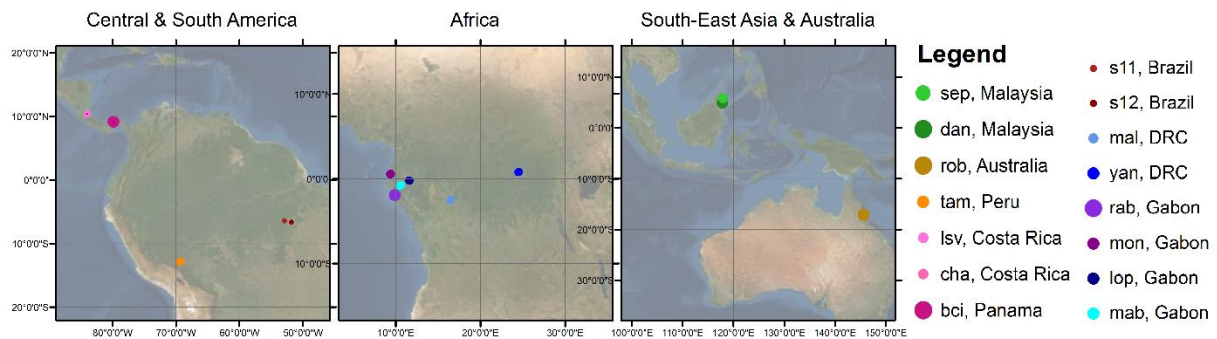
149
150   *Figure 1: Location of field sites across the three continents, colors of each study site are consistent*
151   *throughout the paper. Gridlines indicate 10° intervals in longitudinal and latitudinal directions. The size*
152   *of the place markers represents the size of the total sampled area relative to each other.*

153

154 *Table 1: Information on the original plot size, the amount of total area sampled in the field and the*
155 *source of the data which is either a website where the data are published and/or a publication in which*
156 *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/ (Lopez-Gonzalez *et al.*, 2009, 2011; 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; 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 *et al.*, 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/carbono-project/ (Clark & 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 & Dalling, 2013)

157

158 In this study, we assessed the structure-richness relationship at three spatial resolutions (1.0, 0.25,

159 0.0625 ha) because of the non-linear relationship between the number of tree species (S) and sampled

160 area. We selected squares of 1.0 ha (100 x 100 m) because they are often-used in ecology and it has

161 been shown that the spatial mismatch of plot location and remote sensing products is minimized at this

162 resolution (Réjou-Méchain *et al.*, 2014). We used squares of 0.25 ha (50 x 50 m) because these yielded

163 the best results describing the structure-diversity relationship in Gabon (Marselis *et al.*, 2019), and

164 squares of 0.0625 ha (25 x 25 m) because they correspond to a resolution close to the GEDI footprint

165   size. The datasets were used at one, two or three of the aforementioned resolutions depending on the

166   original plot size and the availability of stem maps or subplots (Table 1, full table in SI1). For each of the

167   field sites we calculated S for the entire dataset and for each plot at each plot size (Table 2). Only live

168   trees with a DBH ≥ 10 cm were included, to ensure consistency among datasets, and we included all

169   plots of each resolution in which more than 80% of the trees were identified to at least the genus level.

170   *Table 2: The total number of species identified at each study site and the average (x̄) and standard*
171   *deviation (s) of the species richness for each of the three plot sizes expressed as x̄ ± s (including only live*
172   *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*	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 ± 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 ± 8	42 ± 6	17 ± 3

173

174   **2.2 Lidar Datasets**

175   Each of the field datasets had coincident discrete return airborne laser scanning (ALS) data, or full-

176   waveform lidar data from the Land Vegetation and Ice Sensor (LVIS), collected over the field plots within

177   5 years of field data collection. We used the GEDI simulator (Hancock *et al.*, 2019) to create lidar

178   waveforms from the ALS data over the field plots. The ALS data was originally collected with a variety of

179   airborne instruments, but the GEDI simulator ensures a reliable GEDI-like waveform with minimal

180   influence of the original instrument-specific characteristics. In this way, all lidar information could be

181   processed consistently across all study sites ensuring a reliable inter-comparison of canopy structure

182   metrics derived from the waveforms and allowing for easy transfer of the developed models to future

183   on-orbit GEDI data. Lidar waveforms were simulated with a 22 m ground footprint (Gaussian distribution

184   of laser energy, σ = 5.5 m). Lidar waveform locations were determined by filling each field plot, using the

185   original field plot size and shape, with footprint center locations 6.25 m from the plot edge and 5 m

186   between footprint center locations (Figure 2). This allowed a reliable measure of canopy structure to be

187   acquired for each plot by averaging lidar metrics from all waveforms inside the plot, instead of using

188   single waveforms in the plot center and evaluating structure-richness relationships based on such

189   potentially unrepresentative waveforms. The following information was extracted from each simulated

190   lidar waveform using mature and published algorithms: canopy height (expressed as the 98th percentile

191   of the relative height metric; RH98), total Plant Area Index (PAI), and Plant Area Index at a 1 m vertical

192   resolution (Drake *et al.*, 2002; Tang *et al.*, 2012; Marselis *et al.*, 2018; Hancock *et al.*, 2019). The 1 m

193   vertical profile was used to compare the canopy structure across the study sites. It was aggregated into

194   a 10 m vertical profile, summing all PAI values in each 10 m vertical bin, to be used in the structure-

195   richness analyses. We chose to use the PAI profile because it is a biophysical variable describing the

196   amount of plant material along the vertical forest axis, thus directly indicating the occupation of vertical

197   space. Marselis *et al.*, (2019) previously showed this information relates well to tree species richness in

198   Africa. The average of each of the resulting metrics from all waveforms within each plot was computed

199   to represent the canopy structure for each plot at each spatial resolution.
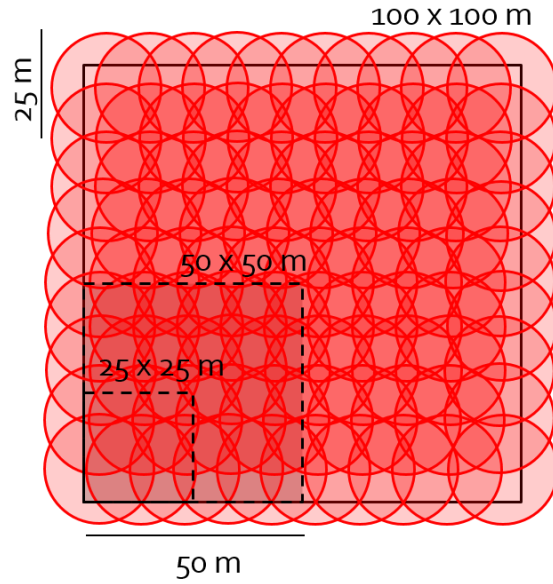
200
201 *Figure 2: Illustration of simulated lidar waveform layout. The waveforms (red circles) have a Gaussian*
202 *energy distribution with σ=5.5 m, resulting in a roughly 22 m diameter footprint. Example of simulated*
203 *footprint distribution locations in a 1.0 (solid outline), 0.25 and 0.0625 ha field plot (dashed outline).*
204 *Note: this footprint distribution was chosen to accurately depict canopy structure within the 0.0625, 0.25*
205 *and 1.0 ha plots but it does not represent the spatial distribution of spaceborne GEDI waveforms.*

206 **2.3 Canopy Structure across the tropics**

207 To evaluate the canopy characteristics across the different study sites we calculated the median plant

208 area volume density profile (composed of the PAI values for each 1 m vertical bin), using all simulated

209 lidar waveforms for each study site. In addition to the median (50th percentile), we calculated the 10th,

210 30th, 70th and 90th percentiles of the PAI values in the same 1 m vertical bins, to provide a representative

211 distribution of the canopy structure across each study site.

