ENVIRONMENTAL RESEARCH
LETTERS

LETTER • OPEN ACCESS

The use of GEDI canopy structure for explaining variation in tree species richness in natural forests

View the article online for updates and enhancements.

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

The use of GEDI canopy structure for explaining variation in tree species richness in natural forests

Suzanne M Marselis[1,*] ⬤, Petr Keil[2], Jonathan M Chase[3,4] and Ralph Dubayah[5]

[1] Institute of Environmental Sciences, Leiden University, Einsteinweg 2, 2333 CC Leiden, The Netherlands
[2] Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Kamýcká 129, Praha, Suchdol 16500, Czech Republic
[3] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstraße 4, 04103 Leipzig, Germany
[4] Institute of Computer Science, Martin Luther University Halle-Wittenberg, 06120 Halle (Saale), Germany
[5] Department of Geographical Sciences, University of Maryland, 7251 Preinkert Dr, College Park, MD 20742, United States of America
* Author to whom any correspondence should be addressed.

E-mail: s.m.marselis@cml.leidenuniv.nl

Abstract

Variables describing the abiotic environment (e.g. climate, topography or biogeographic history) have a long tradition of use as predictors of tree species richness patterns. However, these variables may capture variations in richness related to climate, but not those that are related to soil type or forest disturbance. Canopy structure has previously been shown to provide information on the variation of tree species richness, with richness generally increasing with larger canopy heights and denser foliage. The use of canopy structure is increasingly relevant with the availability of such data from the Global Ecosystem Dynamics Investigation (GEDI), a lidar mission onboard the International Space Station. In this analysis we show that GEDI canopy structure explains up to 66% of the variation in tree species richness in natural forests without a history of recent disturbance across the globe. However, this portion overlaps with the variation (up to 80%) explained by environmental and biogeographical variables. Our results show that relationships between tree species richness on one side and climate and canopy structure on the other side are not as straightforward as we initially expected, and should be further investigated across both natural and disturbed forests.

1. Introduction

Tree species diversity varies greatly across the Earth, ranging from about one species per hectare in boreal forests at high latitudes, to several hundred species per hectare in tropical forests (Keil and Chase 2019). Understanding and predicting this variation is important because tree diversity has been linked to ecosystem services, such as productivity and carbon accumulation (Huang *et al* 2018, Craven *et al* 2020), and accurate predictions of diversity can improve future conservation decisions (Algar *et al* 2009).

Because of logistical and taxonomic challenges, it has so far been impossible to measure tree diversity directly in the field over large, continuous areas. Instead, we only have information from discrete, isolated forest plots (Šímová *et al* 2011, Ricklefs and He 2016, Sullivan *et al* 2017, Keil and Chase 2019). A common practice is to link diversity in

these plots to environmental predictors (e.g. climate, topography, land cover, and other spatial predictors) using parametric statistical models or machine learning algorithms (e.g. Ricklefs and He 2016, Keil and Chase 2019, Večeřa *et al* 2019). These can then be used to predict diversity to unsurveyed areas, and to test hypotheses about the role of various predictors driving variation in diversity. Thus far, climatic variables describing the water-energy balance (such as evapotranspiration, precipitation, temperature, remotely sensed proxies of productivity, and their seasonality) have been particularly effective predictors of local plant and tree diversity over large geographic extent, explaining considerable fractions of variation (see table 1 for examples). However, there is still a margin of improvement, for example, climatic, topographic, and land-use variables only explain a small portion of the variation in richness in some regions (e.g. continental US; Craven *et al* 2020).

Table 1. Several examples of studies modelling the relationship between tree diversity in local samples (plots) and environmental predictors, and the percentage of variation in species richness that the models were able to explain.

Reference	Geographic extent	Spatial grain	Predictors	Modelling method	Explained variation
Craven *et al* (2020)	USA	0.0672 ha, FIA plots	Productivity, topography, climate, land use, area	Random forest	37%
Keil and Chase (2019)	Global	Varying	Climate, topography, space, productivity, area	GAM	88%
Ricklefs and He (2016)	Global	Varying, CTFS plots	Climate, space, # of trees	GLM	92%
Šímová *et al* (2011)	Global	0.1 ha, Gentry plots	Climate, space	GLM	76%
Francis and Currie (1998)	Northern temperate zone	Varying	Climate, area, space	LM	76.5%
Kay *et al* (1997)	South America	Local plots (unclear)	Climate (rainfall)	Local regression	65%
Ganzhorn *et al* (1997)	Madagascar	0.1 ha	Climate (rainfall)	Correlation	50%
Specht and Specht (1994)	Australia	1 ha	Productivity	LM	98%
Clinebell *et al* (1995)	Northern Neotropics	0.1 ha, Gentry plots	Climate, soils	LM	53%

One way to potentially improve the existing models of tree diversity is through the use of forest structure. Canopy structure depends on environmental factors, soil, disturbance and crown plasticity (Frolking *et al* 2009, Jucker *et al* 2015, Pfeifer *et al* 2018, Senf *et al* 2020, Hakkenberg and Goetz 2021) and thus may contain information on tree species diversity that cannot be captured by environmental variables alone. There are theoretical reasons that may explain why forest structure (such as height, total volume, or complexity of canopy) should predict richness. First, due to sampling effects (Šímová *et al* 2011), a more diverse forest could have sampled species (from regional species pool) that fill the upper canopy, as well as shade-tolerant species in the forest understory. Second, canopy height can be considered a proxy for canopy volume, which may correlate with the volume of niche-based opportunities for species to coexist. Both effects can increase diversity in more complex, higher, and voluminous canopy structures, through niche complementarity and resource partitioning (Schoener 1974, Chase and Leibold 2003) and there is empirical evidence that forests with higher canopies have a higher plant species diversity (Gatti *et al* 2017). Furthermore, studies using vertical canopy structure data derived from full-waveform lidar explained more of the variation in tropical tree species richness than canopy height alone (Marselis *et al* 2019, 2020). Such models become increasingly relevant with the availability of more detailed information on not only canopy height, but also the density of canopy material along the vertical forest axis. The Global Ecosystem Dynamics Investigation (GEDI) mission (Dubayah *et al* 2020), launched in December 2018,

is now providing a large set (billions) of such canopy structure measurements. However, the potential of canopy structure, such as observed by GEDI, to predict broad-scale patterns of tree biodiversity remains unknown.

