Mapping tree diversity in the tropical forest region of Chocó-Colombia

J Camilo Fagua∗,1, Patrick Jantz1, Patrick Burns1, Richard Massey1, Jeiner Y Buitrago2, Sassan Saatchi3, Christopher Hakkenberg2 and Scott J Goetz1

1 Global Earth Observation and Dynamics of Ecosystems Lab (GEODE), School of Informatics, Computing, and Cyber Systems (SICCS), Northern Arizona University, Flagstaff, AZ, United States of America
2 Department of Agronomy, Faculty of Agricultural Sciences, Universidad Nacional de Colombia, Bogota, Colombia
3 Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States of America

∗ Author to whom any correspondence should be addressed.
E-mail: Camilo.Fagua@nau.edu

Keywords: Alpha diversity, forest structure, forest inventories, LiDAR, remote sensing

Abstract

Understanding spatial patterns of diversity in tropical forests is indispensable for their sustainable use and conservation. Recent studies have reported relationships between forest structure and α-diversity. While tree α-diversity is difficult to map via remote sensing, large-scale forest structure models are becoming more common, which would facilitate mapping the relationship between tree α-diversity and forest structure, contributing to our understanding of biogeographic patterns in the tropics. We developed a methodology to map tree α-diversity in tropical forest regions at 50 m spatial resolution using α-diversity estimates from forest inventories as response variables and forest structural metrics and environmental variables as predictors. To include forest structural metrics in our modelling, we first developed a method to map seven of these metrics integrating discrete light detection and ranging (LiDAR), multispectral, and synthetic aperture radar imagery (SAR). We evaluated this methodology in the Chocó region of Colombia, a tropical forest with high tree diversity and complex forest structure. The relative errors (REs) of the random forest models used to map the seven forest structural variables ranged from low (6%) to moderate (35%). The α-diversity maps had moderate RE; the maps of Simpson and Shannon diversity indices had the lowest RE (9% and 13%), followed by richness (17%), while Shannon and Simpson effective number of species indices had the highest RE, 27% and 47%, respectively. The highest concentrations of tree α-diversity are located along the Pacific Coast from the centre to the northwest of the Chocó Region and in non-flooded forest along the boundary between the Chocó region and the Andes. Our results reveal strong relationships between canopy structure and tree α-diversity, providing support for ecological theories that link structure to diversity via niche partitioning and environmental conditions. With modification, our methods could be applied to assess tree α-diversity of any tropical forest where tree α-diversity field observations coincident with LiDAR data.

1. Introduction

Decreasing biodiversity loss and ensuring sustainable use of biodiversity are two main targets of the Convention on Biological Diversity, currently signed by 196 countries (CBD 2019). Essential biodiversity variables (EBVs) have been developed to monitor progress towards these goals (Pereira et al 2013, Vihervaara et al 2017). α-, β-, and γ-diversity measures are EBVs typically estimated by in-situ observations (Condit et al 1992, Higuchi et al 2008, Laurance et al 2010, Baraloto et al 2013, Hubbell 2013, Duque et al 2017, McNicol et al 2018, Oliveira et al 2018). However, the potential of remote sensing to predict α- β- γ- and even functional diversity has been evaluated by relating information of multispectral,
hyperspectral, synthetic aperture radar (SAR), and light detection and ranging (LiDAR) data to species inventories (Gould 2000, Foody and Cutler 2003, Carter et al 2005, Carlson et al 2007, Goetz et al 2007, 2015, Rocchini 2007, Rocchini et al 2010, Hernández-Stefanoni et al 2012, Hakkenberg et al 2018a, 2018b, Bae et al 2019, Marselis et al 2020) and to morphological and physiological traits (Schneider et al 2017, 2020).

Although the importance of tree diversity information for the conservation, management, and sustainable use of tropical forests has been recognized (Kessler et al 2009, Anderson-Teixeira et al 2015, Laurance et al 2018), regional spatial models of tree α-diversity (mean species diversity at a local scale (Whittaker 1960)) at detailed spatial resolution are lacking. Some maps have been developed regionally and globally but their relatively coarse spatial resolutions (≥1 km) (Barthlott et al 1996, 2007, Ter Steege et al 2003, Parmentier et al 2011, Keil and Chase 2019), limits their applications for local management, as does the fact that they contain relatively little information on the vegetation itself.

In situ observations have shown that the effect of vegetation structure on α-diversity patterns can differ among taxa (Allouche et al 2012, Simonson et al 2014, Ben-hur and Kadmon 2020). However, positive relationships between plant and animal α-diversity and vegetation structure are common in several vegetation types including forests, shrublands, wetlands, and grasslands (Duivenvoorden 1996, Lundholm 2009, Laurans et al 2014, Bae et al 2019, Maskell et al 2019, Marselis et al 2020, Meyer et al 2020). Two non-exclusive hypotheses could explain this ecological relationship. The niche hypothesis proposes that α-diversity should rise with the increasing niche space produced by taller and more heterogeneous vegetation structures (Hutchinson 1957, Stein et al 2014) until a maximum where it would start to decline due to the reduction of the average amount of effective area available for the species (Ben-hur and Kadmon 2020). Under this hypothesis, taller and more heterogeneous canopies should present higher tree α-diversity in tropical forest. The environmental harshness hypothesis proposes that harsh environments limit the set of viable tree species traits. Vegetation height is assumed to act as an integrative measure of limiting environmental factors, thereby leading to positive relationships between taller canopies and tree diversity (Whittaker 1965, Kraft et al 2015, Marks et al 2016). Causality between diversity and structure may act in the other direction as well, with more diverse forest stands containing a wider array of tree architectures thereby leading to greater structural complexity.

LiDAR, an active sensor that allows to map vegetation structure from the ground to the top of the canopy (Silva et al 2018, Dubayah et al 2020), has been used for assessing tree α-diversity at detailed spatial resolution (∼50 m) in central America (Fricker et al 2015, George-Chacon et al 2019), in Lopé National Park, Gabon, Africa (Marselis et al 2019), Hawaii (Carlson et al 2007), and for 15 study sites distributed across the tropics (Marselis et al 2020). These and other local scale studies demonstrate strong relationships between LiDAR and tree diversity.

Notwithstanding their efficacy for estimating detailed forest structural metrics, airborne LiDAR are expensive to acquire and typically cover limited spatial extents. Although the long-awaited spaceborne Global Ecosystem Dynamics Investigation (GEDI) LiDAR is now operational and has acquired measurements across much of Earth’s temperate and tropical forests, its orbital transects may be separated by hundreds of meters and persistent cloud cover may further reduce coverage (Dubayah et al 2020). Consequently, spatial models of α-diversity that use only LiDAR predictors also generally cover limited spatial extents. To overcome this constraint, LiDAR data can be integrated with other sensor types to obtain wall-to-wall maps of forest structural metrics for large areas (Asner et al 2010, 2012, 2013, Qi and Dubayah 2016, Potapov et al 2020). For example canopy height has been mapped in tropical forests with relatively high accuracy by integrating temporal metric predictors derived from multispectral imagery and SAR together with lidar data (Fagua et al 2019b, Healey et al 2020). Such examples show the potential of building wall-to-wall maps of forest structural metrics for large areas of tropical forest, allowing the subsequent estimation of α-diversity patterns.

Our objectives in this paper were to (a) model the relationships between forest structure and tree diversity for the Chocó-region of Colombia and (b) create the first maps of forest structure and tree α-diversity at relatively detailed spatial resolution (∼50 m) for the region. This region contains the most well preserved forest of the Chocó global ecoregion, one of 25 biodiversity hotspots prioritized for conservation (Fagua and Ramsey 2019, WWF 2019, Fagua et al 2019a). We hypothesized that α-diversity of trees would be related to forest structural metrics and environmental variables at local scales, revealing patterns at larger scales. In the results and discussion of the paper we identify and evaluate these spatial patterns from the perspective of the niche diversity (Whittaker 1965) and environmental harshness hypotheses (Marks et al 2016) in the context of biodiversity conservation in this region.

2. Methods

2.1. Study area

Our analysis focused on the Chocó-region (Iteam et al 2016), a lowland tropical forest zone of 68927 km² located along the Colombian Pacific Coast and the western branch of the Colombian Andes (figure 1). The Chocó-region includes sections of five ecoregions (Olson et al 2001), and contains
Figure 1. Study area. (a) Estimated historical extent of low land tropical forest (LTF). The Chocó Region (ChR) and the distribution of the vegetation plots (Plot of veg. in yellow) for the modelling are shown. (b) The Chocó-Darién Global Ecoregion and its five terrestrial sub-ecoregions: Chocó–Darien moist forest (ChD), Western Ecuador Moist Forest (WEc), Eastern Panamanian montane forests (EPaM), Magdalena- Urabá Moist forests (MgU), South American Pacific (SAP) Mangrove, and Amazon Orinoco-Southern Caribbean (AOSC) Mangrove. (c) Land covers of the Chocó Ecoregion: well-preserved forests (Forest), secondary vegetation (Sec. veg.), wetlands (wetland), water bodies (water body), and land cover with high anthropogenic disturbance, mainly grassland, crops and palm plantations (No natural veg.). (d) LiDAR surveys across the Chocó region (LiDAR covering) and withheld vegetation plots (Plot of veg. in blue) localized across the Chocó region to estimate the tree diversity map errors.
A high percentage of well-preserved forests (Fagua and Ramsey 2019). Because of its unique geological and biological history, the Chocó-region is one of the most rich and endemic areas on Earth in flora and fauna (Gentry 1986, Gregory-Wodzicki 2000, WWF 2019). Additional description of the region in appendix A.

2.2. Plots, α-diversity, forest structure and environmental variables

We measured α-diversity, forest structure, and environmental variables in eighty-five 0.25-hectare plots (50 m × 50 m) located in six ecoregions of lowland moist forests across South America (figure 1(a)). This size of vegetation plot tends to produce lower errors when predicting tropical tree diversity using remote sensing (Fricker et al 2015, Marselis et al 2019) and is an appropriate scale for estimating tree α-diversity (Gentry 1988, Phillips and Miller 2002, Baraloto et al 2013).

For α-diversity, in-situ observations were conducted in each plot including censuses of all live trees, live palms, and live woody lianas with a diameter at breast height (DBH) > 10 cm. These data were provided by BioREDD project for biodiversity research in Colombia (Meyer et al 2019). We then calculated a set of tree α-diversity measures to evaluate and compare tropical forest communities: the Hill numbers, Shannon index, and Simpson index (table 1). Additional details in appendix B.

For forest structure, discrete return LiDAR data were acquired in each plot. We first estimated four vertical forest structural metrics from the canopy profile, then we estimated three horizontal metrics based on the derived canopy height model. A summary of these metrics is provided in table 2. Although other metrics could be estimated (e.g. individual tree attributes and density metrics), we selected these metrics because they are comparable when different LiDAR instruments are used, can be emulated with metrics of waveform LiDAR, and have been related to α-diversity measures (MacArthur and MacArthur 1961, Davies and Asner 2014, Zhang et al 2017). Additional information regarding forest structure data is provided in appendix B and details on the lidar data acquisitions in appendix C.