212 **2.4 Species-area relationships across the tropics**

213 We created species-area relationships, calculating the mean and standard deviation of S for plot sizes

214 ranging between 0.01 and 50 ha, to assess how species richness changes by plot size across our study

215 sites. Each of the original field plots was filled with as many non-overlapping subplots as possible at 17

216 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,

217 25.0, 50.0 ha) with each tree assigned to a subplot at each resolution. The plot sizes used at each study

218   site depended on the original plot size and the availability of stem maps (SI1). We visualized the mean

219   and standard deviation of S for each plot size at each study site to evaluate the differences in species-

220   area curves across the tropics.

221   **2.5 Structure-Richness Analysis**

222   To evaluate the existence of a relationship between vertical canopy structure and tree species richness

223   across the tropics, we developed models at three scales: local, regional, and pan-tropical, because many

224   historical and environmental drivers of (tree) species diversity have stronger or weaker relations

225   depending on the scale of observation (Gaston, 2000; Keil & Chase, 2019) as do different ecosystem

226   functions (Chisholm *et al.*, 2013). Definitions of the scales are presented in the following sections.

227   **2.5.1 Local Analysis**

228   The local analysis focused on the structure-richness relationship within large (25 or 50 ha) plots. We

229   used data from adjacent field plots to evaluate the relationship between S and the canopy structure

230   expressed as canopy height (RH98), total PAI and vertical canopy profile (PAI at 10 m vertical intervals).

231   The local analysis was performed on data collected in *bci* (50 ha), *rab* and *rob* (25 ha). The other 50 ha

232   plot (*dan)* was not suitable for this analysis because the species identification was incomplete at the

233   time of analysis (Table 1). We related the canopy structure with S using a generalized linear model with

234   a Poisson error distribution. We used 5-fold cross-validation, extracting 20% of the data at random in

235   each fold as test data. We first performed feature selection on the training data, choosing the model

236   with the lowest Bayesian Information Criterion (BIC) score, and then constructed the predictive model

237   based on the same training data. We evaluated model performance using R^2, Root Mean Squared

238   Difference as a percentage of the mean (RMSD%) and bias based on the predictions for the test data

239   (Piñeiro *et al.*, 2008). The average and 95% confidence interval of these metrics were recorded for each

240   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 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 ≤ 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, the same methods were applied as in the regional analysis.

261 *Table 3: Number of plots from each dataset used for regional and pan-tropical analysis of the structure-*
262 *richness relationships. Note that one region may not contain the same number of plots across all*
263 *resolutions due to limitations in the availability of subplot and stem map information, limiting the use of*
264 *data from some study sites to only one or two resolutions.*

| Region	Resolution (ha)	Study sites															Total
		sep	dan	rob	lsv	cha	bci	tam	s11	s12	mal	yan	rab	mon	lop	mab	
Africa	1										21	9	4	10	8	10	**62**
	0.25										21	9	4	11	11		**56**
	0.0625											9	4	12	11		**36**
South America	1																**-**
	0.25																**-**
	0.0625							6	8	21							**35**
Central America	1																**-**
	0.25				18	3	6										**27**
	0.0625				18	3	6										**27**
South-East Asia	1	9	2														**11**
	0.25	9	2														**11**
	0.0625																**-**
Pan-tropical	1	9	2	4		1	6	6			21	9	4	10	8	10	**90**
	0.25	9	2	4	18	3	6	6			21	9	4	11	11		**104**
	0.0625		6	4	18	3	6	6	8	21		9	4	12	11		**108**

265

3. Results

3.1 Vertical forest structure across the tropics

268 The vertical canopy structure of forests, in terms of the vertical distribution of plant material varies

269 between tropical regions (Figure 3). Maximum canopy height in our study sites in the Neotropics and

270 Central Africa is typically around 40 m, and slightly lower in Australia, while canopy heights in South-East

271 Asia exceed 60 m. Many sites show a distinct understory layer and a decrease in plant material through

272 the canopy. Relative to the understory, the canopy layer sharply declines in vegetation density (*sep* and

273 *dan*, Malaysia) or steadily declines along the vertical axis (*bci*, Panama; *rab*, Gabon; *mal*, DRC; *rob*,

274 Australia). This vertical distribution of declining vegetation is exacerbated in degraded forests: in *s11,*

275 *s12* (Brazil) and *mon* (Gabon), where the bulk of the vegetation exists close to the forest floor at ~5 m

15

276   height, but remnant trees in some plots may reach 40 m. Other sites, especially undisturbed ones, have

277   distinct canopy layers. In *tam* (Peru) and in the old-growth forest in *lsv* (Costa Rica) there are multiple

278   peaks of high-density vegetation across the vertical strata of the forest. The profiles of *yan* (DRC) and *lop*

279   (Gabon) are characterized by a multiple-peak pattern, with one peak 20-30 m in the canopy and another

280   within 5 m of the ground, reflecting the inherent structure of the forest-savanna mosaic. The less

281   disturbed *mab* (Gabon) forest shows high variability in canopy structure between plots (e.g. the wide

282   shaded area in Figure 3).
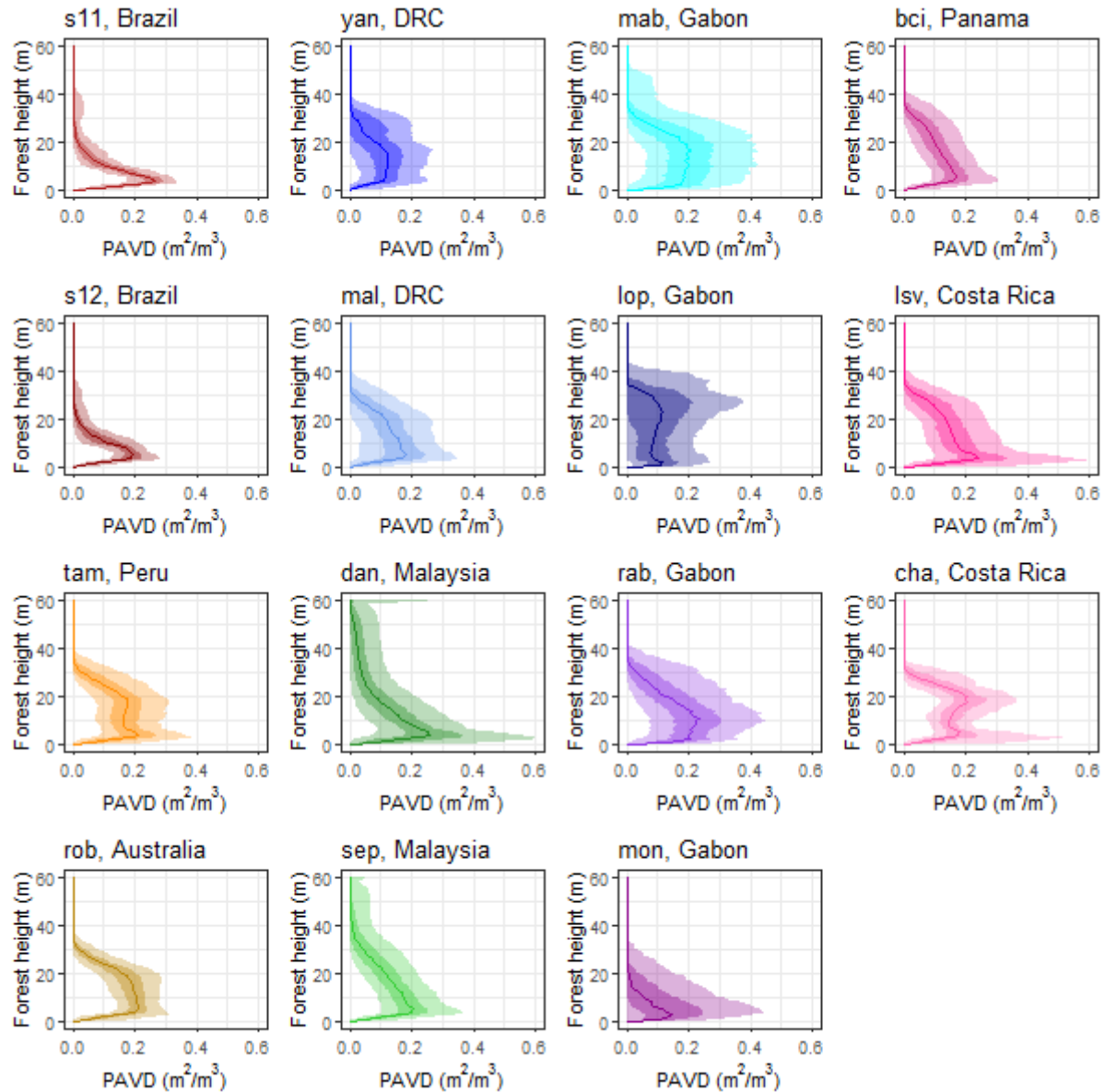
283
284 *Figure 3: Canopy structure expressed as the Plant Area Volume Density profile (PAVD), expressing the*
285 *Plant Area Index for each 1 m vertical bin, displayed as the median of all plots within each study site*
286 *(solid line), the 30th-70th percentile (darker shaded area) and 10th-90th percentile (lighter shaded area).*