Here we evaluate the efficacy of the GEDI-derived canopy structure to explain variation in tree species richness in natural and semi-natural forest plots at a near-global extent covering both the tropical and temperate regions. We also compare the efficacy of canopy structure variables within predictive models of tree species richness with other previously used predictors, namely those related to climate, topography, and biogeographic history, as used in Keil and Chase (2019).

2. Methods

2.1. Data

We used tree species richness data which were compiled from publicly available, previously published studies, by Keil and Chase (2019). Specifically, the dataset consisted of data on species richness from 1166 forest plots in tropical and temperate regions from across the world. These forest plots were originally selected by Keil and Chase (2019) to represent both temperate and tropical natural and semi-natural forests across as many biomes and biogeographic realms as possible given open access data availability. Nevertheless, as with most such analyses, there are considerable data gaps in large parts of Africa and Asia due to limited availability of forest plots from these regions. Only plots from natural or semi-natural forest (e.g. not from plantations, production

forests, or severely disturbed/damaged forests, for example, as a result of logging, hurricanes, or bark beetle outbreaks) were used. These plots were taken from published database compilations (Phillips and Miller 2002, Phillips *et al* 2003, Ramesh *et al* 2010, Anderson-Teixeira *et al* 2015, Coelho de Souza *et al* 2016, Ricklefs and He 2016, Sullivan *et al* 2017), national forest inventory surveys (De Natale *et al* 2005, USDA 2016, IGN 2017), and extracted manually from primary sources (Beals 1965, Yamada 1976, Abbott 1984, Kohyama 1986, Cao and Zhang 1997, Linder *et al* 1997, Ansley and Battles 1998, Narayanan and Parthasarathy 1999, Maycock *et al* 2000, Adam 2001, Kohira *et al* 2001, Cairns *et al* 2003, Enoki 2003, Hirayama and Sakimoto 2003, Cheng-Yang *et al* 2004, Fashing *et al* 2004, Jing-Yun *et al* 2004, Sheil and Salim 2004, Shu-Qing *et al* 2004, Wu *et al* 2004, Bonino and Araujo 2005, Davis *et al* 2005, Sawada *et al* 2005, Splechtna *et al* 2005, Adekunle 2006, Graham 2006, Malizia and Grau 2006, Szwagrzyk and Gazda 2007, Lopes *et al* 2008, Yasuoka 2008, Addo-Fordjour *et al* 2009, Eichhorn 2010, Krishnamurthy *et al* 2010, Lalfakawma *et al* 2010, Nagel *et al* 2010, Namikawa *et al* 2010, Lü *et al* 2010, Eshete *et al* 2011, Round *et al* 2011, Sanchez *et al* 2013, Popradit *et al* 2015, van Do *et al* 2015, Wusheng *et al* 2015). For each plot we used the number of tree species (S) as our focal response variable. From the abovementioned studies we further extracted three covariates related to sampling design, which can all influence the observed S, and thus need to be controlled for in statistical analyses. These covariates were:

- The total area of the sampled plot (km^2).
- The minimum diameter at breast height (minDBH in cm); only trees of equal or larger diameter than minDBH were measured and included in the analysis.
- The number of trees larger than the minDBH in the sampled plot.

For each plot, we considered six variables related to the climate, topography, and biogeographic history which were previously successfully used to explain tree species richness in Keil and Chase (2019). These were:

- WorldClim (Hijmans *et al* 2005) mean annual temperature (BIO1), mean isothermality (BIO3), and mean precipitation seasonality (BIO15), which are 1970–2000 spatially interpolated averages extracted from rasters with a resolution of 30 arcseconds (1 km^2 around equator) available at www.worldclim.org/data/v1.4/worldclim14.html.
- Absolute difference between the lowest and highest elevation in a 30 arc-seconds raster cell within which a given plot is located, derived from Danielson and Gesch (2011).

- Island vs. mainland, a binary factor indicating if a plot is on an island or mainland (from Weigelt *et al* 2013), following a loose classification in which shelf islands (connected to mainland in the past) are classified as islands.
- Classification of biogeographic realms into Afrotropic, Australasian, Easter Palearctic, Indo-Malay, Nearctic, Neotropic, and Western Palearctic, following (Ricklefs and He 2016).
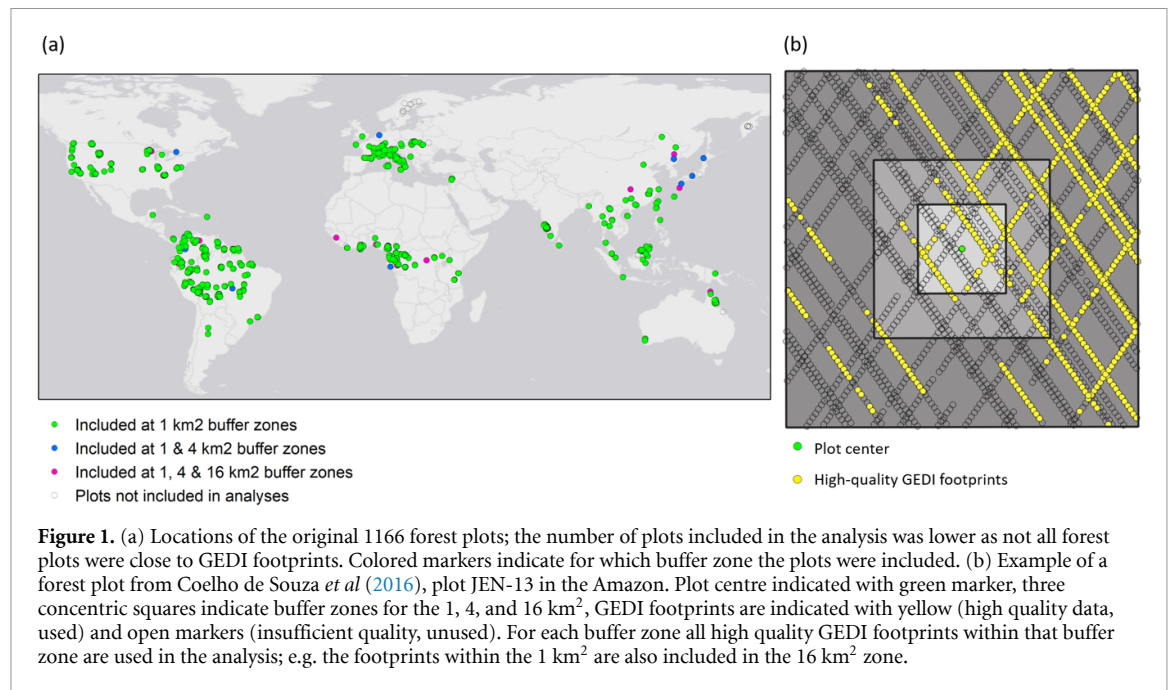
For further details on these environmental data, see supplementary material of Keil and Chase (2019).