For environmental variables, we selected seventy-six, 1 km (or finer) gridded datasets related to climate, soil, topography, and incident radiation which have been shown to be relevant drivers of diversity pattern (Wright et al 1993, Hawkins et al 2003, Laurance et al 2010, Guo et al 2013, Toszogoya and Storch 2019); see table 3 for a summary of these variables and appendix B for additional processing and variable selection information. We did not include any precipitation data between 2001 and 2019 due to their coarse spatial resolution. Furthermore, we did not include slope and aspect as topographic variables because the lowland tropical forests of South America are dominated by relative flat topography and the 85 plots were all in flat areas.

2.3. Plot level modelling

We modelled α-diversity estimated in the 85 plots using the forest structural metrics and environmental variables as predictors. We first assessed the predictive potential of five machine learning algorithms and selected random forest (RF) regressions because RF had the lowest errors (appendix D). With this model in mind, we first selected the set of forest structural metrics and environmental variables that maximized the RF prediction of each α-diversity measure using the Boruta algorithm (Kursa and Rudnicki 2010). We also tuned the number of trees (from 300 to 1000) and the number of variables randomly sampled at each split (from 2 to 10) (Boehmke and Greenwell 2019). Root mean squared error (RMSE), mean absolute error (MAE), and relative error (RE) were used as model-error measures. We performed two error estimations: five-fold cross-validation (CV) and data

| Table 1. Summary of the α-diversity measures computed for each plot. |
|---|---|---|---|---|
| Type | Variable name | Description | Abbreviation | Citation |
| Hill’s numbers | Richness | It is the number of species at locality | Richness | Hill (1973), Jost (2006) |
| Effective number of species-Shannon | It weights species in proportion to their frequency and is calculated as $\exp\left(-\sum p_i \log(p_i)\right)$ where $p_i$ is the proportion of total individuals contributed by the $i$th species. | ENS-Shannon | Hill (1973), Jost (2006) |
| Effective number of species-Simpson | It gives more weight to abundant species and is calculated as $1/\sum (p_i)^2$ | ENS-Simpson | Hill (1973), Jost (2006) |
| Shannon index | Shannon index | It is defined as $H = -\sum p_i \log(p_i)$ | H | Shannon (1948) |
| Shannon index | Simpson index | It is defined as $D = 1 - \sum p_i^2$ | D | Simpson (1949) |
Table 2. Summary of lidar-derived forest structure metrics computed for each plot.

| Type                        | Variable name | Description                                                                 | Abbreviation | Citation                          |
|-----------------------------|---------------|------------------------------------------------------------------------------|--------------|-----------------------------------|
| Vertical metrics            | Canopy height | Height where the 98% of the returns are accumulated                         | CH           | Davies and Asner (2014), Zhang et al (2017) |
|                             | 50th percentile of heights | Height where the 50% of the returns are accumulated                          | z_50         | Davies and Asner (2014), Zhang et al (2017) |
|                             | Vertical canopy heterogeneity | Expressed as $-\sum p_i \log(p_i)$, where $p_i$ is the proportion of returns at a given height | VCH          | Davies and Asner (2014), Zhang et al (2017) |
|                             | Standard deviation of normalized heights | Vertical standard deviation of the returns.                                  | z_sd         | Davies and Asner (2014), Zhang et al (2017) |
| Horizontal metrics         | Standard deviation at canopy height | Vertical standard deviation of the canopy height                             | CH_sd        | Davies and Asner (2014), Zhang et al (2017) |
|                             | Horizontal heterogeneity of canopy height | Expressed as $-\sum v_i \log(v_i)$, where $v_i$ is the proportion of canopy height | CH_H         | Davies and Asner (2014), Zhang et al (2017) |
|                             | Rumple index at canopy height | Ratio between canopy height area and its projected area on the ground       | CH_R         | Jenness (2004)                    |

Table 3. Summary of environmental variables. (a) WorldClim acronyms: BIO1 (annual mean temperature), BIO2 (mean diurnal range), BIO3 (isothermality), BIO4 (temperature seasonality), BIO5 (max temperature of warmest month), BIO6 (min temperature of coldest month), BIO7 (temperature annual range), BIO8 (mean temperature of wettest Quarter), BIO9 (min temperature of driest month), BIO10 (mean temperature of warmest Quarter), BIO11 (mean temperature of coldest Quarter), BIO12 (annual precipitation), BIO13 (precipitation of wettest month), BIO14 (precipitation of dry month), BIO15 (precipitation seasonality), BIO16 (precipitation of wettest Quarter), BIO17 (precipitation of driest Quarter), BIO18 (precipitation of warmest Quarter), and BIO19 (precipitation of coldest Quarter). (b) Soil acronyms: nitrogen content (N), percentage of silt content (Silt), percentage of sand content (Sand), percentage of clay content (Clay), soil pH (pH), soil organic carbon content (C), soil bulk density (Bulk), and soil water content (W).

| Type            | Variable name       | Description                                                                 | Abbreviation | Time of estimation | Citation                        |
|-----------------|---------------------|------------------------------------------------------------------------------|--------------|--------------------|---------------------------------|
| Climate         | WorldClim (version 2.0) | Nineteen climate variables Average of annual mean temperature from MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global. | BIO1, BIO2, ..., and BIO19. Tem | 1970–2000          | Hijmans et al (2005)         |
|                 | Annual mean temperature | Average of annual mean temperature from MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global. |              | 2000–2019          | Wan et al (2015)                |
| Topography      | Altitude            | SRTM-30 digital elevation model                                              | Alt          | 2000               | Jarvis et al (2008)             |
| Incident radiation | Fraction of photosynthetically active radiation | Annual mean of the MODIS product MOD15A2H | FPAR         | Agree with the vegetation survey times | Myneni and Kniazhikhin (2015) |
| Soil            | SoilGrids250 m      | Nine main variables at six soil depths (0 cm, 10 cm, 30 cm, 60 cm, 100 cm, 200 cm): | N, C, Silt, Sand, Clay, pH, Bulk, and Water | Not considered | Hengl et al (2017). |

We applied a 5-fold cross-validation (DP) in which 70% of the data are used for training and 30% are used for testing. We ran 100 iterations per method producing a distribution of RMSE, MAE, and RE. We also reported predictor variable importance generated by the Boruta algorithm, which reduces the bias of the Gini index, a common indicator of variable importance (Nicodemus and Malley 2009). Boruta compares predictor importance with shadow predictors (predictor copies generated by random permutations of their own values) in numerous RF regressions, 100 in our case. Predictors with significantly larger or smaller importance than shadow predictors are declared as important or unimportant for the modelling (Kursa and Rudnicki 2010). We also calculated linear correlation (Pearson-r) between the $\alpha$-diversity
measures and the predictors as a descriptive evaluation.

2.4. Mapping forest structure

We mapped the seven forest structural metrics integrating LiDAR, multispectral and SAR data. We first obtained the forest structural metrics at 50 m spatial resolution for 578 km\(^2\) of LiDAR surveys across the Chocó (figure 1(d)), taking the LiDAR metrics as ‘ground truth’ for training models based on the image data (Davies and Asner 2014, Guan et al 2014, Zhang et al 2017, McRoberts et al 2018, Huang et al 2019, Healey et al 2020). The LiDAR surveys were distributed randomly to cover representative samples of 36 forest strata with similar environmental characteristics (Meyer et al 2019). We followed the same methods to process the LiDAR surveys as described in section 2.2 and appendix B.

We then built temporal mean (mean-X) and standard deviation (sd-X) mosaics of multispectral and SAR data for the Chocó region using the pixel values of all imagery available between 1 January 2017 and 31 December 2019. Temporal metrics improve the modelling of forest structure variables by increasing sensitivity and reducing errors (Fagua et al 2019b, Potapov et al 2019). A summary of these metrics is provided in table 4 and additional details and codes in appendix E.

Finally, we used the RF algorithm to model and map the relationship between the forest structural metrics and the 30 multispectral and SAR predictors.

We selected 67 654 training sites (50 × 50 m pixels) in the 578 km\(^2\) of LiDAR surveys to estimate the seven forest structural metrics, using a spatial filter of 100 m\(^2\) to reduce spatial autocorrelation effects. We evaluated predictor importance using Boruta, tuned the parameters of the RF regressions, and evaluated model errors using 100 iterations for CV and DP as described above.

2.5. Mapping tree \(\alpha\)-diversity

We used the RF regressions at the plot level for modelling maps of the \(\alpha\)-diversity measures at the regional level, using stacks of predictor variables at the spatial scale of 50 m (plot size). Because the RF algorithm selects only about 66% of the training data in an analysis (Liaw 2018), we performed 30 iterations (each iteration is a RF regression) to estimate 30 maps for each \(\alpha\)-diversity measure; a final map per \(\alpha\)-diversity measure was obtained by averaging the 30 maps. Map errors (RMSE, MAE and RE) were calculated using predicted diversities against diversity values of a reserved group of fourteen 0.25 hectare plots in the Chocó region (figure 1(d)) that were not used for model training.

2.6. Spatial distribution of biodiversity

As an additional validation and evaluation of our diversity maps, we randomly selected 100 000 sites to assess differences in the spatial distribution of the tree \(\alpha\)-diversity against the terrestrial ecoregions (Olson et al 2001) and natural land covers of the Chocó.

| Remote sensing data | Satellite | Band, index name, or backscatter coefficient | Wavelength or definition |
|---------------------|-----------|--------------------------------------------|--------------------------|
| Multispectral       | Landsat 8 | Blue                                       | 0.45–0.52 \(\mu\)m        |
|                     |           | Green                                      | 0.52–0.60 \(\mu\)m        |
|                     |           | Red                                        | 0.63–0.69 \(\mu\)m        |
|                     |           | Near Infrared—NIR                         | 0.77–0.90 \(\mu\)m        |
|                     |           | Short wave infrared1—                     | 1.55–1.75 \(\mu\)m        |
|                     |           | Short wave infrared2—                     | 2.08–2.35 \(\mu\)m        |
|                     |           | NDVI—normalized difference vegetation index| (NIR-Red)/([NIR + Red])   |
|                     |           | EVI—enhanced vegetation index              | G × (NIR – Red)/([NIR + C1 × RED–C2 × Blue + L]) |
|                     |           | SAVI—soil adjusted vegetation index        | (1 + L) × (NIR-Red)/([NIR + Red + L]) |
|                     |           | SVVI—spectral variability vegetation index | Sd (Blue, Green, Red, NIR, SWIR1, SWIR2)—Sd(NIR, SWIR1, SWIR2) |
| SAR                 | Sentinel 1| VV of C band                               | 5.6 cm (5.405 GHz)        |
|                     |           | VH of C band                               | 5.6 cm (5.405 GHz)        |
|                     |           | VH/VV of C band                            | 5.6 cm (5.405 GHz)        |
|                     |           | VV–VH of C band                            | 5.6 cm (5.405 GHz)        |
|                     | ALOS-2-   | HV of the L band                           | 23.62 cm (1270 MHz)       |
| SAR                 | PALSAR    | HH of the L band                           | 23.62 cm (1270 MHz)       |
The biodiversity patterns of inventories. Higher ENS-Shannon than ENS-Simpson index values reflect higher evenness of tree species distributions with lower dominance of abundant species.