287

3.2 Species-area relationships

289 The number of species increases with plot size, but the rate of increase varies across study sites (Figure

290 4). For example, in *rob* (Australia) 82-117 species occur in a 1.0 ha plot compared to 16-44 species in

291 0.0625 ha plots. By contrast, *tam* (Peru) contains 154-185 species/ha, but only 11-35 species in a 0.0625

292   ha plot, similar to *rob*. Thus, species' composition of adjacent 0.0625 ha plots in *tam* must be more

293   dissimilar from each other than adjacent 0.0625 ha plots in *rob* (Australia), in other words, the β

294   diversity of the plots in *tam* is higher than in *rob*. The species-area curves vary in shape across study

295   sites, with the highest total species richness in *tam* and lowest species richness in the African sites

296   (Figure 4). Curves that are initially steep and decrease in slope at larger plot sizes indicate a high α

297   diversity but a lower β diversity (e.g. when the area is increased, the same species are encountered).
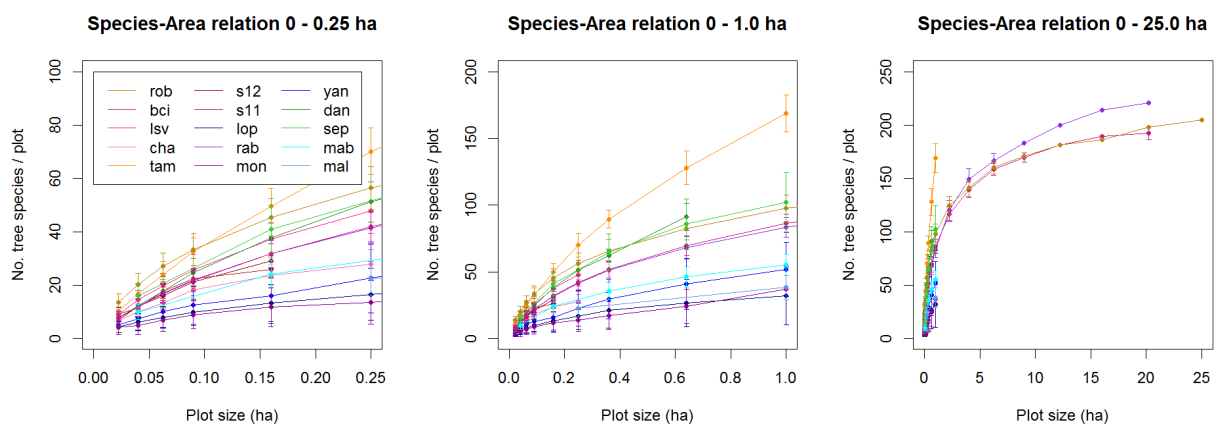

298
299   *Figure 4: Relationships between tree species richness and area for each study site (note the change in y-*
300   *axis across panels from left to right).*

301

## 3.3 Structure-richness relationships

303   Pulling together the information on tree species richness and canopy structure (RH98 and Total PAI),

304   species richness generally increases with increasing canopy height and increasing total Plant Area Index

305   across the tropics (Figure 5).
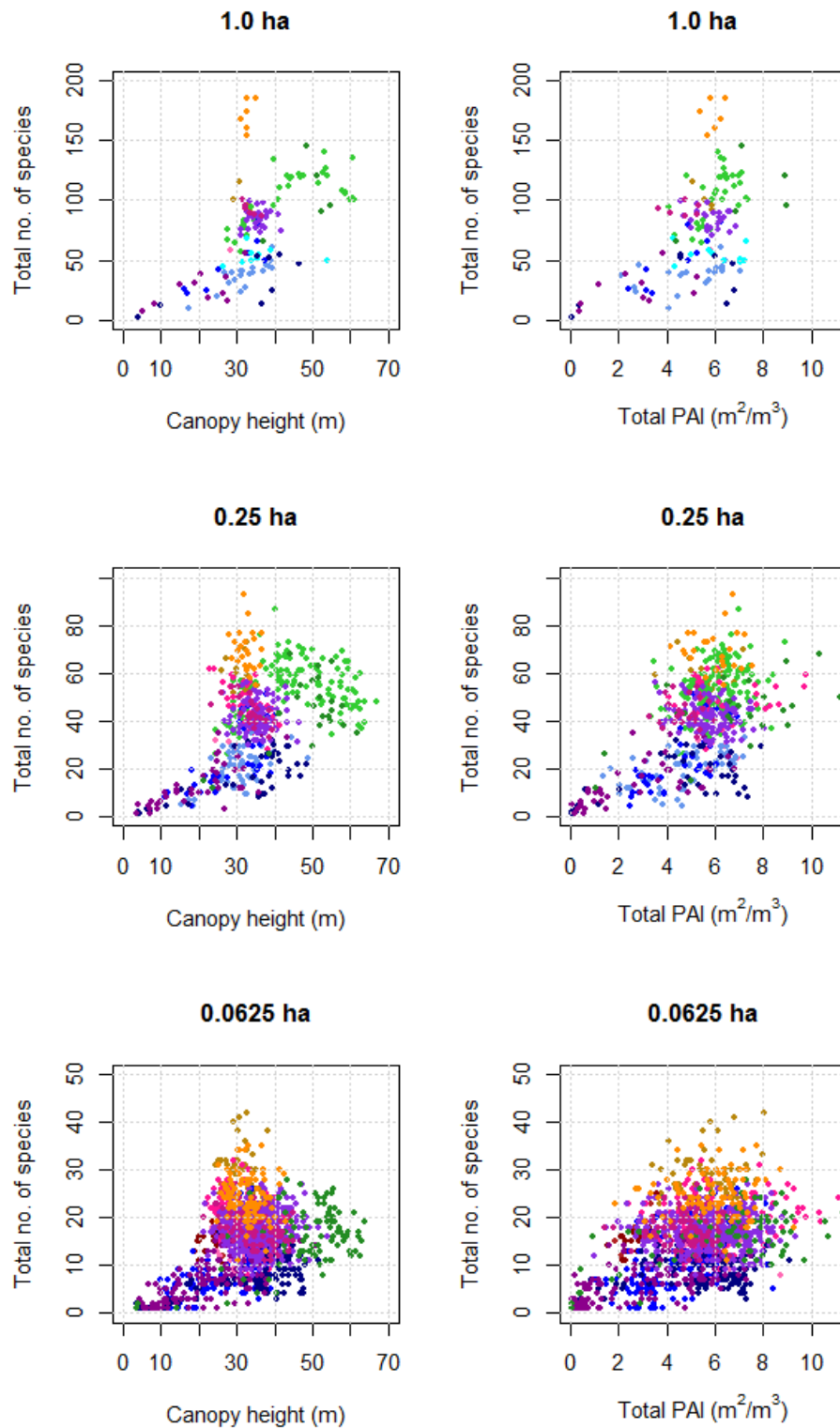
306
307 *Figure 5: Relation between canopy height (left) and total PAI (right) across three spatial scales for all*
308 *study sites across the tropics. Each point represents one plot at the specific resolution. Dots are colored*
309 *by study site corresponding according to legend in Figure 1.*