We used GEDI-derived variables describing canopy structure from the first two years of data collection: April 18, 2019–May 14, 2021. GEDI is a sampling instrument mounted to the International Space Station, collecting lidar data between 51.6° N and S. GEDI has three lasers that are split into four beams total. The beams are dithered across track, resulting in eight beams along-track. The footprint-spacing along track is 60 m and the across-track spacing is 600 m. At each sampling location, GEDI collects a full-waveform with a nominal footprint size of ∼25 m (Dubayah *et al* 2020). During the first two years of the mission, GEDI was programmed to collect data with the least spatial overlap, acquiring the densest possible sampling at the footprint level. Because of the sampling method, footprints do not necessarily overlap exactly with the sampled forest plots from Keil and Chase (2019), thus we used the information from GEDI footprints in square buffer zones around the original plot locations. This also ensures that we capture the canopy structure of the forest plots for those plots with a limited spatial location accuracy. GEDI footprints themselves have a geolocation accuracy of about 10 m. We included plots that have at least one GEDI footprint within the buffer zone around the plot centre. We evaluated three buffer zones (figure 1(b)):

- 1 × 1 km buffer zone (1 km^2, 888 plots).
- 2 × 2 km buffer zone (4 km^2, 978 plots).
- 4 × 4 km buffer zone (16 km^2, 1053 plots).

There were some plots where no GEDI shots intersected any of the buffers because of its sampling method. Moreover, the number of plots available at the different buffer zones increases with the size of the buffer zone (varying from 888 to 1053 plots out of the original 1166 plots considered, figure 1(a)).

According to Marselis *et al* (2019, 2020), the most important GEDI metrics related to tree species richness were: canopy height, total plant area index and the plant area index along the vertical forest axis divided into 10 m vertical bins. We adopted this same set of metrics. However, to confirm that other canopy structure variables do not explain more of the variation in richness, we included the

Figure 1. (a) Locations of the original 1166 forest plots; the number of plots included in the analysis was lower as not all forest plots were close to GEDI footprints. Colored markers indicate for which buffer zone the plots were included. (b) Example of a forest plot from Coelho de Souza *et al* (2016), plot JEN-13 in the Amazon. Plot centre indicated with green marker, three concentric squares indicate buffer zones for the 1, 4, and 16 km², GEDI footprints are indicated with yellow (high quality data, used) and open markers (insufficient quality, unused). For each buffer zone all high quality GEDI footprints within that buffer zone are used in the analysis; e.g. the footprints within the 1 km² are also included in the 16 km² zone.

results of a principal component analysis (PCA) analysis performed on a large set of canopy structure metrics (180 metrics, appendix A). The final set of canopy metrics used here consists of the following variables:

- Canopy Height, measured as the relative height at which 98% of the energy was returned (RH98).
- Total plant area index (PAI) along the vertical forest axis
- The total PAI between 0 and 10 m vertically, m²/10 m³
- The total PAI between 10 and 20 m vertically, units in m²/10 m³
- The total PAI between 20 and 30 m vertically, units in m²/10 m³
- The total PAI between 30 and 40 m vertically, units in m²/10 m³
- The total PAI between 40 and 50 m vertically, units in m²/10 m³
- The total PAI above 50 m vertically, units in m²/10 m³
- The ground elevation

All these metrics were derived from the GEDI Level 2A and 2B datasets version 2 (Dubayah *et al* 2021a, 2021b). The L2A GEDI dataset contains the relative height (RH) products of the processed, geolocated waveform, including RH98 (essentially canopy height) and the ground elevation. The L2B dataset contains the vertical profile metrics; including the total PAI and the PAI along the vertical axis. GEDI L2A data contains information from four processing algorithms applied to the geolocated waveforms. We used the canopy height and ground elevation from the processing algorithm flagged as the 'best' algorithm

for each shot in the GEDI dataset (Dubayah *et al* 2021a, 2021b). We also only included waveforms that were collected in the leaf on season, using the '*leaf off flag*' in the GEDI data products. Lastly, only waveforms that were flagged '1' for the metric '*successful processing at the 2b data level*' (vertical canopy structure, Dubayah *et al* 2021b) were considered reliable and used in this analysis.

For each of the 16 km², 4 km² and 1 km² areas, we calculated the average of the nine canopy structure metrics from all footprints within the buffer zone and joined them to the field and environmental data, creating three final datasets (one for each buffer zone).

2.2. Analysis

2.2.1. Raw relationships

In a first exploration, we examined the general relation between the observed tree species richness and canopy structure using just two selected canopy structure metrics, canopy height and total PAI, as these are descriptors of the general canopy structure and have previously shown significant relationships with tree species richness in the tropics (Marselis *et al* 2020). This preliminary analysis gives insight in variation in the used data, and it shows naïve relationships when failing to control for the effects of other variables.