Diversity patterns changed significantly across terrestrial ecoregions; Hill’s numbers, $H$ and $D$ were higher in the three forest ecoregions than in the mangrove ecoregions (398 > $F$ values < 227, $p$ values < 0.001) (figure 5). Diversity patterns also varied significantly with natural land cover; Hill’s numbers, $H$ and $D$ were highest in forest, lower in secondary vegetation, and lowest in wetlands (980 > $F$ values < 1800, $p$ values < 0.001) (figure 6). High tree diversity is located along the Pacific Coast from the centre to the north of the region and in non-flooding areas along the boundary between the Chocó region and the Andes. The north-eastern and south-western portions of the Chocó Region tend to have lower diversity (figure 3).

### 4. Discussion

#### 4.1. Spatial modelling
Integrating in situ forest inventories and remote sensing data, we developed an accurate methodology to map tree $\alpha$-diversity metrics in tropical forests at detailed spatial resolution. We used mostly free and open-access data; therefore, our methods can be adopted in other regions to evaluate diversity. The only restricted data source was airborne LiDAR data, however our methods can be adapted to use publicly available GEDI LiDAR and airborne LiDAR datasets are becoming more common. Our methods open new possibilities for mapping and monitoring EBVs at regional to global scales as well as new prospects to understand and conserve the biodiversity of tropical forests.

Although previous work has explored relationships between tree $\alpha$-diversity and remote sensing data in tropical forest (Carlson and Ripley 1997,
Fricker et al. 2015, Schäfer et al. 2016, Marselis et al. 2019), we were able to use these relations to map tree α-diversity over a relatively large extent by using aircraft LiDAR as the key link between forest inventories and continuous remote sensing imagery. LiDAR data allowed accurate estimations of forest structural metrics that were related to tree diversity at the level of forest inventories. These forest structural metrics were also related to continuous remote sensing imagery, allowing the construction of accurate regional maps of forest structure, a primary factor driving diversity at fine scales in tropical forests.

We found forest structural metrics were highly correlated to tree α-diversity and were important for their prediction. CH is an indicator of successional stage (Finegan 1996, Guariguata and Ostertag 2001) and a proxy for canopy volume and thus niche availability (Gatti et al. 2017, Marselis et al. 2019), explaining its high correlation to tree α-diversity. Higher $z_{50}$ should correspond to high proportions of mature trees and low density of understory, traits related to late successional forests with high diversity (Nicotra et al. 1999, Guariguata and Ostertag 2001). VCH and $z_{sd}$ show the vertical variations of the LiDAR impulse density; consequently, these metrics are sensitive to canopy stratification (Hirata et al. 2009, Bakx et al. 2019, Oliveira and Scheffers 2019). Higher VCH and $z_{sd}$ thus suggest more canopy stratification and more tree species because trees tend to occupy different vertical strata due to light competition. This hierarchy of the canopy is considered one of the main mechanisms that maintains the high tree diversity of tropical forests (Sheil et al. 2006). Higher canopy stratification also suggests more niches and thus diversity as niche hypothesis proposes. Evidence of vertical niche partitioning has been related to diversity of the maximum plant height trait, providing support for the statement that niche partitioning is an important mechanism for maintaining high diversity (Laurans et al. 2014). CH_sd, CH_H, and CH_R are indicators of canopy heterogeneity; higher structural heterogeneity suggests more potential for ecological niches and thus higher diversity (Torresani et al. 2020).

Climate (Hawkins et al. 2003), elevation (Guo et al. 2013) and radiation (Toszogyova and Storch 2019) drive global diversity patterns, explaining their importance in our RF modelling. Precipitation...
Figure 3. Maps of tree-diversity for the Chocó Region: (a) map of species richness (Richness), (b) map of Shannon's entropy index (ENS-Shannon), (c) map of Simpson's concentration index (ENS-Simpson), (d) map of Shannon diversity index ($H$), and (e) map of Simpson diversity index ($D$). Protected areas of the Chocó region are outlined in black.

Seasonality (BIO15) and temperature were important and negatively correlated to $\alpha$-diversity, indicating that longer periods of seasonal water deficit plus higher temperatures decrease tree $\alpha$-diversity in tropical forests as predicted by the harshness hypothesis. The southern and northern portions of Amazon forests present this combination of harsh conditions, resulting in fire increases and diversity reduction (Davidson et al 2012, Brando et al 2014). Although most soil predictors did not influence our modeling, some predictors related to soil structure (sand, clay, and bulk density) from 0 cm to 30 cm tended to have low but significant importance. Sand and clay proportion at low depths regulates water content by reducing water loss at high precipitation seasonality and temperature in the tropical forest (Maia et al 2020). Therefore, these soil variables are related to tree diversity, temperature, and BIO15 in correspondence with our results.

Two aspects of our approach were key to minimizing model errors: First, we used seven forest structural metrics to predict $\alpha$-diversity measures, which contributed significantly to the prediction by capturing different aspects of forest structure. Second, integrating LiDAR metrics with temporal SAR and multispectral metrics improved the accuracy of forest structure maps. The smaller numerical scale of Shannon and Simpson indices produced lower errors compared with maps of ENS-Shannon and ENS-Simpson.

Our map estimations of tree $\alpha$-diversity fall within the range of field estimates recorded for the ecoregions and land cover of the study region.
Figure 4. Correlations between Hill’s numbers and richness for the 85 plots (a) and for 10000 random points of the $\alpha$ diversity maps for Chocó (b). Correlations between Shannon index and Simpson index for the 85 plots (c) and for 10000 random points of the $\alpha$ diversity maps for Chocó (d).

(Rangel 2011, Alvarez-Davila et al 2016) and detected the classical pattern of high evenness of tree species with lower dominance of few abundant species (Condit et al 1996, Tovo et al 2017). However, the high tree species turnover (species dissimilarity from 30% to 45% in adjacent hectares and from 85% to 99% in hectares separated by 50 km (Condit et al 2002)) linked to large numbers of tree species in the tropical forests (8000 species are estimated in the Choco alone (WWF 2019)) mean that our tree $\alpha$-diversity maps cannot be interpolated to a different spatial resolution.

Our results add to a growing literature that show strong positive relationships between canopy height, canopy vertical structure, and tree diversity at plot scales for a range of forest types (Dubayah et al 2020, Walter et al 2021). Our study area sits at an extreme end of the tree diversity gradient, recognized as one of the richest forests in the world (Gentry 1986, Keil and Chase 2019). Other studies conducted in temperate ecosystems with far fewer tree species, greater seasonality, lower rainfall, and lower productivity have also observed strong structure diversity relationships (Lopatin et al 2016, Walter et al 2021). This suggests that our approach may be applicable to forest ecosystems with wide range of environmental conditions and biogeographic histories and could yield similarly useful insights on ecoregion scale patterns of tree diversity. Testing the approach in different biogeographic realms would also aid in understanding how historical factors and environmental constraints mediate relationships between structure and diversity.

4.2. Distribution of tree $\alpha$-diversity

Our maps agree with field estimates of tree $\alpha$-diversity in ecoregions and by natural land cover category. This result can be interpreted as indirect validation of map accuracy and as partial support for the harshness hypothesis. The highest $\alpha$-diversity values were in non-flooding areas of EPaM and ChD ecoregions. The highest $\alpha$-diversity was in non-flooding forest followed by non-flooding secondary vegetation. These areas also correspond to tall and heterogeneous canopies. This strong relation between well-drained upland areas, CH, and tree diversity has also been observed in other lowland forest of South America and has been associated with stable environmental conditions for the establishment and growth of trees (Duivenvoorden 1996, Ter Steege et al 2003). Conversely, the lowest $\alpha$-diversity values were found in areas with permanent or seasonal flooding and adverse conditions like high salinity and anoxic soils, such as wetlands or mangroves. These areas also corresponded to shorter and less heterogeneous canopies. Similar results have been recorded in the
Figure 5. (a) Hill’s numbers and (b) diversity indices of trees for the ecoregions of the Chocó region. Species richness (Richness), Shannon’s entropy index (ENS-Shannon), Simpson’s concentration index (ENS-Simpson), Shannon diversity index (H), and Simpson diversity index (D). Ecoregions of the Chocó Chocó–Darien moist forest (ChD), Western Ecuador Moist Forest (WEC), Eastern Panamanian montane forests (EPaM), South American Pacific (SAP) Mangrove, and Amazon Orinoco–Southern Caribbean (AOSC) Mangrove.

other areas of the Amazon (Duivenvoorden 1996) and Chocó (Rangel 2011, Alvarez-Davila et al. 2016). The highly variable conditions of the flooded areas would select for species with a narrower subset of traits than found in the regional pool, thus reducing tree diversity relative to upland forest.

Our maps suggest a zone of highest tree diversity between the Pacific Coast and the wetlands of the Atrato River, from around 5.5–8.6° of longitude. This non-flooding forest zone has the lowest human population density of the Chocó region and roads are absent, conferring it special value in terms of conservation. This zone was identified as the highest biomass forest in the Chocó region (Meyer et al. 2019), increasing its value in terms of co-benefits if preserved. However, the proximity of Quibdó, the capital of Chocó department, and minimal protected area coverage are the main threats to its preservation. The other high diversity zone identified by our maps corresponds to the non-flooding areas along the border between the Chocó region and the Andes. Scholars have highlighted the high biodiversity of the Andes forest/lowland forest ecotone in South America (Dehling et al. 2014).

Although wetlands and mangroves presented lower tree $\alpha$-diversity than the non-flooding forest areas of the Chocó-region, this does not indicate low absolute diversity. Some wetlands of Chocó-region show comparable $\alpha$-diversity values (Shannon index until 3.1) to other tropical vegetation, such as savannas (Morandi et al. 2018). These wetlands are also key habitat for endemic flora and fauna (Rangel 2011), increasing their contribution to regional and global biodiversity.
There are no directly comparable maps of tree $\alpha$-diversity for the Chocó-region. A global tree $\alpha$-diversity model was recently published using a minimum 1 cm DBH threshold to define trees (Keil and Chase 2019), compared to 10 cm in our study. Because a large proportion of diversity in the region is comprised of small stemmed trees (Gentry 1986), the use of different thresholds alone would likely lead to significant differences in diversity estimates between these studies. However, in terms of internal validation metrics, our model is comparable, with a 4.4% relative error computed from five-fold CV compared to $\sim$90% of deviance explained in the global richness model (Keil and Chase 2019) and 80% variance explained in a model of Amazon tree $\alpha$-diversity (Ter Steege et al 2003).