310   The cross-validation results of the local models reveal weak structure-richness relationships. Of the

311   three large plots (25 and 50 ha), only the models for *bci* (50 ha) show evidence of a significant

312   relationship between the predicted and observed values (R^2=0.32 at 1.0 ha, SI2). Even though species

313   richness within all three large plots can be predicted with a root mean squared error between 7-20% of

314   the mean species richness, the low RMSD% found only indicates that the predictions at the local scale

315   are close to the mean species richness, however in *rab* and *rob* the canopy structure is insensitive to the

316   local variation in tree species richness (see for example Figure SI2-1).

317   Regional structure-richness models generally show much better performance (Figure 6) than the local

318   models in terms of the variance in species richness that can be explained with the canopy structure

319   information (mostly significant models and higher R^2 values). However, prediction error (as percentage

320   of the mean species richness) is generally higher, partly due to the larger range in species richness in

321   these regional datasets. Regions of Africa and South America (Table 3) show the best model

322   performance whereas regions including the Costa Rica datasets show much poorer performance

323   (regions indicated with *centralamerica*). Results from an additional analysis on the compositional

324   similarity (Bray-Curtis; Faith *et al.*, 1987, SI3) of the Costa Rica dataset showed that, even though species

325   richness varies in Costa Rica (Table 2), the plots share many species, i.e. the composition is similar. In the

326   *africa* and *southamerica* datasets the variation in species richness is accompanied by a much larger

327   variation in species composition (SI3). The variation of the model performance for *seasia* is very high

328   because of the low number of plots available for this region and at the 0.25 ha resolution it was not

329   possible to create a significant model >95% of the Monte-Carlo iterations (Table 3). The model

330   performance does not provide clear results on the effect of the different resolutions, given the

331   overlapping error bars for models in the same region at multiple resolutions and the inability to create

332   each regional model at each spatial resolution (Figure 6).
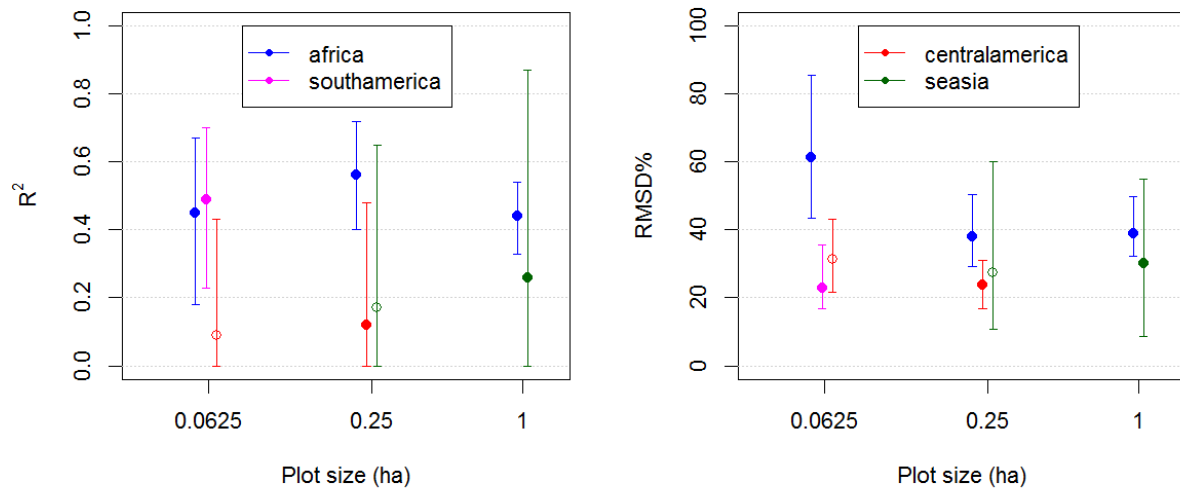
333
334 *Figure 6: Cross-validated model performance of regional structure-richness models. Error bars indicate*
335 *the 95% range of values for each performance metric. Solid dots indicate >95% of the generated models*
336 *was statistically significant, open circles indicate a lower percentage was significant.*

337 Pan-tropical structure-richness models show varying performance across the spatial resolutions with

338 mean R^2 ranging between 0.25 and 0.39 and RMSD% between 66 and 43% for the plot sizes from 1.0

339 and 0.0625 ha (Figure 7). However, the error bars of the model performance at different resolutions are

340 overlapping, indicating that no resolution has a statistically better performance. Around 39% of the

341 variation in tree species richness can be explained using canopy structure metrics alone at the 0.25 ha

342 resolution at the pan-tropical scale. Sites with extremely high values of observed species richness are

343 generally predicted poorly (SI4).

344
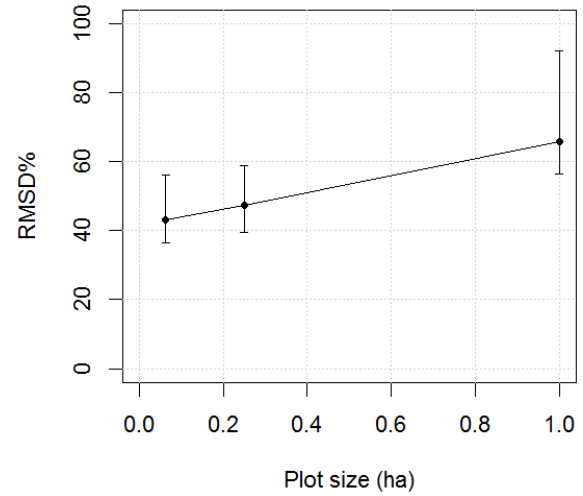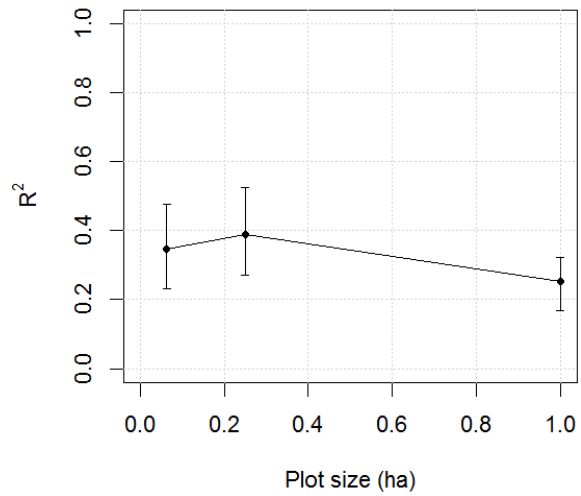
345

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

348

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, SI1 and SI3). 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. 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. 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 *et al.*, 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 (SI3).

23

372 Increasing the scale, we found that regions consisting of sites exhibiting a large variation in species

373 composition among plots, but with a similar biogeographical history, show a much stronger structure-

374 richness relationship. However, we note that model performance differed quite drastically across

375 regions. The forest in *lsv,* Costa Rica, consists of largely similar species composition, whereas species

376 composition is much more varied in regions where the structure-richness models perform better (South-

377 America, Africa), supporting the result from local scale models that species richness can be better

378 predicted from canopy structure in areas with greater β diversity.

379 At the pan-tropical scale we find a significant relationship between canopy structure and tree species

380 richness across all spatial resolutions. At the intermediate resolution (0. 25 ha) this relationship appears

381 to be slightly stronger than at the higher and lower resolutions, but no significant difference was found.