First, to show the relationships in the raw data, we fitted single-term generalized linear models, with S as a response (quasi-Poisson errors, log link function) and canopy height or total PAI as predictors. We performed a second exploratory step as S can be affected by sampling protocols that vary among plots. Thus, we examined these relationships after controlling for the effects of covariates related to sampling protocol. To do this, we first fitted a 'covariate' random forest (RF) model (see below in the variation partitioning

section), with plot area, number of tree individuals, and minDBH as predictors of S. We calculated the residuals as (observed $\log_{10}$ S) − (predicted $\log_{10}$ S), and fitted a generalized linear model (normal errors, identity link function) to these residuals as a function of canopy height and total PAI. We chose this model because, unlike the raw counts of species (S), residuals centre around 0 and thus cannot be modelled by the distributions designed to model count data (e.g. Poisson, or quasi-Poisson).

2.2.2. RFs
We used RFs (Breiman 2001) as the main analytical tool to evaluate the relationships between the main response variable tree species richness (S) and all the other predictors. Specifically, we used RF setting with 1000 trees and the default settings of the 'randomForest' function implemented in R package (randomForest v 4.6–14), to model S as a function of study design covariates, and environmental, biogeographical, and GEDI-derived predictors (details are below).

2.2.3. Collinearity and variation partitioning
Our data are observational and can potentially suffer from collinearity problems. Specifically, forests in tropical warm and humid areas may have similar canopy structures, which are different from forests in other biomes (e.g. cold and/or dry climates). This means that species diversity can potentially be explained either by environmental predictors (precipitation, temperature, primary productivity, topography, biogeographical history), or by GEDI canopy structure, and we may potentially have no way to tease these two apart. We show the relation among predictors in appendix B; negative relations are found between mean elevation with annual temperature and isothermality, and positive correlations are mostly found between canopy height and the total PAI values. Another correlation can be found between the PAI of one layer with its subsequent layer (e.g. PAI between 0–10 m and PAI between 10–20 m). To account for the collinearity, we performed variation partitioning (Borcard *et al* 1992, Keil and Chase 2019, Viana *et al* 2021). Variation partitioning separates the independent fractions of variation in response (in our case species richness, S) explained by groups of predictors. We looked for fractions attributable exclusively to environmental predictors ($R_{\text{envi\_only}}$) or to GEDI-derived predictors($R_{\text{GEDI\_only}}$). We also assessed the shared fraction (R_{shared}) that cannot unequivocally be attributed to one of these two groups of predictors, and emerges as a result of the collinearity. Importantly, if R_{shared} is larger than one or both of the independent fractions $R_{\text{envi\_only}}$ and $R_{\text{GEDI\_only}}$, then there is too much collinearity in the data to say which group of predictors is more important than the other.

To do the variation partitioning, we fitted four separate random forests:

(a) RF_{cov} with covariates describing size of the plot, number of trees and minDBH, but no environmental or GEDI predictors.
(b) $RF_{\text{envi\_only}}$ with covariates and environmental predictors.
(c) $RF_{\text{GEDI\_only}}$ with covariates and GEDI predictors.
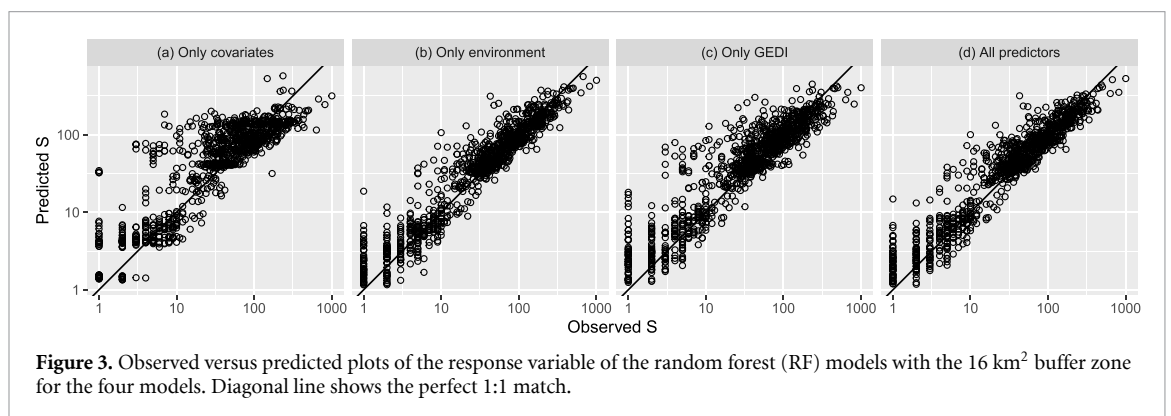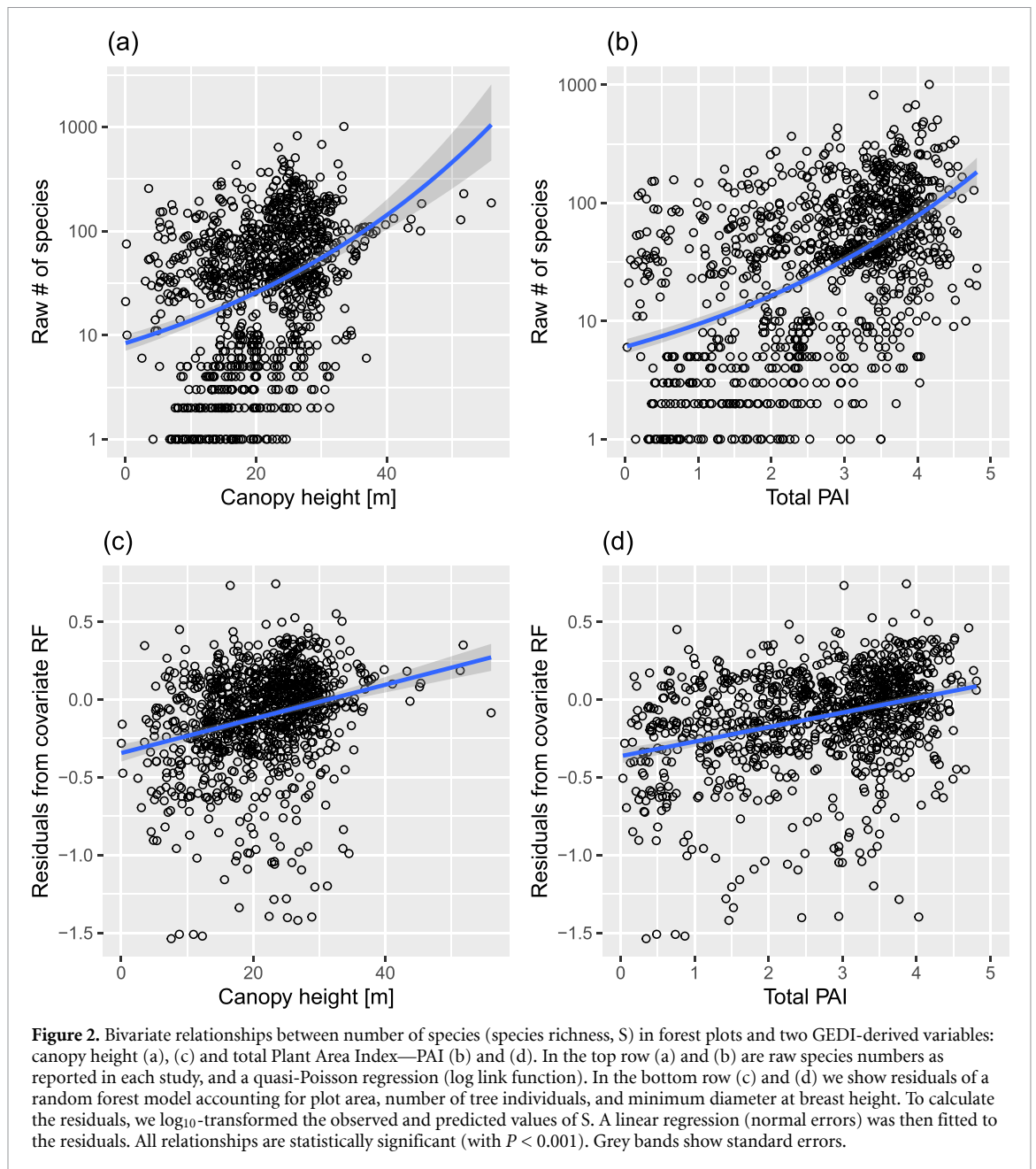(d) RF_{full} with covariates, environmental predictors, and GEDI predictors.