5. Conclusion

We show it is possible to integrate inventories of vegetation and remote sensing data to build accurate maps of tree $\alpha$-diversity in tropical forest regions. The use of LiDAR data was essential for deriving the forest structural metrics that were associated with optical and SAR image data that was, in turn, used to create maps of $\alpha$-diversity. Although we used the forest structural maps to evaluate tree diversity, these maps could be used to evaluate other processes and vegetation traits that are linked to vegetation structure. Furthermore, the variability in tree diversity observed across ecoregions and land covers with our study region provides evidence for interactions between the niche and harshness hypotheses to explain tree diversity.

Figure 6. (a) Hill’s numbers and (b) diversity indices of trees for natural land cover of the Chocó region. Species richness (Richness), Shannon’s entropy index (ENS-Shannon), Simpson’s concentration index (ENS-Simpson), Shannon diversity index ($H$), and Simpson diversity index ($D$). Land cover of the Chocó: well-preserved forests (Forest), secondary vegetation (Secon. veg.), and wetlands.
diversity. While we did not use GEDI LiDAR data, the similarity in spatial resolution between our analysis (50 m) and GEDI (25 m), indicate the potential to produce wall-to-wall structure maps trained on GEDI Lidar data. Thus, our results establish a baseline for predicting and mapping tree α-diversity in additional regions using widely available LiDAR data.

Data availability

Maps of diversity, maps of forest structure, data, and codes that support the findings of this study are openly available at the following URL/DOI: (https://zenodo.org/record/4697615). Likewise, the Landsat imagery processing using Google Earth Engine codes is in the next link (https://code.earthengine.google.com/ed2613adc4c3beb6ad4f02128ec6ee59) and the Sentinel-1 imagery processing using Google Earth Engine codes is in next link (https://code.earthengine.google.com/8ed106adc2976b23224761dca47faa65f).

Acknowledgments

We acknowledge NASA Group on Earth Observations Work Programme grant 80NSSC18K0338 for support of CF and PJ, and NASA GEDI Science Definition Team grant NNL15AA03C for support of PB and SJG. We also acknowledged COLCIENCIAS-Colombia (529–2011) and Fulbright-US for particularly support of CF. We really appreciate the effort and comments done by three hidden reviewers to improve this paper.

APPENDICES

Appendix A. Additional information on study area

The study area of our analysis is called the Chocó-region by Colombian environmental authorities and is considered one of the five primary natural regions of the country (Ideam et al 2016) (figure 1(a)). In terms of the global map of terrestrial ecoregions (Olson et al 2001), 86.5% of this area is considered Chocó–Darien moist forest (59672 km²), 3.4% is considered Western Ecuador Moist Forest in the south (2364.6 km²), and 1.2% is Eastern Panamanian montane forests (871.3 km²) in the north. The Chocó-region also includes the 7.8% of South American Pacific Mangrove (5444.3 km) and 0.8% of Amazon Orinoco–Southern Caribbean Mangrove (574.7 km) (figure 1(b)). In terms of recent land cover, circa 2015 (Fagua and Ramsey 2019), most of the Chocó-region is formed by natural covers such as well-preserved forests (72.2%; 49803 km), secondary vegetation (natural regrowth of vegetation after deforestation: 13.5%; 9535.4 km), wetlands (7.5%; 5213.1 km), and water bodies (3.6%; 2539 km). Anthropogenic land cover, such as grassland, crops and palm plantations, covered less than 1% of the region (figure 1(c)).

The forests of the Chocó-region became separated from the forests of the amazon basin by the uplift of the Andes beginning around 25 million years ago (Gregory-Wodzicki 2000). Consequently, endemic species emerged in a significant pulse of diversification. The Isthmus of Panamá later formed roughly three million years ago, establishing a land bridge for fauna and flora from North and South America (Gentry 1986, WWF W W F 2019) and further enriching regional diversity. The Chocó-region presents a rainfall ranging from 8000 mm to 13000 mm across the region (Poveda and Mesa 2000); the highest rainfalls are recorded in the municipality of Lloró (Chocó Department—Colombia) with 12700 mm in average between 1952 and 1960.

Appendix B. Summary of α-diversity measures, forest structural metrics, and environmental variables

B.1. α-diversity measures

The five measures of α-diversity used in this work are defined as: the three Hill’s numbers (also termed effective number of species (ENS)), where \( q = 0 \) is species richness (richness or \( \alpha \)D), \( q = 1 \) is ENS-Shannon or \( \langle \alpha \rangle \) Hill (1973), Jost (2006). Richness \( \langle 0 \rangle \) is simply the number of species. ENS-Shannon \( \langle 2 \rangle \) weights species in proportion to their frequency and is calculated as \( \exp(\sum p_i \log(p_i)) \). ENS-Simpson \( \langle 2 \rangle \) gives more weight to abundant species and is calculated as \( 1/\sum (p_i)^2 \) (Chao et al 2014). We also estimated the Shannon diversity index \( H \) defined as \( H = – \sum p_i \log(p_i) \) (Shannon 1948, Hill 1973) and the Simpson diversity index \( D \) defined as \( D = 1 – \sum p_i^2 \) (Simpson 1949). In all cases, \( p_i \) is the proportional abundance of the \( i \)th species. We used the R package ‘mobsim’ (May et al 2018) to calculate these metrics of diversity.

B.2. Forest structure metrics

Vertical metrics: to obtain the first set of vertical metrics, the LiDAR surveys were processed by summing the returns within a footprint estimated at 50 m spatial resolution, as the 0.25 hectare plots of vegetation, and 1 m vertical resolution (Dubayah et al 2010, Clark et al 2011). We selected the ground returns and produced 1 m DTMs (digital terrain models) using the k-nearest neighbour method with the inverse-distance
weighting. We then obtained normalized heights for each return of the point cloud and selected the highest returns in the 50 m spatial resolution to compute the vertical metrics as: (a) canopy height (CH) or the cumulative height above ground at the 98th percentile. (b) 50th percentile of heights (z_50) or the cumulative height above ground at the 50th percentile (also referred to as height of median energy). (c) vertical canopy heterogeneity (VCH) expressed as \(-\sum p_i \log(p_i)\), where \(p_i\) was the sum of the returns in a voxel calculated at 50 m spatial resolution and 1 m vertical resolution. (d) Standard deviation of normalized heights (z_sd) where the standard deviation was calculated in voxels of 50 m spatial resolution and 1 m vertical resolution (Davies and Asner 2014, Zhang et al 2017). We also estimated the coefficient of variation of normalized heights (z_cv) from the LiDAR data, but we eliminated this metric from our analysis because we could not model the z_cv spatially. Horizontal metrics: to obtain these metrics, we initially created 1 m CH models using the previous procedure. We used the 1 m CH model to calculate three horizontal metrics as: (a) standard deviation (CH_sd) or standard deviation of the pixel values. (b) Horizontal heterogeneity (CH_H) expressed as \(-\sum v_i \log(v_i)\), where \(v_i\) was the proportion of the pixel value. (c) Rumple Index (CH_R) that computes the roughness of CH as the ratio between its area and its projected area on the ground using Jenness’s algorithm (Jenness 2004). The R packages ‘Raster’ and ‘lidR’ were used to implement these procedures (Hijmans et al 2016, Roussel et al 2019).

B.3. Environmental variables
The 76 environmental variables were obtained and processed thus. (a) The 19 WorldClim variables (BIO1, BIO1… and BIO19) were used without modification (Hijmans et al 2005). (b) The annual mean temperature between 2001 and 2019 (\(T_{em}\)) was obtained from the MODIS product MOD11A1 v006 (MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global) (Wan et al 2015). This product provides daily land surface temperature and emissivity in °K at 1 km spatial resolution. We built mosaics of annual mean values using Google Earth Engine (Gorelick et al 2017) and average these daily mosaics to obtain a final mosaic. (c) We obtained altitude (Alt) from the SRTM-30 digital elevation model (NASA Shuttle Radar Topographic Mission) (Jarvis et al 2008). (d) Photosynthetically Active Radiation (FPAR) was obtained from the MODIS product MOD15A2H (Myneni and Knyazikhin 2015); an 8 days composite dataset at 500 m resolution that chooses the best pixel available from all the acquisitions of the Terra sensor from within the 8 days period. We built annual mean mosaics of this product corresponding to the years of the vegetation surveys of each 0.5 hectare plot, using Google Earth Engine. (e) The soil variables were obtained from the SoilGrids250m (Hengl et al 2017). This soil product offers nine main variables at six soil depths (0 cm, 10 cm, 30 cm, 60 cm, 100 cm, 200 cm): Nitrogen content (N) in cg kg\(^{-1}\), percentage of silt content (Silt), percentage of sand content (Sand), percentage of clay content (Clay), soil pH (pH), soil organic carbon content (C) in 5 g kg\(^{-1}\), soil bulk density (Bulk) in 10 kg m\(^{-3}\), and soil water content (Water) at 33 kPa of field capacity.

Appendix C. Description of the LiDAR instruments
The LiDAR instruments used in the eighty-five 0.25 hectare plots across tropical forest of South America (figure 1(b)) and the Chocó region (figure 1(d)) were similar models of Teledyne Optech: (a) ALTM 3100 (Optech Inc.) instruments were used in 44 plots, (b) ALTM Orion M-20 instruments (Optech Inc.) in 31 plots, and (c) ALTM 3033 (Optech Inc.) instruments in the other ten plots. The flights were between 700 and 1000 m above ground and had a swath side lap of 65%. A minimum return density (\(\geq 4 \text{ m}^2 \text{ over } 99.5\% \text{ of the study area}\)) was required to avoid inconsistencies and biases in estimated forest structural metrics associated with low return density (Jakubowski et al 2013, Longo et al 2016). These LiDAR data have been used previously in other studies (EMBRAPA 2019, Meyer et al 2019). The table below shows the details for the LiDAR instrument by regions.
Appendix D. Selection of the machine learning algorithm.

We evaluate five different regression algorithms to model the relationship between the \( \alpha \)-diversity measures and the 77 predictors (seven forest structural metrics and seventy environmental variables). These algorithms use different statistical approaches to predict a response variable: (a) Lasso and Elastic-Net Regularized Generalized Linear Models (GLM.net), (b) Random Forest (RF), (c) support vector machine (SVM), (d) Multivariate Adaptive Regression Splines (MARS), and (e) bagged regression tree (BRT). We used root mean squared error (RMSE) and mean absolute error (MAE) to evaluate the error of the models. RMSE and MAE express average model prediction error in units of the variable of interest; however, RMSE gives a relatively high weight to larger errors and MAE penalizes greater variances of the errors. We performed 100 iterations for each of the five regression models using the \textit{R} \texttt{Package} \text{‘caret’} (Kuhn et al 2019). Each iteration included a five-fold cross-validation to estimate RMSE and MAE. We tuned the parameters of each regression algorithms to obtain the best predicted models using the train function of the \textit{caret} \texttt{R} \texttt{package} and following Boehmke and Greenwell (2019) (Boehmke and Greenwell 2019). RF tended to produce the lowest errors (RMSE and MAE) for the five diversity measures: the Hill’s numbers (Richness, ENS_Shannon and ENS_Simpson), the Shannon diversity index, and the Simpson diversity index (figure D1).