382 However, the observed difference may be attributed to the lower sensitivity of species richness to rare

383 species at smaller plot sizes. For example, *tam* (Peru) plots have very high species richness at the 1.0 ha

384 resolution (Table 2), whereas at the 0.0625 ha resolution the species richness ranges between 11-35

385 species, which is still higher than most other sites but much less than at the 1.0 ha plot size. Because the

386 1.0 ha plot size captures more rare species in each plot, the 1.0 ha pan-tropical model predictions for

387 *tam* contain highly erroneous predictions that are not present in 0.0625 ha models (SI4). Rare species do

388 not contribute much to the canopy structure, thereby complicating the relationship between structure

389 and richness at a scale at which they contribute largely to species richness numbers.

390 **4.2 Limitations**

391 This research could be significantly improved by using more coincident lidar and field data to thoroughly

392 evaluate the existence and strength of the structure-richness relationship across all tropical regions.

393 However, the collection of such data is costly and time-consuming. Here, we were able to exploit 15

394 independently collected datasets (SI1), but data gaps exist, especially in the Amazon basin, high biomass

395   forests of Central Africa, the mainland of South-East Asia, New Guinea and Australia as well as the dry

396   tropics and montane ecosystems. Apart from the spatial representation problem, the low number of

397   plots for certain regions likely influences the observed variability in model performance. The pan-

398   tropical models (with n ≥ 90) show more stable performance than models of regions with low numbers

399   of plots (e.g. *seasia*). A training dataset that does not fully represent the range of structure in the full

400   dataset can lead to biased predictions for some of the test plots. Such errors are exacerbated by the

401   logarithmic link model in Poisson regression because errors can increase exponentially. Even so,

402   negative predictions are possible with linear regression and the risk of underestimating tree species

403   richness is higher for diverse areas. Hence, we chose to use Poisson regression, knowing that it may lead

404   to extreme predictions in some cases that should be accounted for when operationalizing this method.

405   Species diversity can be identified in many different ways (Gotelli & Colwell, 2001; Colwell, 2009) and

406   there are risks and pitfalls using just one metric. In this study we only used 'species richness' (S), defined

407   by the number of different tree species in a defined area (the plot, with different sizes), as this metric is

408   easy to interpret and a prediction of the number of species/area can probably be used most directly by

409   ecosystem managers. Hereby we did not control for the number of stems in the plot, nor for the

410   abundance of the different species. Such information can be considered, for example, by using the

411   Shannon diversity index or rarefaction curves. Moreover, depending on the type of metric, a different

412   model may need to be selected to describe the structure-richness relationship as different metrics are

413   related differently to canopy structure information. For example, a generalized linear regression with a

414   Poisson error distribution, as used here, is more suitable for estimated tree species richness as this is

415   count data, whereas a linear model with a Gaussian error distribution will be better suited for estimating

416   Shannon diversity. Hence, we chose to focus on one metric of diversity to test the structure-richness

417   relationships, while acknowledging other metrics may provide better, worse, or more useful predictions

418   of tree species diversity and these should be considered in the future.

419   This study serves as a first attempt to study the pan-tropical structure-richness relationship and should

420   be improved and further developed when more data become available. Additionally, the characteristics

421   of each dataset differed widely because all data were collected by different researchers and institutions.

422   We accounted for this as much as possible by using datasets only at reliable plot and subplot

423   resolutions, including only trees ≥ 10 cm DBH and including only plots with less than 20% of unidentified

424   trees at the genus level. Nonetheless, we acknowledge that the quality of the species identification

425   varied and may have affected our models as species identification in the tropics can be challenging due

426   to the vast variety of tree species and the fact that new species are still encountered. Species

427   identification of new and existing data could be improved using more botanists or genetic tests in the

428   lab, which has been done for some of the datasets used here, but is not yet feasible for all datasets.

429   Additionally, including information on species for trees with DBH ≥ 10 cm omits large diversity found in

430   the understory. Fricker *et al.* (2015) showed that especially this diversity variation in small trees related

431   well to the canopy structure. Future research should examine if these findings are consistent across the

432   tropics.

433   The availability of stem maps and subplots in each study site determined the spatial resolutions at which

434   datasets could be used. This resulted in the inclusion of different datasets for each region (Table 3). This

435   makes the comparison of model performance in the same region at different resolutions unreliable

436   because the models were not always built on the same data (plots and study sites), but we weighed this

437   decision to maximize the sizes of the datasets used to build the structure-richness models. Hence, no

438   conclusion can be drawn about the optimal resolution for the structure-richness relationships.

439   Accurate geolocation of field plots is key for the development of reliable species-richness models

440   (Fricker *et al.*, 2015). However, geolocation of field plots in the tropical forest can be challenging due to

441   difficulties receiving a reliable GPS signal under dense canopy. This should be taken into account,

442   especially when evaluating the performance of models build with small field plots, where the effects of

443   such geolocation errors will be larger (Réjou-Méchain *et al.*, 2014).

444   We included data from a range of forest stages, including old-growth forest, successional stages,

445   disturbed forest and even low tree density savanna sites. The relationships we found are partially driven

446   by this gradient (Figure 5). However, we deemed it essential to include data from across this range of

447   forest types, because if this method is to be operationalized using canopy structure information from

448   across the tropics, we will encounter all these different stages of forest (Lewis *et al.*, 2015). We

449   acknowledge that climatic, edaphic, and topographic variables could also impact tree species richness

450   across the tropics, such as mean annual temperature and precipitation (Keil & Chase, 2019) and slope

451   and elevation (Robinson *et al.*, 2018). However, in this study we specifically focused on the relation

452   between canopy structure and tree species diversity, in light of the recently launched GEDI mission. We

453   recognize that including such information on topographic and environmental variables may further

454   improve the mapping of tree species richness across the tropics.

455   **4.3 Future research & Applications**

456   Our results provide confidence regarding the existence of regional and pan-tropical structure-richness

457   relationships that may be used to map pan-tropical tree species richness. The most accurate predictions

458   seem to be achieved at the regional scale when adequate data are available and when forested areas

459   are grouped by regions of similar biogeographical history. However, in the absence of such data it may

460   be of more immediate interest to further develop pan-tropical models that were shown to explain up to

461   39% of variation in tree species richness. At the time of writing, GEDI is collecting canopy structure

462   information close to the finest resolution tested here (0.0625 ha) and thus these data may be well suited

463   for mapping tree species richness across the tropics. GEDI is a sampling mission in which lidar

464   waveforms with 25 m diameter footprints are collected across 8 tracks with 600 m between-track

465   spacing, 60 m along-track spacing (Figure 8). By the end of its nominal two-year mission, GEDI will have