We measured the amounts of variation in S explained by these random forests using R_{cov}, $R_{\text{envi\_only}}$, $R_{\text{GEDI\_only}}$ and R_{full} respectively, where each of these is a coefficient of determination (R^2) from the respective RF model. We thus get $R_{\text{envi\_only}} = R_{\text{full}} - R_{\text{GEDI}}$, $R_{\text{GEDI\_only}} = R_{\text{full}} - R_{\text{envi}}$, and $R_{\text{shared}} = R_{\text{full}} - (R_{\text{GEDI\_only}} + R_{\text{envi\_only}})$. We did the RFs for each of the three buffer zones, resulting in 12 global RFs. The data and code are at: https://figshare.com/s/82d033d5592889af48bc (Marselis and Keil 2022).

3. Results

Examination of the raw relationships (using a 16 km^2 buffer zone) revealed a significant positive relationship between the number of tree species with both canopy height and total PAI (figure 2). This relationship was significant in the raw data (figures 2(a) and (b)), and also after accounting for the effects of study design covariates: plot area, total number of individuals, and minimum DBH (figures 2(c) and (d)). However, the relationships were weaker in the residuals than in the raw values, and there was a substantial fraction of variation around the fitted lines (figure 2).
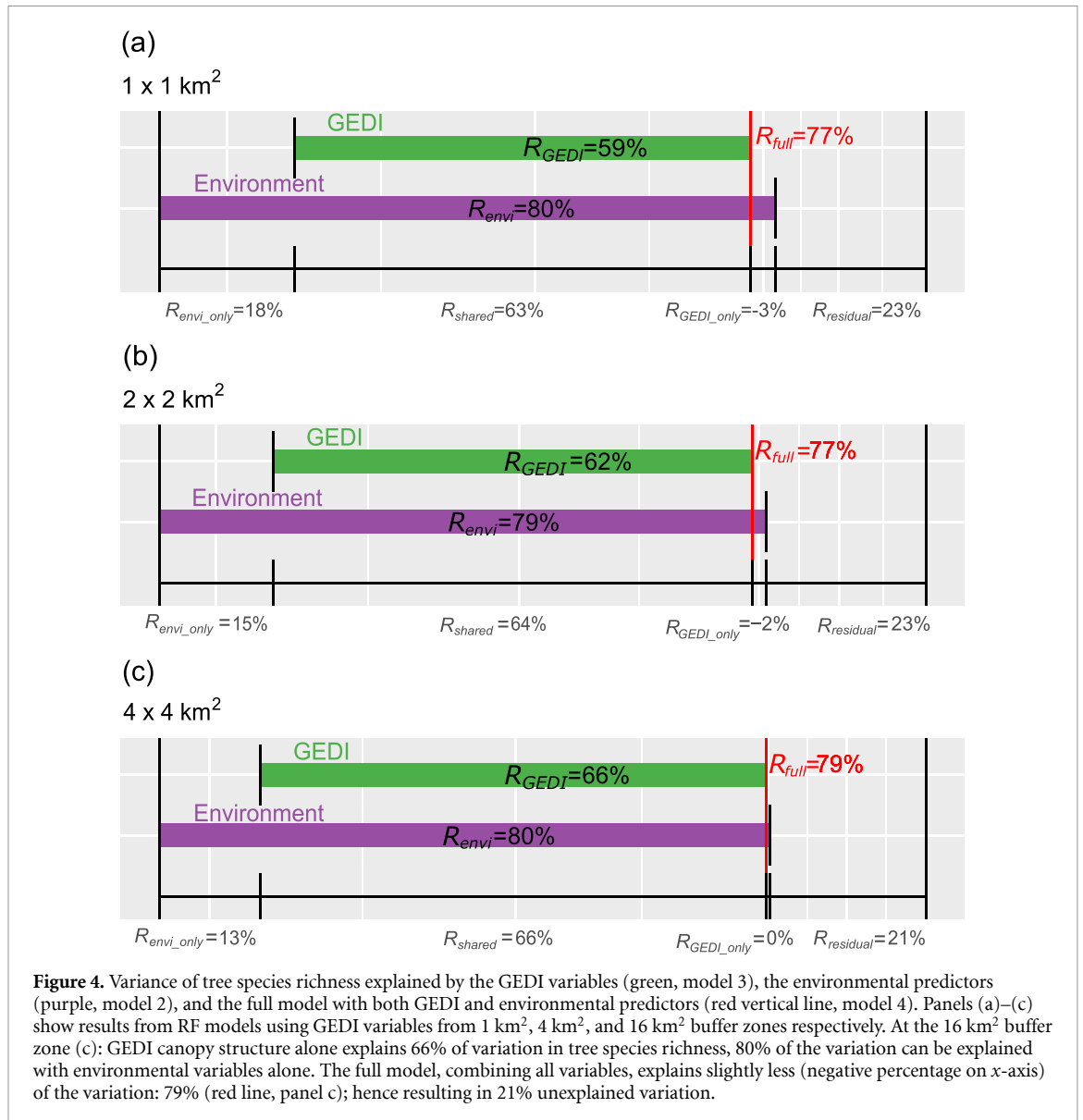
The observed vs. predicted plots for the four RF models show that both environmental and GEDI predictors explain substantial fractions of variation in tree species richness at the 16 km^2 buffer zone (figure 3), although the environmental predictors alone explain more than the GEDI predictors alone (figure 4). Specifically, 80% of the variation in tree species richness was explained by the covariates and environmental variables (R_{envi}, figure 4(a)). GEDI variables, in combination with the covariates, explained 59% of this variation (R_{GEDI}, figure 4(a)). A random forest model that used both GEDI canopy structure and environmental variables explained just 77% of the variation (R_{full}, figure 4(d)). Importantly, the fraction of variation explained by GEDI overlaps with the fraction explained by the environmental variables (figure 4(a)). For partial dependence plots of the RF models, see appendix D.