For GLM.net, we tuned \( \lambda \) and \( \alpha \). For RF, we selected the predictors that fit the best models using the algorithm Boruta and tuned the number of trees to grow (from 300 to 1000) as well as the number of variables randomly sampled as candidates at each split of each tree. For SVM, we assessed three kernel types and tuned their parameters: Liner kernel (cost was tuned), polynomial kernel (cost and degree were tuned) and radial kernel (cost and sigma were tuned). Liner kernel produced the best models; thus, we only include their results in this appendix. For MARS (an extension of linear regression that captures nonlinearities and interactions between variables), we tuned the maximum number of terms (including intercept) in the pruned model and the degree. For BRT, we excluded highly correlated predictors based on their variance inflation factors (VIFs).

Appendix E. Building multispectral and SAR mosaics.

We built: (a) ten temporal mean (\textit{mean-X}) and ten standard deviation (\textit{sd-X}) mosaics for seven bands and three indices using Landsat data; (b) four \textit{mean-X} and four \textit{sd-X} mosaics of backscattering coefficients of the C-band using Sentinel-1 data; and (c) two \textit{mean-X} mosaics of backscattering coefficients of the L-band using ALOS-2-PALSAR data. These mosaics were built in Google Earth Engine (Gorelick et al 2017). For the Landsat data, we first utilized the product USGS Landsat-8 Surface Reflectance Tier-1 to estimate temporal mean-\( X \) and temporal 5d-\( X \) mosaics. This products is atmospherically corrected using LaSRC (Landsat 8 Surface Reflectance algorithm) that includes cloud, shadow, water and snow masks (Usgs 2019). We only included confident pixel-values (low cloud and atmospheric distortion) using a mask based on the pixel quality attributes generated from the CFMASK correction.
algorithm (pixel_qa values) and the radiometric saturation pixel (radsat_qa values). The mosaics were resampled to 50 m using bilinear interpolation to match the spatial resolution of the 0.25 hectare plots. To reduce the blanks in the previous Landsat-8 mosaics, a second mosaics was built combining two Landsat products to fill these blanks: USGS Landsat-8 Surface Reflectance Tier-1 and USGS Landsat-7 Surface Reflectance Tier 1 collections. In this second mosaics, the Landsat-8 pixel-values were rescaled to the Landsat-7 pixel-values to estimate mean-\(X\) and Sd-\(X\) mosaics following (Roy et al 2016). For the Sentinel-1 data, we used the Sentinel-1 SAR GRD product, which offers calibrated and ortho-corrected data (ESA 2019). We applied an angular-based radiometric slope correction to Sentinel-1 data following (Vollrath et al 2020), using the correction model suggested for vegetation applications and the ALOS World 3D global digital surface model to estimate surface slope. The spatial resolution of this Sentinel-1 product is 15 m which we resampled to 50 m using bilinear interpolation. For the ALOS-2-PALSAR data, we used the years 2017 and 2018 of the Global PALSAR-2/PALSAR Yearly Mosaic (Shimada et al 2014) to estimate X-means. We did not estimate X-Sds because this is an annual product with and only two years of data were available. The spatial resolution of this product is 25 m which we resampled to 50 m using bilinear interpolation.

Figure D1. Errors of five algorithms used to estimate five \(\alpha\)-diversity measures. Hill’s numbers (Richness, ENS_Shannon and ENS_Simpson) are shown in (a) and (c) while Shannon diversity index and the Simpson diversity index are shown in (b) and (d).
Appendix F

Figure F. Importance of predictors of Random Forest regressions for five tree-diversity metrics: (a) richness, (b) ENS-Shannon, (c) ENS-Simpson, (d) H, and (e) D. Blue lines show the importance range for the predictors after 100 iterations where each iteration was a RF regression. The magenta dash line indicates the third quantile or Q3 (75%) for the importance values.

Appendix G

Table G. Correlations (Pearson-r) between tree α-diversity measures against forest structural metrics and environmental variables. Significant P values range; P < 0.001 (***) , P < 0.01 (**) , and P < 0.05 (*). Bold highlights significant P values. Variable acronyms were described previously.

|                      | Richness  | ENS-Shannon | ENS-Simpson | H       | D       |
|----------------------|-----------|-------------|-------------|---------|---------|
| Forest structure     |           |             |             |         |         |
| CH                   | 0.96***   | 0.96***     | 0.90***     | 0.96*** | 0.94*** |
| VCH                  | 0.97***   | 0.93***     | 0.86***     | 0.95*** | 0.94*** |
| z_50                 | 0.85***   | 0.88***     | 0.88***     | 0.79*** | 0.73*** |
| z_sd                 | 0.95***   | 0.94***     | 0.85***     | 0.95*** | 0.92*** |
| CH_sd                | 0.80***   | 0.77***     | 0.65***     | 0.87*** | 0.89*** |
| CH_ent               | 0.94***   | 0.88***     | 0.79***     | 0.96*** | 0.97*** |
| CH_R                 | 0.93***   | 0.86***     | 0.73***     | 0.94*** | 0.95*** |
| Incident radiation   |           |             |             |         |         |
| FPAR                 | 0.17      | 0.15        | 0.08        | 0.34*** | 0.41*** |
| Climate              |           |             |             |         |         |
| Tem                  | -0.50***  | -0.36***    | -0.29**     | -0.41***| -0.42***|
| BIO1                 | -0.14     | -0.09       | 0.04        | -0.25*  | -0.32** |
| BIO2                 | -0.20     | -0.20       | -0.26*      | -0.05   | 0.02    |
| BIO3                 | -0.62***  | -0.71***    | -0.72***    | -0.55***| -0.46***|
| BIO4                 | 0.49***   | 0.58***     | 0.59***     | 0.41*** | 0.32*** |
| BIO5                 | -0.28**   | -0.16       | -0.12       | -0.12   | -0.09   |
| BIO6                 | 0.06      | 0.07        | 0.16        | -0.08   | -0.15   |
| BIO7                 | -0.10     | -0.09       | -0.16       | 0.05    | 0.10    |
| BIO8                 | -0.18     | -0.15       | -0.02       | -0.28** | -0.34***|
| BIO9                 | 0.11      | 0.15        | 0.23*       | -0.03   | -0.11   |
| BIO10                | -0.01     | 0.06        | 0.18        | -0.13   | -0.21   |
| BIO11                | -0.17     | -0.14       | -0.01       | -0.28** | -0.34***|
| BIO12                | 0.01      | -0.13       | -0.14       | -0.09   | -0.07   |
| BIO13                | -0.13     | -0.28**     | -0.30**     | -0.24*  | -0.21   |
| BIO14                | 0.05      | -0.08       | -0.09       | -0.07   | -0.06   |

(Continued.)
| Richness  | ENS-Shannon  | ENS-Simpson  | $H$  | $D$  |
|----------|--------------|--------------|------|------|
| BIO15    | $-0.29^{**}$ | $-0.20$      | $-0.22$ | $-0.20$ | $-0.20$ |
| BIO16    | $-0.12$ | $-0.26^*$ | $-0.28^{**}$ | $-0.23^*$ | $-0.20$ |
| BIO17    | $0.06$ | $-0.07$ | $-0.08$ | $-0.05$ | $-0.04$ |
| BIO18    | $0.04$ | $-0.10$ | $-0.12$ | $-0.03$ | $0.01$ |
| BIO19    | $-0.33^{***}$ | $-0.44^{***}$ | $-0.42^{**}$ | $-0.39^{***}$ | $-0.36^{***}$ |