466   sampled roughly 4% of total land area.
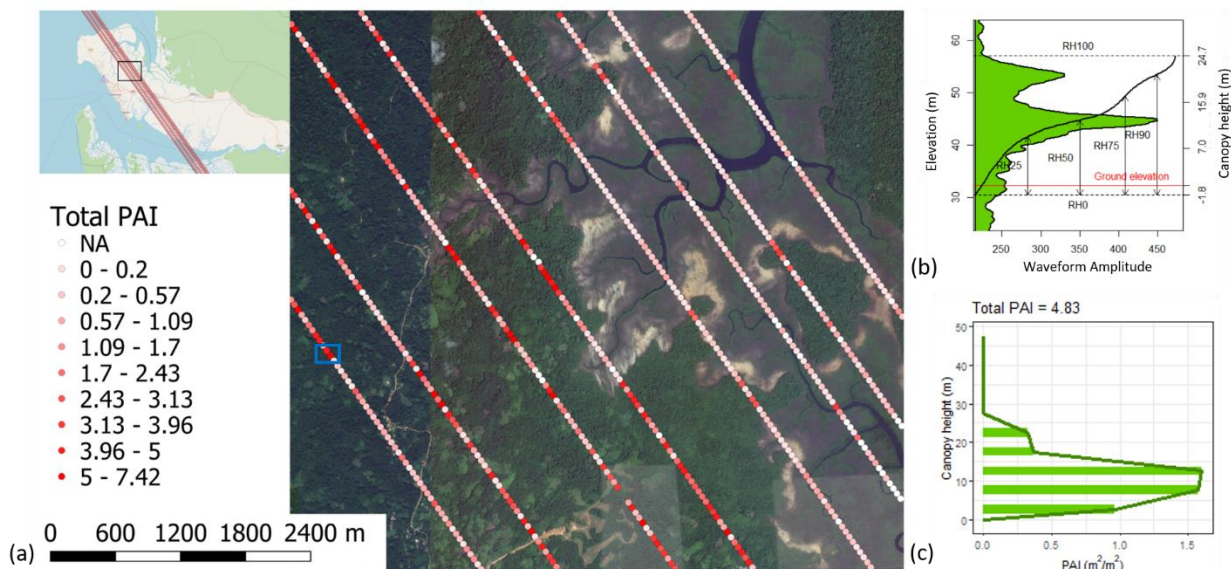
467
468   *Figure 8: (a) Example of GEDI data captured over the east of Mondah forest, north-west of Libreville, in*
469   *Gabon, Africa. The lidar waveforms are collected along-track with 8 tracks, a between-track spacing of*
470   *600 m and an along-track spacing of 60 m. (b) shows an example GEDI waveform with Relative Height*
471   *metrics (shot number = 31151116800411054, orbit = 03115, track = 05633); at the location indicated*
472   *with the blue box on (a)).(c) shows the accompanying PAI profile at 5 m vertical intervals from the Level-*
473   *2 data product.*

474

475   The footprint-level GEDI information on vertical canopy structure is stored in the Level-2 data products

476   which are publicly available from the NASA Land Processes Distributed Active Archive Center (LPDAAC)[1]

477   (Dubayah *et al.*, 2020b, a,c). GEDI gridded data products will have a 1 km^2 or finer resolution (Dubayah

478   *et al.*, 2020d). Our local scale models show that predictions of adjacent 0.0625 ha plots (or in the future,

479   footprints) are on average correct, but they will not detect local nuances in species richness within

480   forests of uniform composition. We suggest that the species richness predictions could potentially be

481   used in a similar way as gridded GEDI data products by estimating the average number of

482   species/0.0625 ha within a 1 km^2 cell, as such information may still be of interest to local land managers.

---

[1] https://lpdaac.usgs.gov/

483   Given the variable species-area relationships, it is not easy to translate species richness predictions at

484   0.0625 ha resolution to the expected number of tree species in 1 km^2. Also, the amount of variance in

485   species richness explained is limited. Therefore, we propose two future research avenues of interest:

486   fusion with spectral and/or radar data and using an environmental framework. Both spectral data and

487   radar data have previously been shown to predict some of the variance in tree species richness (Foody &

488   Cutler, 2006; Wolf *et al.*, 2012; Schäfer *et al.*, 2016; Bae *et al.*, 2019; Bongalov *et al.*, 2019; Marselis *et*

489   *al.*, 2019) and may improve our models and allow for more accurate predictions of tree species richness

490   across the tropics and the creation of wall-to-wall data products at higher spatial resolution. Especially

491   data from the hyperspectral HISUI (Matsunaga *et al.*, 2013) instrument, that is soon to be launched to

492   the International Space Station, the radar BIOMASS mission (Le Toan *et al.*, 2011), the ICESat-2 mission

493   (Duncanson *et al.*, 2020) the TanDEM-X mission (Qi *et al.*, 2019) and Landsat (Saarela *et al.*, 2018), may

494   be highly relevant for such applications. Alternatively, we believe that the inclusion of structural data

495   within previously developed environmental and biogeographical frameworks will help to predict tree

496   species diversity (Keil & Chase, 2019) as such frameworks already display intrinsic differences in tree

497   species diversity. Such frameworks could benefit from GEDI lidar data providing information on the

498   occupation of the vertical niche space and likely improve predictions of tree species richness across the

499   tropics, which could then be compared to existing predictions such as from Slik *et al.* (2015). Moreover,

500   it has previously been shown that lidar data can provide interesting information about the diversity of

501   other taxa as well (Huang *et al.*, 2014; Rappaport *et al.*, 2020) and future avenues for using lidar data to

502   provide information on a holistic measure of species diversity, including many taxa, could be of

503   incredible value.

5. Conclusions

504

In this study we evaluated the existence of local, regional and pan-tropical relationships between

505

vertical canopy structure and tree species richness in the tropics at three spatial resolutions: 1.0, 0.25,

506

and 0.0625 ha. Full-waveform lidar data provides detailed information on the differences in vertical

507

canopy structure between forests across the tropics. Our results show that canopy structure can explain

508

a significant percentage of variation in tree species richness across different biogeographical regions. A

509

full set of regional structure-richness models will most likely aid accurate pan-tropical species richness

510

mapping, but the development of such a set of models is contingent on the availability of sufficient

511

coincident field & lidar data across the tropics. Using one single predictive model at a pan-tropical scale,

512

39% of the variation in tree species richness could be explained using the vertical canopy structure.

513

Given this canopy structure is measured directly from GEDI waveforms at the footprint level, this

514

provides an interesting avenue for mapping tree species richness at high spatial resolution.

515

Alternatively, canopy structure information from GEDI could be included in existing modeling

516

frameworks, combining structural with spectral, environmental and topographic information to create

517

more accurate tree species richness predictions.

518

References

519

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

525 Bastin, J.F., Barbier, N., Réjou-Méchain, M., Fayolle, A., Gourlet-Fleury, S., Maniatis, D., De Haulleville, T.,
526   Baya, F., Beeckman, H., Beina, D., Couteron, P., Chuyong, G., Dauby, G., Doucet, J.L., Droissart, V.,
527   Dufrêne, M., Ewango, C., Gillet, J.F., Gonmadje, C.H., Hart, T., Kavali, T., Kenfack, D., Libalah, M.,
528   Malhi, Y., Makana, J.R., Pélissier, R., Ploton, P., Serckx, A., Sonké, B., Stevart, T., Thomas, D.W., De
529   Cannière, C. & Bogaert, J. (2015) Seeing Central African forests through their largest trees. *Scientific*
530   *Reports*, **5**.

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

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

537 Bradford, M.G., Metcalfe, D.J., Ford, A., Liddell, M.J. & McKeown, A. (2014) Floristics, stand structure
538   and aboveground biomass of a 25-ha rainforest plot in the wet tropics of Australia. *Journal of*
539   *Tropical Forest Science*, **26**, 543–553.

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

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

552 Clark, D.B. & Clark, D.A. (2000) Landscape-scale variation in forest structure and biomass in a tropical
553   rain forest. *Forest ecology and management*, **137**, 185–198.

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

556 Condit, R., Ashton, P.S., Baker, P., Bunyavejchewin, S., Gunatilleke, S., Gunatilleke, N., Hubbell, S.P.,
557   Foster, R.B., Itoh, A., LaFrankie, J. V., Lee, H.S., Losos, E., Manokaran, N., Sukumar, R. & Yamakura,
558   T. (2000) Spatial patterns in the distribution of tropical tree species. *Science*, **288**, 1414–1418.

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

561   Drake, J.B., Dubayah, R.O., Knox, R.G., Clark, D.B. & Blair, J.B. (2002) Sensitivity of large-footprint lidar to
562   canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment*, **81**, 378–
563   392.

564   Dubayah, R., Hofton, M., Blair, J.B., Armston, J., Tang, H. & Luthcke, S. (2020a) GEDI L2A Elevation and
565   Height Metrics Data Global Footprint Level V001 [Data set].