The results for the models developed for the 4 and 16 km^2 buffer zones were similar to those at 1 km^2 zone. GEDI variables alone explained respectively

Figure 2. Bivariate relationships between number of species (species richness, S) in forest plots and two GEDI-derived variables: canopy height (a), (c) and total Plant Area Index—PAI (b) and (d). In the top row (a) and (b) are raw species numbers as reported in each study, and a quasi-Poisson regression (log link function). In the bottom row (c) and (d) we show residuals of a random forest model accounting for plot area, number of tree individuals, and minimum diameter at breast height. To calculate the residuals, we $\log_{10}$-transformed the observed and predicted values of S. A linear regression (normal errors) was then fitted to the residuals. All relationships are statistically significant (with $P < 0.001$). Grey bands show standard errors.

Figure 3. Observed versus predicted plots of the response variable of the random forest (RF) models with the 16 km² buffer zone for the four models. Diagonal line shows the perfect 1:1 match.

62% and 66% of the variation, the environmental variables 79% and 80% and the full models 77% and 79% (figures 4(b) and (c)). The fraction of variation explained by GEDI overlapped with the fraction explained by the environmental variables for both buffer zones.
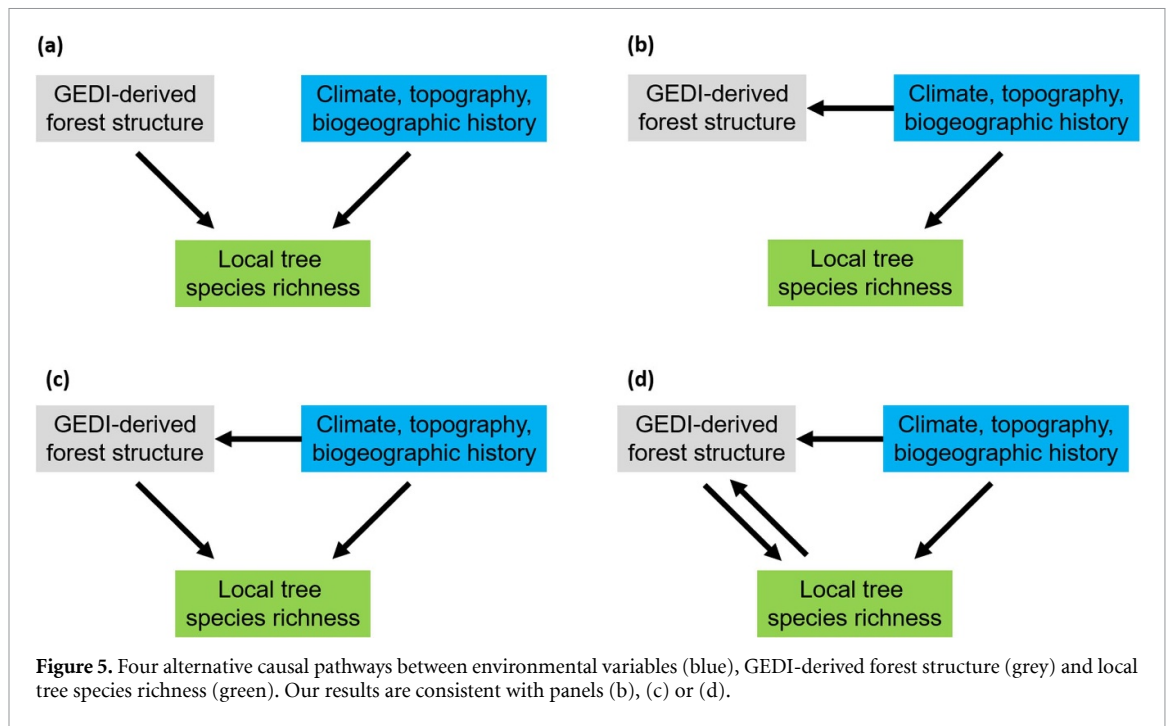
Figure 4. Variance of tree species richness explained by the GEDI variables (green, model 3), the environmental predictors (purple, model 2), and the full model with both GEDI and environmental predictors (red vertical line, model 4). Panels (a)–(c) show results from RF models using GEDI variables from 1 km^2, 4 km^2, and 16 km^2 buffer zones respectively. At the 16 km^2 buffer zone (c): GEDI canopy structure alone explains 66% of variation in tree species richness, 80% of the variation can be explained with environmental variables alone. The full model, combining all variables, explains slightly less (negative percentage on *x*-axis) of the variation: 79% (red line, panel c); hence resulting in 21% unexplained variation.