**Table G. (Continued.)**

| Topography | Alt | 0.36^{***} | 0.33^{***} | 0.21 | 0.48^{***} | 0.53^{***} |

| Soil | N$_0$ cm | 0.06 | $-0.10$ | $-0.15$ | $-0.02$ | $0.02$ |
| N$_{10}$ cm | 0.14 | $-0.01$ | $-0.05$ | $0.04$ | $0.06$ |
| N$_{30}$ cm | 0.10 | $-0.05$ | $-0.09$ | $-0.02$ | $0.00$ |
| N$_{60}$ cm | 0.07 | $-0.10$ | $-0.14$ | $-0.04$ | $-0.02$ |
| N$_{100}$ cm | 0.02 | $-0.13$ | $-0.16$ | $-0.08$ | $-0.06$ |
| N$_{200}$ cm | 0.04 | $-0.12$ | $-0.16$ | $-0.07$ | $-0.05$ |
| Silt$_0$ cm | $-0.23^*$ | $-0.39^{***}$ | $-0.45^{***}$ | $-0.21$ | $-0.12$ |
| Silt$_{10}$ cm | $-0.25^*$ | $-0.41^{***}$ | $-0.48^{***}$ | $-0.23^*$ | $-0.14$ |
| Silt$_{30}$ cm | $-0.22$ | $-0.39^{***}$ | $-0.47^{***}$ | $-0.20$ | $-0.11$ |
| Silt$_{60}$ cm | $-0.24^*$ | $-0.41^{***}$ | $-0.48^{***}$ | $-0.21$ | $-0.12$ |
| Silt$_{100}$ cm | $-0.21$ | $-0.38^{***}$ | $-0.46^{***}$ | $-0.18$ | $-0.09$ |
| Silt$_{200}$ cm | $-0.22^*$ | $-0.39^{***}$ | $-0.47^{***}$ | $-0.19$ | $-0.10$ |
| Sand$_0$ cm | $-0.61^{***}$ | $-0.39^{***}$ | $-0.18$ | $-0.51^{***}$ | $-0.55^{***}$ |
| Sand$_{10}$ cm | $-0.60^{***}$ | $-0.39^{***}$ | $-0.17$ | $-0.49^{***}$ | $-0.53^{***}$ |
| Sand$_{30}$ cm | $-0.61^{***}$ | $-0.40^{***}$ | $-0.17$ | $-0.50^{***}$ | $-0.54^{***}$ |
| Sand$_{60}$ cm | $-0.52^{***}$ | $-0.29^{***}$ | $-0.05$ | $-0.42^{***}$ | $-0.48^{***}$ |
| Sand$_{100}$ cm | $-0.54^{***}$ | $-0.33^{***}$ | $-0.09$ | $-0.45^{***}$ | $-0.50^{***}$ |
| Sand$_{200}$ cm | $-0.57^{***}$ | $-0.36^{***}$ | $-0.12$ | $-0.48^{***}$ | $-0.52^{***}$ |
| pH$_0$ cm | $-0.43^{***}$ | $-0.57^{***}$ | $-0.59^{***}$ | $-0.39^{***}$ | $-0.30^{***}$ |
| pH$_{10}$ cm | $-0.41^{***}$ | $-0.55^{***}$ | $-0.58^{***}$ | $-0.37^{***}$ | $-0.28^{***}$ |
| pH$_{30}$ cm | $-0.33^{***}$ | $-0.48^{***}$ | $-0.52^{***}$ | $-0.29^{***}$ | $-0.19$ |
| pH$_{60}$ cm | $-0.22$ | $-0.38^{***}$ | $-0.43^{***}$ | $-0.18$ | $-0.08$ |
| pH$_{100}$ cm | $-0.22$ | $-0.37^{***}$ | $-0.41^{***}$ | $-0.17$ | $-0.07$ |
| pH$_{200}$ cm | $-0.20$ | $-0.35^{***}$ | $-0.39^{***}$ | $-0.14$ | $-0.04$ |
| C$_0$ cm | 0.18 | 0.01 | $-0.05$ | 0.07 | 0.09 |
| C$_{10}$ cm | 0.17 | $-0.01$ | $-0.07$ | 0.07 | 0.10 |
| C$_{30}$ cm | 0.21 | 0.04 | $-0.04$ | 0.09 | 0.11 |
| C$_{60}$ cm | 0.10 | $-0.06$ | $-0.10$ | $-0.02$ | $-0.01$ |
| C$_{100}$ cm | 0.06 | $-0.09$ | $-0.13$ | $-0.06$ | $-0.05$ |
| C$_{200}$ cm | 0.22 | 0.10 | 0.08 | 0.09 | 0.08 |
| Clay$_0$ cm | $0.76^{***}$ | $0.69^{***}$ | $0.54^{***}$ | $0.65^{***}$ | $0.62^{***}$ |
| Clay$_{10}$ cm | $0.75^{***}$ | $0.68^{***}$ | $0.54^{***}$ | $0.63^{***}$ | $0.59^{***}$ |
| Clay$_{30}$ cm | $0.75^{***}$ | $0.69^{***}$ | $0.54^{***}$ | $0.64^{***}$ | $0.60^{***}$ |
| Clay$_{60}$ cm | $0.69^{***}$ | $0.64^{***}$ | $0.48^{***}$ | $0.58^{***}$ | $0.55^{***}$ |
| Clay$_{100}$ cm | $0.68^{***}$ | $0.64^{***}$ | $0.49^{***}$ | $0.57^{***}$ | $0.54^{***}$ |
| Clay$_{200}$ cm | $0.69^{***}$ | $0.65^{***}$ | $0.51^{***}$ | $0.58^{***}$ | $0.54^{***}$ |
| Bulk$_0$ cm | $-0.52^{***}$ | $-0.47^{***}$ | $-0.47^{***}$ | $-0.43^{***}$ | $-0.41^{***}$ |
| Bulk$_{10}$ cm | $-0.56^{***}$ | $-0.51^{***}$ | $-0.52^{***}$ | $-0.47^{***}$ | $-0.44^{***}$ |
| Bulk$_{30}$ cm | $-0.58^{***}$ | $-0.54^{***}$ | $-0.54^{***}$ | $-0.49^{***}$ | $-0.46^{***}$ |
| Bulk$_{60}$ cm | $-0.58^{***}$ | $-0.55^{***}$ | $-0.56^{***}$ | $-0.48^{***}$ | $-0.44^{***}$ |
| Bulk$_{100}$ cm | $-0.49^{***}$ | $-0.50^{***}$ | $-0.54^{***}$ | $-0.38^{***}$ | $-0.32^{***}$ |
| Bulk$_{200}$ cm | $-0.49^{***}$ | $-0.52^{***}$ | $-0.58^{***}$ | $-0.39^{***}$ | $-0.32^{***}$ |
| W$_0$ cm | $0.36^{***}$ | $0.23^*$ | 0.13 | $0.41^{***}$ | $0.48^{***}$ |
| W$_{10}$ cm | 0.23 | 0.07 | $-0.08$ | $0.28^{***}$ | $0.37^{***}$ |
| W$_{30}$ cm | 0.08 | $-0.08$ | $-0.23^*$ | 0.14 | $0.25^*$ |
| W$_{60}$ cm | $0.34^{***}$ | 0.16 | $-0.02$ | $0.39^{***}$ | $0.48^{***}$ |
| W$_{100}$ cm | $0.42^{***}$ | $0.25^*$ | 0.07 | $0.47^{***}$ | $0.56^{***}$ |
| W$_{200}$ cm | $0.41^{***}$ | 0.21 | 0.02 | $0.41^{***}$ | $0.50^{***}$ |
Appendix H

Table H. Error estimation of random forest regressions to predict seven structural forest variables estimated from 578.25 km² of LiDAR surveys using 67654 training sites. The vertical structural forest variables were Canopy Height (CH), Vertical Canopy Heterogeneity (VCH), 50th percentile of heights (z50), and standard deviation of normalized heights (zSD). The horizontal structural forest variables were standard deviation (CHsd), horizontal heterogeneity (CHent) and Rumple Index (CHR). For each error estimation method, cross validation (CV) and data partition (DP), we performed 100 iterations where each iteration estimated RMSE, MAE, and RE. Mean and Standard deviation (Sd) of each group of 100 iterations are showed in the table.

|                  | Cross validation (CV) | Partition 30/70 (DP) |
|------------------|-----------------------|----------------------|
|                  | RMSE Sd | MAE Sd | RE% Sd | RMSE Sd | MAE Sd | RE% Sd |
| CH               | 7.83 ± 0.39 | 3.82 ± 0.1 | 24.5 ± 11.7 | 8.8 ± 0.27 | 4.13 ± 0.06 | 24.9 ± 5.9 |
| CH sd            | 4.7 ± 0.79 | 2.64 ± 0.02 | 35.1 ± 20 | 4.85 ± 0.75 | 2.71 ± 0.02 | 36.5 ± 5.7 |
| zSD              | 0.23 ± 0.01 | 0.13 ± 0.00 | 6.5 ± 1.7 | 0.25 ± 0.00 | 0.14 ± 0.00 | 10.1 ± 3.8 |
| VCH              | 3.3 ± 1.72 | 1.05 ± 0.03 | 19.7 ± 4.29 | 3.52 ± 0.99 | 1.14 ± 0.02 | 23.7 ± 10.9 |
| CH sd            | 1.59 ± 0.00 | 1.16 ± 0.00 | 34.4 ± 4.29 | 1.7 ± 0.00 | 1.23 ± 0.00 | 38.7 ± 6.32 |
| CH_H             | 0.37 ± 0.00 | 0.25 ± 0.00 | 18.4 ± 9.05 | 0.38 ± 0.00 | 0.27 ± 0.00 | 19.6 ± 6.5 |
| CH_R             | 0.64 ± 0.00 | 0.46 ± 0.00 | 22.3% ± 8.6 | 0.67 ± 0.00 | 0.48 ± 0.00 | 21.4% ± 6.2 |

Appendix I

Figure I. Importance of predictors for the random forest regressions for seven forest structural metrics: (a) canopy height (CH) map, (b) 50th percentile of heights (z50) map, (c) vertical Canopy Heterogeneity (VCH) map, (d) standard deviation of normalized heights (zsd) map, (e) horizontal standard deviation of CH (CHsd), (f) horizontal heterogeneity of CH (CHH) and (g) rumple index of CH (CHR). Blue lines show the importance range for the predictors after 100 iterations where each interaction was a RF regression. The magenta dashed line indicates the third quantile or Q3(75%) for the importance values.
References

Allouche O, Kahuzi M, Moreno-Rueda G, Pizarro M and Kadmon R 2012 Area-heterogeneity tradeoff and the diversity of ecological communities Proc. Natl Acad. Sci. USA 109 17495–500

Alvarez-Davila E, Jaramillo-Giraldo G C, Cesar Cogollo-Rivera C, Martinez-Higuera H, Rojas E and Fernandez-Mendez F 2016 Structure and diversity of the three plant associations in the San Juan river delta, Choco, Colombia Rev. Avrere 40 833–43

Anderson-Teixeira K J et al 2015 CTFS-ForestGEO A Worldwide Network Monitoring Forests in an Era of Global Change Glob. Chang. Biol. 21 528–49

Asner G P et al 2010 High-resolution forest carbon stocks and emissions in the Amazon Proc. Natl Acad. Sci. USA 107 16738–48

Asner G P et al 2012 High-resolution mapping of forest carbon stocks in the Colombian Amazon Biogesioscience 9 2683–96

Asner G P et al 2013 High-fidelity national carbon mapping for resource management and REDD+ Carbon Balance Manage. 8 7

Bae S et al 2019 Radar vision in the mapping of forest biodiversity from space Nat. Commun. 10 1–10

Bakx T R M, Koma Z, Seijmonsbergen A C and Kissling W D 2019 Radar vision in the mapping of forest biodiversity from space Nat. Commun. 10 1–10

Barloto C, Molto Q, Rabaud S, Herault B, Valencia R, Blanc L, Fine P V A and Thompson J 2013 Rapid simultaneous estimation of aboveground biomass and tree diversity across neotropical forests: a comparison of field inventory methods Biotropica 45 288–98

Barthlott W, Hostert A, Kier G, Koper W, Kreft H, Matusek J, Radsigpoor M D and Sommer J H 2007 Geographic patterns of vascular plant diversity at continental to global scales Erdkunde 61 305–15

Barthlott W, Lauer W and Placke A 1996 Global distribution of species diversity in vascular plants: towards a world map of phytodiversity (Globale Verteilung der Artenvielfalt Höherer Pflanzen: vorarbeiten zu einer Weltkarte der Phytodiversität) Erdkunde 50 317–27

Ben-hur E and Kadmon R 2020 Heterogeneity–diversity relationships in sessile organisms A Unified Framework Ecol. Lett. 23 193–207

Boehmke B and Greenwell B M 2019 Hands-On Machine Learning with R (available at: https://bradleyboehmke.github.io/HOML/)

Brando P M et al 2014 Abrupt increases in Amazonian tree mortality due to drought–fire interactions Proc. Natl Acad. Sci. 111 6347–52

Carlson K M, Asner G P, Hughes R F, Ostertag R and Martin R E 2007 Hyperspectral remote sensing of canopy biodiversity in Hawaiian Lowland rainforests Ecosystems 10 536–49

Carlson T N and Ripley D A 1997 On the relation between NDVI, fractional vegetation cover, and leaf area index Remote Sens. Environ. 62 241–52

Carter G A, Knapp A K, Anderson J E, Hoch G A and Smith M D 2005 Indicators of plant species richness in AVIRIS spectra of a mesic grassland Remote Sens. Environ. 98 304–16