566   Dubayah, R., Luthcke, S., Blair, J.B., Hofton, M., Armston, J. & Tang, H. (2020b) GEDI L1B Geolocated
567   Waveform Data Global Footprint Level V001 [Data set].

568   Dubayah, R., Tang, H., Armston, J., Luthcke, S., Hofton, M. & Blair, J.B. (2020c) GEDI L2B Canopy Cover
569   and Vertical Profile Metrics Data Global Footprint Level V001 [Data set].

570   Dubayah, R.O., Blair, J.B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner,
571   J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P.,
572   Qi, W. & Silva, C. (2020d) The Global Ecosystem Dynamics Investigation: High-resolution laser
573   ranging of the Earth's forests and topography. *Science of Remote Sensing*, **1**.

574   Duncanson, L., Neuenschwander, A., Hancock, S., Thomas, N., Fatoyinbo, T., Simard, M., Silva, C.A.,
575   Armston, J., Luthcke, S.B., Hofton, M., Kellner, J.R. & Dubayah, R. (2020) Biomass estimation from
576   simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California.
577   *Remote Sensing of Environment*, **242**, 111779.

578   Engone Obiang, N.L., Kenfack, D., Picard, N., Lutz, J.A., Bissiengou, P., Memiaghe, H.R. & Alonso, A.
579   (2019) Determinants of spatial patterns of canopy tree species in a tropical evergreen forest in
580   Gabon. *Journal of Vegetation Science*, **20**.

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

583   Fatoyinbo, T.E., Pinto, N., Simard, M., Armston, J., Duncanson, L., Hofton, M., Saatchi, S., Lavalle, M.,
584   Lou, Y., Denbina, M., Dubayah, R., Marselis, S.M., Tang, H., Hancock, S. & Hensley, S. (2017) The
585   2016 NASA AfriSAR campaign: airborne SAR and Lidar measurements of tropical forest structure
586   and biomass in support of future satellite missions. *IEEE Journal of Selected Topics in Applied Earth*
587   *Observations and Remote Sensing*, 4286–4287.

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

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

592   Fricker, G.A., Wolf, J. a., Saatchi, S.S. & Gillespie, T.W. (2015) Predicting spatial variations of tree species
593   richness in tropical forests from high resolution remote sensing. *Ecological Applications*, **25**,
594   150218095111005.

595   Gaston, K.J. (2000) Global patters in biodiversity. *Nature*, **405**, 220–227.

596   Gatti, R.C., Di Paola, A., Bombelli, A., Noce, S. & Valentini, R. (2017) Exploring the relationship between
597   canopy height and terrestrial plant diversity. *Plant Ecology*, **218**, 899–908.

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

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

601 Hancock, S., Armston, J., Hofton, M., Sun, X., Tang, H., Duncanson, L.I., Kellner, J.R. & Dubayah, R. (2019)
602   The GEDI Simulator: A Large-Footprint Waveform Lidar Simulator for Calibration and Validation of
603   Spaceborne Missions. *Earth and Space Science*, **6**, 294–310.

604 Huang, Q.Y., Swatantran, A., Dubayah, R. & Goetz, S.J. (2014) The Influence of Vegetation Height
605   Heterogeneity on Forest and Woodland Bird Species Richness across the United States. *Plos One*, **9**,
606   10.

607 Jucker, T., Asner, G.P., Dalponte, M., Brodrick, P.G., Philipson, C.D., Vaughn, N.R., Arn Teh, Y., Brelsford,
608   C., Burslem, D.F.R.P., Deere, N.J., Ewers, R.M., Kvasnica, J., Lewis, S.L., Malhi, Y., Milne, S., Nilus, R.,
609   Pfeifer, M., Phillips, O.L., Qie, L., Renneboog, N., Reynolds, G., Riutta, T., Struebig, M.J., Svátek, M.,
610   Turner, E.C. & Coomes, D.A. (2018) Estimating aboveground carbon density and its uncertainty in
611   Borneo's structurally complex tropical forests using airborne laser scanning. *Biogeosciences*, **15**,
612   3811–3830.

613 Kearsley, E., De Haulleville, T., Hufkens, K., Kidimbu, A., Toirambe, B., Baert, G., Huygens, D., Kebede, Y.,
614   Defourny, P., Bogaert, J., Beeckman, H., Steppe, K., Boeckx, P. & Verbeeck, H. (2013) Conventional
615   tree height-diameter relationships significantly overestimate aboveground carbon stocks in the
616   Central Congo Basin. *Nature Communications*, **4**.

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

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

621 Labrière, N., Tao, S., Chave, J., Scipal, K., Le Toan, T., Abernethy, K., Alonso, A., Barbier, N., Bissiengou, P.,
622   Casal, T. & others (2018) In Situ Reference Datasets From the TropiSAR and AfriSAR Campaigns in
623   Support of Upcoming Spaceborne Biomass Missions. *IEEE Journal of Selected Topics in Applied
624   Earth Observations and Remote Sensing*, 1–11.

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

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

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

642   Lopez-Gonzalez, G., Lewis, S.L., Burkit, M. & Pillips, O.L. (2011) ForestPlots.net: a web application and
643   research tool to manage and analyse tropical forest plot data. *Journal of Vegetation Science*, **22**,
644   610–613.

645   Lopez-Gonzalez, G., Lewis, S.L., Burkitt, M., Baker, T.R. & Phillips, O.L. (2009) ForestPlots.net database.
646   *www.forestplots.net*, Date of Extraction [01,12,17].

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

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

652   Marselis, S.M., Tang, H., Armston, J.D., Calders, K., Labrière, N. & Dubayah, R. (2018) Distinguishing
653   vegetation types with airborne waveform lidar data in a tropical forest-savanna mosaic: A case
654   study in Lopé National Park, Gabon. *Remote Sensing of Environment*, **216**, 626–634.

655   Matsunaga, T., Iwasaki, A., Tsuchida, S., Tanii, J., Kashimura, O., Nakamura, R., Yamamoto, H.,
656   Tachikawa, T. & Rokugawa, S. (2013) *Current status of Hyperspectral Imager Suite (HISUI)*.
657   *International Geoscience and Remote Sensing Symposium (IGARSS)*,.

658   Memiaghe, H.R., Lutz, J.A., Korte, L., Alonso, A. & Kenfack, D. (2016) Ecological Importance of Small-
659   Diameter Trees to the Structure, Diversity and Biomass of a Tropical Evergreen Forest at Rabi,
660   Gabon. *PLoS ONE*, **11**.

661   Moles, A.T., Warton, D.I., Warman, L., Swenson, N.G., Laffan, S.W., Zanne, A.E., Pitman, A., Hemmings,
662   F.A. & Leishman, M.R. (2009) Global patterns in plant height. *Journal of Ecology*, **97**, 923–932.

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

665   Newnham, G.J., Armston, J.D., Calders, K., Disney, M.I., Lovell, J.L., Schaaf, C.B., Strahler, A.H. & Danson,
666   F.M. (2015) Terrestrial laser scanning for plot-scale forest measurement. *Current Forestry Reports*,
667   **1**, 239–251.

668   Palace, M.W., Sullivan, F.B., Ducey, M.J., Treuhaft, R.N., Herrick, C., Shimbo, J.Z. & Mota-E-Silva, J. (2015)
669   Estimating forest structure in a tropical forest using field measurements, a synthetic model and
670   discrete return lidar data. *Remote Sensing of Environment*, **161**, 1–11.

671   Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W.,
672   Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith, D.P., Freyhof, J.,
673   Gregory, R.D., Heip, C., Höft, R., Hurtt, G., Jetz, W., Karp, D.S., McGeoch, M.A., Obura, D., Onoda,
674   Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M. &
675   Wegmann, M. (2013) Essential biodiversity variables. *Science*, **339**, 277–278.