4. Discussion

4.1. GEDI vs. environmental variables

The most important result is that, in local forest plots at a global extent, we were unable to unequivocally attribute variation in species diversity to GEDI-derived forest structure, nor to other environmental predictors. More specifically, even though forest structure (together with covariates describing study design) explained up to 66% of variation in tree species diversity, environmental predictors related to climate, topography, and history can explain even more variation in diversity (up to 80%), and there is no increase in explained variation when GEDI and environmental predictors are used together in one model. This means that GEDI-derived forest structure can indeed be used to explain local tree species diversity, which is in line with the relationship between canopy height and tree species diversity found by Gatti *et al* (2017), and it underpins the

hypothesis that canopy structure data may serve to predict variation in tree species richness (Marselis *et al* 2020). However, if other environmental predictors are available they may provide even better diversity estimates than the GEDI-derived variables, and we found no clear advantage in combining the strengths of the two groups of predictors for the datasets we examined.

Even though this is a negative result from a practical point of view–the GEDI-derived forest structure does not improve existing correlative near-global models of tree species diversity–there is a valuable conceptual inference. I*ndependent* effects of forest structure and environment on forest diversity (figure 5(a)) are unlikely, because otherwise, including forest structure should have unequivocally improved the overall models. Our results are more consistent with an alternative (causal) pathway in which the abiotic environment drives both forest structure and species diversity (figure 5(b)); with

Figure 5. Four alternative causal pathways between environmental variables (blue), GEDI-derived forest structure (grey) and local tree species richness (green). Our results are consistent with panels (b), (c) or (d).

the possibility of a secondary effect of forest structure on species diversity (figure 5(c)) or the possibility of a two-way effect where tree species richness also drives forest structure itself (figure 5(d)). However, our correlative models cannot reject or confirm this secondary link, since the overlapping fraction of explained variation (figure 4) is too large. Thus, we conclude that the three pathways (figures 5(b)–(d)) are all plausible. This is in line with other studies (Frolking *et al* 2009, Pfeifer *et al* 2018, Senf *et al* 2020, Hakkenberg and Goetz 2021) which hypothesize that canopy structure should depend on environmental factors, soil and disturbance. Potential evidence for the secondary effect (figure 5(c)) was found by Hakkenberg and Goetz (2021) who studied the relation between canopy structure, climate and tree species richness in forests of the United States. Their study showed that a small portion of the remaining variance was explained by *climate x structure* interaction terms. Such a result could not be reproduced at this global scale using GEDI canopy structure, potentially due to the differences in canopy structure metrics available as Hakkenberg and Goetz (2021) used high resolution discrete return airborne lidar data at 20×20 m plot level, which points to the importance of spatial scale, discussed below.

4.2. Limitations

Our tree species richness dataset only included data on natural or semi-natural forests; there were no (severely) disturbed forests, orchards, production forests, or plantations/monocultures. Thus, our forest plots mostly have a compact and well-developed

canopy, which potentially reduces the variation of forest structure across the dataset, limiting the scope for the GEDI-derived structure to explain diversity. In contrast, human-affected forests have distinctly 'unnatural', regular, or loose canopies (Tropek *et al* 2014). If such forests were also included in the analysis, we might expect that the GEDI-derived forest structure should explain more variation in richness, and could even outperform the other environmental predictors. Here we see a potential for joint use of environment and GEDI-derived variables to predictions of *actual* richness of all forests, including those modified by humans, as opposed to *potential* richness of natural plots as in this study and in Keil and Chase (2019).

Furthermore, GEDI footprints are not randomly distributed in our plots' buffer zones and they may not exactly overlap with the plots. For example, GEDI measurements may only cover a part of the buffer zone (e.g. figure 1(b)). More local, plot-specific forest structure information could provide additional/different information on the canopy structure than our buffered and spatially averaged metrics, but unfortunately the GEDI sampling density is not high enough to assert this. Encouragingly, however, we did not find considerable differences in model performance between the smallest buffer zone areas (1 km² and 4 km²) and a small improvement with the largest buffer zone area (16 km²). This is likely because structural variables derived from these different buffer areas are similar (appendix C), which indicates that averaging a handful of GEDI footprints in a 1 km² buffer zone is already representative of

general forest structure in a larger area. This could be in line with previous findings suggesting that the canopy structure in tropical forests becomes uncorrelated over just hundreds of meters (Réjou-Méchain *et al* 2014), suggesting that all of our buffer zones describe a *general forest structure* instead of the *plot-level forest structure*. In this case, the larger number of canopy structure samples in 16 km² zones, compared to the smaller zones, would likely result in a more accurate measure of the general forest structure which may explain the improved model performance at this larger buffer zone.