CBD C on B D 2019 Aichi biodiversity targets (available at: www.cbd.int/sp/targets/)

Chao A, Gotelli N J, Hsieh T C, Sander E L, Ma K H, Colwell R K and Ellison A M 2014 Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies Ecol. Monogr. 84 45–67

Clark M L, Roberts D A, Ewel J J and Clark D B 2011 Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors Remote Sens. Environ. 115 2931–42

Condit R et al 2002 Beta-diversity in tropical forest trees Science 295 666–9

Condit R, Hubbell S P and Foster R B 1992 Short-term dynamics of a neotropical forest Biocience 42 822–8

Condit R, Hubbell S P and Foster R B 1996 Changes in tree species abundance in a Neotropical forest: impact of climate change J. Trop. Ecol. 12 231–56

Davidson E A et al 2012 The Amazon basin in transition Nature 481 321

Davies A B and Asner G P 2014 Advances in animal ecology from 3D-LiDAR ecosystem mapping TRENDS Ecol. Evol. 29 681–91

Dehling D M, Fritz S A, Toepfer T, Pauckert M, Estler P, Boehning-Gaese K and Schleuning M 2014 Functional and phylogenetic diversity and assemblage structure of frugivorous birds along an elevational gradient in the tropical Andes Ecography 37 1047–55

Dubayah R O, Sheldon S L, Clark D B, Hofton M A, Blair J B, Hurtt G C and Chazdon R L 2010 Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica J. Geophys. Res. 115 1–17

Dubayah R et al 2020 The global ecosystem dynamics investigation: high-resolution laser ranging of the Earth’s forests and topography Sci. Remote Sens. 1 1–14

Duivenvoorden J F 1996 Patterns of tree species richness in rain forests of the middle Caqueta area, Colombia, NW Amazonia Biotropica 28 142–58

Duque A, Saldarriaga J, Meyer V and Saatchi S 2017 Structure and allometry in tropical forests of Choco, Colombia For. Ecol. Manage. 485 309–16

EMBRAPA 2019 Paisagens Sustentáveis Brasil (available at: www.paisagenslidar.cnptia.embrapa.br/webgis/)

ESA European Space Agency 2019 Sentinel-1 SAR GRD: c-band synthetic aperture radar ground range detected, log scaling Earth Engine Data Cat. (available at: https://developers.google.com/earth-engine/datasets/catalog/COBeRNICUS_S1_GRD)

Fagiu L C, Baggio J A and Ramsey R D 2019a Drivers of forest cover changes in the Chocó–Darien global ecoregion of South America Ecosystems 10 e02648

Fagiu L C, Jantz P, Rodriguez-Buritica S, Duncanson L and Goetz S J 2019b Integrating LiDAR, multispectral and SAR data to estimate and map canopy height in tropical forests Remote Sens. 11 1–20

Fagiu L C and Ramsey R D 2019 Geospatial modeling of land cover change in the Chocó–Darien global ecoregion of South America; one of most biodiverse and rainy areas in the world PLoS One 14 e0211324

Finegan B 1996 Pattern and process in neotropical secondary rain forests: the first 100 years of succession TRENDS Ecol. Evol. 11 119–24
Foody G M and Cutler M E J 2003 Tree biodiversity in protected and logged Bornean tropical rain forests and its measurement by satellite remote sensing J. Biogeogr. 30 1053–66

Fricke G A, Wolf J A, Saatchi S S and Gillespie T W 2015 Predicting spatial variations of tree species richness in tropical forests from high-resolution remote sensing Ecol. Appl. 25 1776–89

Gatti R C, Di Paola A, Bombelli A, Noce S and Valentini R 2017 Exploring the relationship between canopy height and terrestrial plant diversity Plant Ecol. 218 899–908

Gentry A H 1986 Species richness and floristic composition of Choco region plant communities Caldasia 15 5

Gentry A H 1988 Tree species richness of upper Amazonian forests Proc. Natl Acad. Sci. 85 156–9

George-Chacon S P, Dupuy J M, Peduzzi A and Hernandez-Stefanoni J L 2019 Combining high resolution satellite imagery and lidar data to model woody species diversity of tropical dry forests Ecol. Indic. 101 975–84

Goetz S J, Hansen M, Houghton R A, Walker W, Laporte N and Busch J 2015 Measurement and monitoring needs, capabilities and potential for addressing reduced emissions from deforestation and forest degradation under REDD Environ. Res. Lett. 10

Goetz S, Steinberg D, Dubayah R and Blair B 2007 Laser remote sensing of canopy heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA Remote Sens. Environ. 108 254–63

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D and Moore R 2017 Google Earth Engine: planetary-scale geographic analysis for everyone Remote Sens. Environ. 202 18–27

Gould W 2000 Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots Ecol. Appl. 10 1861–70

Gregory-Wodzicki K M 2000 Uplift history of the central and South-east Brazil Ecol. Appl. 10 218–27

Guo Q, Kelt D A, Sun Z, Liu H, Ren H and Wen J 2013 DEM generation from LiDAR and hyperspectral remote sensing Int. J. Remote Sens. 35 927–48

Guariguata M R and Ostertag R 2001 Neotropical secondary forest succession: changes in structural and functional characteristics For. Ecol. Manage. 148 185–206

Guo Q, Kelt D A, Sun Z, Liu H, Ren H and Wen J 2013 Global variation in elevational diversity patterns Sci. Rep. 3 1–7

Hakkenberg C R, Peet R K, Urban D L and Song C 2018a Modeling plant composition as community continua in a forest landscape with LiDAR and hyperspectral remote sensing Ecol. Appl. 28 177–90

Hakkenberg C R, Zhu K, Peet R K and Song C 2018b Mapping multi-scale vascular plant richness in a forest landscape with integrated LiDAR and hyperspectral remote-sensing Ecology 99 474–87

Hawkins B A et al 2003 Energy, water, and broad-scale geographic patterns of species richness Ecology 84 3105–17

Healey S P, Yang Z, Gorelick N and Ilyushchenko S 2020 Highly local model calibration with a new GEDE LiDAR asset on Google Earth Engine reduces landsat forest height signal saturation Remote Sens. 12 1–10

Hengl T et al 2017 SoilGrids250m: global gridded soil information based on machine learning Plos One 12 1–14

Hernández-Stefanoni J L, Gallardo-Cruz J A, Meave J A, Rocchini D, Bello-Pineda J and Martínez J O 2012 Modeling α- and β-diversity in a tropical forest from remotely sensed and spatial data Int. J. Appl. Earth Obs. Geoinf. 19 359–68

Higuchi P, Oliveira-Filho A T, Bebbner D P, Brown N D, Silva A C and Machado E L M 2008 Spatio-temporal patterns of tree community dynamics in a tropical forest fragment in South-east Brazil Plant Ecol. 199 125–35

Hijmans R J, Cameron S E, Parra J L, Jones P G and Jarvis A 2005 Very high resolution interpolated climate surfaces for global land areas Int. J. Climatol. 25 1965–78

Hijmans R, Van E J, Cheng J, Mattiuzzi M, Summer M, Greenberg J, Perpinan O, Bevan A, Racine E and Shortridge A 2016 Package ‘raster’ (r-project.org)

Hill M O 1973 Diversity and evenness: a unifying notation and its consequences Ecology 54 427–32

Hirata A, Kamijo T and Saito S 2009 Host trait preferences and distribution of vascular epiphytes in a warm-temperate forest Plant Ecol. 201 247–54

Huang W, Dolan K, Swatantrad A, Johnson K, Tang H, O’Neill-Dunne J, Dubayah R and Hurtt G 2019 High-resolution mapping of aboveground biomass for forest carbon monitoring system in the Tri-State region of Maryland, Pennsylvania and Delaware, USA Environ. Res. Lett. 14 1–16

Hubbell S P 2013 Tropical rain forest conservation and the twin challenges of diversity and rarity Ecol. Evol. 3 3263–74

Hutchinson G E 1957 Concluding remarks cold spring Harth. Symp. Quant. Biol. 22 413–27

Ideam I D H M Y E A, Invermar I D I M Y C J B V D A, Iiap I D I M Casallas, I D P Guardiola and Y A Moreno (Bogota DCC IDEAM) (available at: www.andi.com.co/Uploads/Informe.del.estado.del.ambiente.y.los.recursos.naturales.renovables.pdf)

Jakubowski M K, Guo Q and Kelly M 2013 Tradeoffs between lidar pulse density and forest measurement accuracy Remote Sens. Environ. 130 245–53

Jarris A, Reuter H, Nelson A and Guevara E 2008 Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (available at: http://srtm.csi.cgiar.org/)

Jenness J S 2004 Calculating landscape surface area from digital elevation models Wildl. Soc. Bull. 32 829–39

Jost L 2006 Entropy and diversity Oikos 113 363–75

Kelly P and Chase J M 2019 Global patterns and drivers of tree diversity integrated across a continuum of spatial grains Nat. Ecol. Evol. 3 390–9

Kessler M et al 2009 Alpha and beta diversity of plants and animals along a tropical land-use gradient Ecol. Appl. 19 2142–56

Kraft N J B, Adler P B, Godoy O, James E C, Fuller S and Levine J M 2015 Community assembly, coexistence and the environmental filtering metaphor Funct. Ecol. 29 592–9

Kuhn M et al 2019 Package ‘caret’: classification and regression training (available at: https://github.com/topepo/caret/)

Kursa M B and Rudnicki W R 2010 Feature selection with the ensemble approach J. Mach. Learn. Res. 11 157–80

Kwok Q, Provini S, Wang P and Lambin E F 2010 Geospatial analysis of tree species richness and forest cover loss in an agricultural frontier landscapes: a case study in Amazonia J. Vég. Sci. 21 1159–72

Laurens S W G, Laurance W F, Andrade A, Fearnside P M, Harms K E, Vicentini A and Luizao R C C 2010 Influence of soils and topography on Amazonian tree diversity: a landscape-scale study J. Veg. Sci. 21 96–106

Laurens W F, Camargo J L G, Fearnside P M, Lovejoy T E, Williamson G B, Mesquita R C G, Meyer C F J, Bobrowiec P E D and Laurance S G W 2010 An Amazonian rainforest in the making: spatial analyses as a laboratory of global change Biol. Rev. 93 231–47

Laurans M, Herault A, Vicentini G and Vincent G 2014 Vertical stratification reduces competition for light in dense tropical forests For. Ecol. Manage. 329 79–88

Liao A 2018 Package ‘randomForest’: breiman and cutler’s random forests for classification and regression (Berkeley, CA) (available at: www.stat.berkeley.edu/~breiman/)

Longo M, Keller M, Dos-santos M N, Leitold V, Pinage É R, Baccini A, Saatchi S, Nogueira E M, Batistella M and Morton D C 2016 Aboveground biomass variability across intact and degraded forests in the Brazilian Amazon Global Biogeochem. Cycles 30 1639–60