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

680   Century. *Science*, **330**, 1496–1501.

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

683 Qi, W., Saarela, S., Armston, J., Stahl, G. & Dubayah, R. (2019) Forest biomass estimation over three
684   distinct forest types using TanDEM-X InSAR data and simulated GEDI lidar data. *Remote Sensing of*
685   *Environment*, **232**.

686 Rappaport, D.I., Royle, J.A. & Morton, D.C. (2020) Acoustic space occupancy: Combining ecoacoustics
687   and lidar to model biodiversity variation and detection bias across heterogeneous landscapes.
688   *Ecological Indicators*, **113**.

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

700 Robinson, C., Saatchi, S., Clark, D., Hurtado Astaiza, J., Hubel, A.F. & Gillespie, T.W. (2018) Topography
701   and Three-Dimensional Structure Can Estimate Tree Diversity along a Tropical Elevational Gradient
702   in Costa Rica. *Remote Sensing*, **10**, 629.

703 Rocchini, D., Boyd, D.S., Féret, J.B., Foody, G.M., He, K.S., Lausch, A., Nagendra, H., Wegmann, M. &
704   Pettorelli, N. (2016) Satellite remote sensing to monitor species diversity: potential and pitfalls.
705   *Remote Sensing in Ecology and Conservation*, **2**, 25–36.

706 Saarela, S., Holm, S., Healey, S.P., Andersen, H.E., Petersson, H., Prentius, W., Patterson, P.L., Næsset, E.,
707   Gregoire, T.G. & Ståhl, G. (2018) Generalized hierarchical model-based estimation for aboveground
708   biomass assessment using GEDI and landsat data. *Remote Sensing*.

709 Schäfer, E., Heiskanen, J., Heikinheimo, V. & Pellikka, P. (2016) Mapping tree species diversity of a
710   tropical montane forest by unsupervised clustering of airborne imaging spectroscopy data.
711   *Ecological Indicators*, **64**, 49–58.

712 Skidmore, A.K., Pettorelli, N., Coops, N.C., Geller, G.N., Hansen, M., Lucas, R., Mucher, C.A., O'Connor, B.,
713   Paganini, M., Pereira, H.M., Schaepman, M.E., Turner, W., Wang, T.J. & Wegmann, M. (2015) Agree
714   on biodiversity metrics to track from space. *Nature*, **523**, 403–405.

715 Slik, J.W.F., Arroyo-Rodríguez, V., Aiba, S.-I., Alvarez-Loayza, P., Alves, L.F., Ashton, P., Balvanera, P.,
716   Bastian, M.L., Bellingham, P.J., van den Berg, E., Bernacci, L., da Conceição Bispo, P., Blanc, L.,
717   Böhning-Gaese, K., Boeckx, P., Bongers, F., Boyle, B., Bradford, M., Brearley, F.Q., Breuer-
718   Ndoundou Hockemba, M., Bunyavejchewin, S., Calderado Leal Matos, D., Castillo-Santiago, M.,
719   Catharino, E.L.M., Chai, S.-L., Chen, Y., Colwell, R.K., Chazdon, R.L., Clark, C., Clark, D.B., Clark, D.A.,
720   Culmsee, H., Damas, K., Dattaraja, H.S., Dauby, G., Davidar, P., DeWalt, S.J., Doucet, J.-L., Duque, A.,
721   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.

769 Ter Steege, H., Pitman, N.C.A., Killeen, T.J., Laurance, W.F., Peres, C.A., Guevara, J.E., Salomão, R.P.,
770   Castilho, C. V, Amaral, I.L., de Almeida Matos, F.D. & others (2015) Estimating the global
771   conservation status of more than 15,000 Amazonian tree species. *Science advances*, **1**, e1500936.

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

790 Tang, H., Dubayah, R., Swatantran, A., Hofton, M., Sheldon, S., Clark, D.B. & Blair, B. (2012) Retrieval of
791   vertical LAI profiles over tropical rain forests using waveform lidar at La Selva, Costa Rica. *Remote
792   Sensing of Environment*, **124**, 242–250.

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

796 Watson, J.E.M., Darling, E.S., Venter, O., Maron, M., Walston, J., Possingham, H.P., Dudley, N., Hockings,
797   M., Barnes, M. & Brooks, T.M. (2016) Bolder science needed now for protected areas. *Conservation
798   Biology*, **30**, 243–248.

799 Watson, J.E.M., Evans, T., Venter, O., Williams, B., Tulloch, A., Stewart, C., Thompson, I., Ray, J.C.,
800   Murray, K., Salazar, A., McAlpine, C., Potapov, P., Walston, J., Robinson, J.G., Painter, M., Wilkie, D.,
801   Filardi, C., Laurance, W.F., Houghton, R.A., Maxwell, S., Grantham, H., Samper, C., Wang, S.,
802   Laestadius, L., Runting, R.K., Silva-Chávez, G.A., Ervin, J. & Lindenmayer, D. (2018) The exceptional
803   value of intact forest ecosystems. *Nature Ecology and Evolution*, **2**, 599.

804 Wolf, J.A., Fricker, G.A., Meyer, V., Hubbell, S.P., Gillespie, T.W. & Saatchi, S.S. (2012) Plant species
805   richness is associated with canopy height and topography in a neotropical forest. *Remote Sensing*,
806   **4**, 4010–4021.

807

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 datasets can be requested as described below. We have grouped the data in four groups: (i) LVIS lidar data, (ii) ALS lidar data, (iii) field data and (iv) GEDI lidar data.

(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/.

The ALS data over *yan* and *mal* is available through ArcGIS online at

https://www.arcgis.com/home/item.html?id=a6095e77541d4ad88dc6f0945639d089.

The ALS data over *bci* is can be downloaded directly using the following download link:

http://www.life.illinois.edu/dalling/lidar_data.tgz.

The ALS data over *tam* is not publicly available online as it is 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.

829   The ALS data over *dan* and *sep* is currently in the process of being made available through the Centre for

830   Environmental Data Analysis (CEDA) https://www.ceda.ac.uk/.

831   **(iii) Field data**

832   Field data from *rob* has been published through the Terrestrial Ecosystem Research Network (TERN)

833   data portal linked from https://supersites.tern.org.au/supersites/fnqr-robson.

834   The *dan*, *rab and bci* field data are all available on request through the Forestgeo website at

835   https://forestgeo.si.edu/explore-data: https://forestgeo.si.edu/explore-data/rabi-

836   termsconditionsrequest-form, https://forestgeo.si.edu/explore-data/barro-colorado-island-

837   termsconditionsrequest-forms, https://forestgeo.si.edu/explore-data/danum-valley-

838   termsconditionsrequest-forms.

839   The *sep, lop*, *tam* and *yan* field data are all available upon request through forestplots.net and can be

840   found under the project names 'sepilok', 'lope', 'tambopata' and 'yangambi' at

841   https://www.forestplots.net/en/.

842   The *mon* field data is archived through the NASA data archiving center and available at DOI:

843   https://doi.org/10.3334/ORNLDAAC/1580.

844   The *s11* and *s12* were available through the data portals of the sustainable landscapes projects and can

845   be found under the field data from the São Félix do Xingu region collected in 2011 and 2012 in the

846   following data portal: http://www.paisagenslidar.cnptia.embrapa.br/webgis/.

847   The *cha* field dataset can be requested here http://neoselvas.wordpress.uconn.edu/data/.

848   The *lsv* data can be accessed through the following website: https://tropicalstudies.org/carbono-

849   project/#1554994367217-6bb19222-75b7.

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

851   amap/centrafriplots.

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

853   **(iv) GEDI lidar data**

854   The different lidar data products from GEDI used to create figure 8 can be download through

855   https://doi.org/10.5067/GEDI/GEDI01_B.001, https://doi.org/10.5067/GEDI/GEDI02_A.001, and

856   https://doi.org/10.5067/GEDI/GEDI02_B.001.