Lopatin J, Dolos K, Hernández H J, Galleguillos M and Fassnacht F E 2016 Comparing generalized linear models
and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile Remote Sens. Environ. 173 200–100

Lundholm J T 2009 Plant species diversity and environmental heterogeneity: spatial scale and competing hypotheses J. Veg. Sci. 20 577–91

MacArthur R H and MacArthur J W 1961 On bird species diversity Ecology 42 594–8

Maia V, De Souza C B, De Aguiar-campos N, Fagundes N C. A, Santos A B M, De Paula G P G, Santos P F, Silva W B, De Oliveira Menino G C and Dos Santos R M 2020 Interactions between climate and soil shape tree community assembly and above-ground woody biomass of tropical dry forests For. Ecol. Manage. 474 118348

Marks C O, Muller-Landau H C and Tilman D 2016 Tree diversity, tree height and environmental harshness in eastern and western North America Ecol. Lett. 19 743–51

Marselis S M et al 2019 Exploring the relation between remotely sensed vertical canopy structure and tree species diversity in Gabon Environ. Res. Lett. 14 1–13

Marselis S M et al 2020 Evaluating the potential of full-waveform lidar for mapping pan-tropical tree species richness Glob. Ecol. Biogeogr. 29 1799–816

Maskell L C et al 2019 Exploring relationships between land use intensity, habitat heterogeneity and biodiversity to identify and monitor Biodiversity of high nature value farmland Conserv. 231 30–38

May F, Gerstner K, McClinn D J, Xiao X and Chase J M 2018 mobism: an R package for the simulation and measurement of biodiversity across spatial scales methods Ecol. Evol. 9 1401–8

McNicol I M, Ryan C M, Dexter K G, Ball S M J and Williams M 2018 Aboveground carbon storage and its links to stand structure, tree diversity and floristic composition in south-eastern Tanzania Ecosystems 21 740–54

McRoberts R E, Chen Q, Gormanson D D and Walters B F 2018 The shelf-life of airborne laser scanning data for enhancing forest inventory inferences Remote Sens. Environ. 206 254–9

Meyer L, Diniz-Filho J S, Lohmann L G, Hortal J, Barreto E, Rangel T and Kissling W D 2020 Canopy height explains species richness in the largest clade of Neotropical lianas Glob. Ecol. Biogeogr. 29 26–37

Meyer V, Saatchi S, Ferraz A, Xu L, Duque A, Garcia M and Chavez J 2019 Forest degradation and biomass loss along the Choco region of Colombia Carbon Balance Manage. 14 1–15

Morandi P S et al 2018 Tree diversity and above-ground biomass in the South America Cerrado biome and their conservation implications Biodivers. Conserv. 29 1539–56

Myeni R and Knyazikhin Y 2015 MODIS1A2H MODIS/terra leaf area index/FPAR 8-day L4 global 500m SIN grid V006 [Data set] NASA EOSDIS L. Process. DAAC (https://doi.org/10.5067/MODIS/MODIS1A2H.006)

Nicodemus K K and Malley J D 2009 Predictor correlation impacts machine learning algorithms: implications for genomic studies Bioinformatics 25 1884–90

Nicotra A B, Chaudoin R L and Iriarte J 2018 Spatial heterogeneity of light and woody seedling regeneration in tropical wet forests Ecology 90 1908–26

Oliveira B F and Schepers B R 2019 Vertical stratification influences global patterns of biodiversity Ecography 42 249

Oliveira W L, Medeiros M B, Simon M F, Hay J and Ter Steege H 2018 The role of recruitment and dispersal limitation in tree community assembly in Amazonian forests Plant Ecol. Divers. 11 1–12

Olson D M et al 2001 Terrestrial ecoregions of the worlds: a new map of life on Earth Bioscience 51 933–8

Parmentier I et al 2011 Predicting alpha diversity of African rain forests: models based on climate and satellite-derived data do not perform better than a purely spatial model J. Biogeogr. 38 1164–76

Pereira H M et al 2013 Essential biodiversity variables Science 339 277–8

Phillips O and Miller J S 2002 Global Patterns of Plant Diversity: Alwyn H. Gentry’s Forest Transect Data Set (Monographs in Systematic Botany from the Missouri Botanical Garden: Missouri Botanical Press) p 319 (available at: https://books.google.com.co/books?hl=es&id=yKwsAQAAMAJ)

Potapov P et al 2019 Annual contiguous fields of woody vegetation structure in the Lower Mekong region from 2000–2017 Landsat time-series Remote Sens. Environ. 232 111278

Potapov P et al 2020 Mapping global forest canopy height through integration of GEDI and Landsat data Remote Sens. Environ. 112165 (available at: www.sciencedirect.com/science/article/pii/S0034425720305381)

Poveda G and Mesa O J 2000 On the existence of Lloro (the rainiest locality on earth): enhanced ocean-land–atmosphere interaction by a low-level jet Geophys. Res. Lett. 27 1675–8

Qi W and Dubayah R O 2016 Combining Tandem-X InSAR and simulated GEDI lidar observations for forest structure mapping Remote Sens. Environ. 187 253–66

Rangel O 2011 La Vegetación Del Chocó Biogeográfico De Colombia Y Zonas Cordilleranas Aledañas—áncoras Colombia Diversidad Biótica Iv ed O Range (Bogota, Universidad Nacional de Colombia) pp 769–815 (available at: www.colegiobioltda.biota.com/Sitoio_web/ LIRBOS_DEL_I_AL_IV/Entradare2011/3/28_COLOMBIA_DIVERSIDAD_BIOTICA_IV.html)

Rocchini D 2007 Effects of spatial and spectral resolution in estimating ecosystem α-diversity by satellite imagery Remote Sens. Environ. 111 423–34

Rocchini D et al 2010 Remotely sensed heterogeneity as a proxy of species diversity: recent advances and open challenges Ecol. Inform. 5 318–29

Roussel J-R, David A, De Boissieu F and Meador A S 2019 Package ‘lidR’: airborne LiDAR data manipulation and visualization for forestry applications (Québec: R Core Team) (available at: https://cran.r-project.org/web/packages/lidR/lidR.pdf)

Roy D P, Kovalsky V, Zhang H K, Vermote E F, Yan L, Kumar S S and Egorov A 2016 Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity Remote Sens. Environ. 185 57–70

Schäfer E, Heiskanen J, Heikinheimo V and Pelikka P 2016 Mapping tree species diversity of a tropical montane forest by unsupervised clustering of airborne imaging spectroscopy data Ecol. Indic. 64 49–58

Schneider F D, Morsdorf F, Schimid B, Petch J, O L, Hueni A, Schimel D S and Schaepman M E 2017 Mapping functional diversity from remotely sensed morphological and physiological forest traits Nat. Commun. 8 1441

Schneider F, Ferraz A, Hancock S, Duncanson I L, Dubayah R, Pavlick R and Schimel D 2020 Towards mapping the diversity of canopy structure from space with GEDI. Environ. Res. Lett (available at: http://iopscience.iop.org/10.1088/1748-9326/ab9e99)

Shannon C, Read T, Bucholz R and Iriarte J F 1984 A mathematical theory of communication Bell Syst. Tech. J. 27 737–423

Sheil D, Salim A, Chavez J R, Vanclay J and Hawthorne W D 2006 Exploring relationships between land use and forest functional form in eastern and western North America For. Ecol. Manage. 215 494–507

Shimada M, Itoh T, Motooka T, Watanabe M, Shiraishi T, Potapov P, Sheil D, Salim A, Chave J R, Vanclay J and Hawthorne W D 2006 Exploring the relation between remotely sensed vertical canopy structure and tree species diversity in Gabon Environ. Res. Lett. 14 1–13

Silva C A et al 2018 Comparison of small- and large-footprint lidar characterization of tropical forest aboveground structure and biomass: a case study from central Gabon IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 3512–26

Simonson W D, Allen H D and Coomes D A 2014 Applications of airborne lidar for the assessment of animal species diversity Methods Ecol. Evol. 5 719–29
Simpson E H 1949 Measurement of Diversity Nature 163 688
Stein A, Gerstner K and Kreft H 2014 Environmental
heterogeneity as a universal driver of species richness across
taxa, biomes and spatial scales Ecol. Lett. 17 866–80
Ter Steege H et al 2003 A spatial model of tree α-diversity
and tree density for the Amazon Biodivers. Conserv. 12 2259–77
Torresani M, Rocchini D, Sommenschin R, Zebisch M,
Hauffe H C, Heym M, Pretzsch H and Tonon G 2020 Height
variation hypothesis: a new approach for estimating forest
species diversity with CHM LiDAR data Ecol. Indic. 117 1–9
Toszogyova A and Storch D 2019 Global diversity patterns are
modulated by temporal fluctuations in primary productivity
Glob. Ecol. Biogeogr. 28 1827–38
Tovo A, Suweis S, Formentin M, Favretti M, Volkov J, Banavar J R,
Azaele S and Maritan A 2017 Upscaling species richness and
abundances in tropical forests Sci. Adv. 3 1–11
USGS U G S 2019 USGS landsat 8 surface reflectance tier 1 earth
data cat (available at: https://developers.google.com/
earth-engine/datasets/catalog/LANDSAT_LC08_
C01_T1_SR)
Vihervaara P et al 2017 How essential biodiversity variables and
remote sensing can help national biodiversity monitoring
Glob. Ecol. Conserv. 10 43–59
Vollrath A, Mullissa A and Reiche J 2020 Angular-based
radiometric slope correction for Sentinel-1 on Google Earth
engine Remote Sens. 12 1–14
Walter J A, Stovall A E L and Atkins J W 2021 Vegetation
structural complexity and biodiversity in the Great Smoky
Mountains Ecosphere 12 e03590
Wan Z, Hook S and Hulley G 2015 MOD11A1 MODIS/terra land
surface temperature/emissivity daily L3 global 1 km SIN
grid V006 NASA EOSDIS L Process. DAAC (https://doi.org/
10.5067/MODIS/MOD11A1.006)
Whittaker R H 1960 Vegetation of the Siskiyou mountains,
Oregon and California Ecol. Monogr. 30 279–338
Whittaker R H 1965 Dominance and diversity in land plant
communities Science 147 250–60
Wright D H, Currie D J and Maurer B A 1993 Energy supply and
patterns of species richness on local and regional scales
Species Diversity in Ecological Communities: Historical
And geographical Perspective ed R E Ricklefs and D Schluter
(Chicago, IL: University of Chicago Press) pp 66–74
WWF W W F 2019 Choco-darien moist forests (available at:
http://wwf.panda.org/about_our_earth/ecoregions/
chocodarien_moist_forests.cfm)
Zhang Z, Cao L and She G 2017 Estimating forest structural
parameters using canopy metrics derived from airborne
LiDAR data in subtropical forests Remote Sens. 9 1–16