The temporal mismatches between collection of the field data, the environmental data, and the canopy structure data should also be kept in mind. Forest growth and (small) natural disturbances alter the forest over the shorter and longer term (Senf *et al* 2020), resulting in a mismatch of the forest structure and representation of the tree species diversity. This is a limitation that is difficult to overcome because of the time-consuming and expensive field data collection. Here we assume that the data used here (with the GEDI data from 2019 to 2021, climatic data from 1970 to 2000, and forest plot data collected between 1962 and 2016) represent the long-term situation in these natural and semi-natural forests. We assume that changes of climate and species temporal turnover in these forests are slow enough to still allow a correlative analysis involving the more recent GEDI data. The high explanatory power of both the climate and GEDI-derived variables support this assumption, although the real consequences of this temporal mismatch are unknown and should be investigated in the future. Additionally, it is yet unknown how the temporal dynamics of GEDI observations acquired from different times within the same leaf-on season, as well as across leaf-on seasons, may affect our results.

4.3. Future research
Future research should further explore the more realistic hypotheses (i.e. figures 5(b)–(d)). We need to step back to properly examine the relationship between abiotic conditions and forest structure (such as recently performed in the US by Hakkenberg and Goetz (2021)). Now that the GEDI data are available, a global analysis of this relationship using its massive amount of remotely sensed structure data is possible. This should enable us to further untangle the (causal) relationships between climate, forest structure, and disturbance, improve our understanding of ecosystem functioning, and enable the analysis of the real usefulness of canopy structure for mapping tree species richness in both natural and disturbed forests.

This also points to the importance of spatial scale in general. The question of the role of structure vs. environment in driving patterns of tree species richness should be explored as a function of scale. This could, for example, help to determine whether there are characteristic scales of analysis where structure

dominates beyond environmental variables or vice versa. Such an analysis would require high quality species richness data (e.g. at 25 m²), across a landscape, along with vertical canopy structure data at the same resolution in spatially continuous grids. Answering this question is exceptionally difficult because these kinds of continuous field data are rare. Furthermore, it has been shown that grain of disturbances is spatially very small, with mean disturbance sizes of 10–20 ha across continents, but with a median strongly skewed towards smaller patches (∼0.1 ha) (Taubert *et al* 2018), and not inconsistent with the decorrelation of structure over hundreds of meters noted above. Given the importance of small-scale disturbance to tree diversity, this suggests the need for fine-grained information on canopy structure at global scales. GEDI provides the basis for obtaining such observations when used in combination with other continuous remote sensing data at fine (30–100 m) resolutions, such as may be obtained from passive optical sensors (e.g. Landsat; Potapov *et al* 2021) or from radar sensors that measure 3D structure (e.g. TanDEM-X; Qi *et al* 2019). GEDI data can provide the basis for machine learning and other fusion methods that predict not only height and biomass at fine scales, but potentially the imputation of entire waveforms, opening the door to pervasive canopy structure predictions, at fine scales, across the landscape.

Some of the complex interactions can perhaps also be teased apart by focusing on fully mapped forest plots such as those in the CTFS/ForestGeo network (Anderson-Teixeira *et al* 2015), which offer rich data on exact location and growth form of every individual tree, collected over multiple time periods, with growth form and species of each known individual. If these were connected with remotely sensed forest structure using fusion with GEDI, an analysis teasing apart the interplay between forest structure, diversity, and abiotic conditions across a variety of scales may be possible.

Lastly, alternative analytical approaches that accommodate complex causal pathways, for example, structural equation modelling (Iriondo *et al* 2003, Lam and Maguire 2012), can be adopted. A similar analysis has been done to evaluate the direct and indirect effects of forest taxonomical and structural diversity, soil fertility and climate on aboveground carbon storage in neotropical forests (Poorter *et al* 2015).

5. Conclusion

Overall, our analysis showed a significant relationship between canopy structure and tree species diversity for natural and semi-natural forest undisturbed forests. However, we found that GEDI canopy structure cannot be unequivocally linked to the variation in tree species richness in natural undisturbed

forests because the portion of the variance explained by GEDI-derived canopy structure overlapped with the variance explained by environmental and biogeographical variables. Hence, to estimate tree species richness across broad spatial scales, GEDI-derived data, while useful, do not appear to provide any additional information for natural forests. Nevertheless, GEDI predictors may still be useful for explaining tree species richness in more disturbed forests that were not part of the current analyses, because environmental variables are less likely to contain information on changes in tree species richness that are linked to disturbance, and across fine spatial scales. Future research should include information on such disturbance gradients and further investigate the relationship between climate, forest canopy structure and tree species richness.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://figshare.com/s/82d033d5592889af48bc.

Table A1. R^2 of random forests built on GEDI metrics (GEDI), 5 (PCA 5), 10 (PCA 10) and 15 (PCA 15) principal components calculated from 159 GEDI metrics.

Buffer zone	GEDI	PCA 5	PCA 10	PCA 15
1 km^2	0.60	0.51	0.52	0.53
4 km^2	0.62	0.53	0.54	0.55
16 km^2	0.66	0.53	0.54	0.55

Appendix A. Principal Component Analysis on GEDI data

A PCA was performed to ensure that the chosen GEDI metrics (Marselis *et al* 2019, 2020) contained all information on canopy structure necessary to generate the structure-richness models. For this analysis the following set of GEDI metrics was extracted from all GEDI footprints and calculated at the 1 km^2, 4 km^2, and 16 km^2, buffer zones:

- Mean elevation lowest mode (ground elevation)
- Foliage height diversity (FHD)
- Canopy cover
- Total plant area index
- Relative height at 0, 25, 50, 75 and 98% (RH0, RH25, RH50, RH75, RH98)
- PAI at 5 m vertical intervals
- Cumulative PAI at 5 m vertical intervals

For each of the buffer zones we calculated the mean, standard deviation, maximum and minimum of all footprints falling within the buffer zone. This resulted in a total of 159 metrics. The results show (table A1) that the model performance is worse at all buffer zones than when using the carefully selected canopy structure metrics based on their biophysical meaning (Marselis *et al* 2019, 2020).

Appendix B. Collinearity between GEDI and environmental variables

Plant Area Indices at the various height layers tend to strongly correlate to the previous height layer as usually the amount of plant material higher in the canopy depends on the amount of plant material lower in the canopy (figure B1).
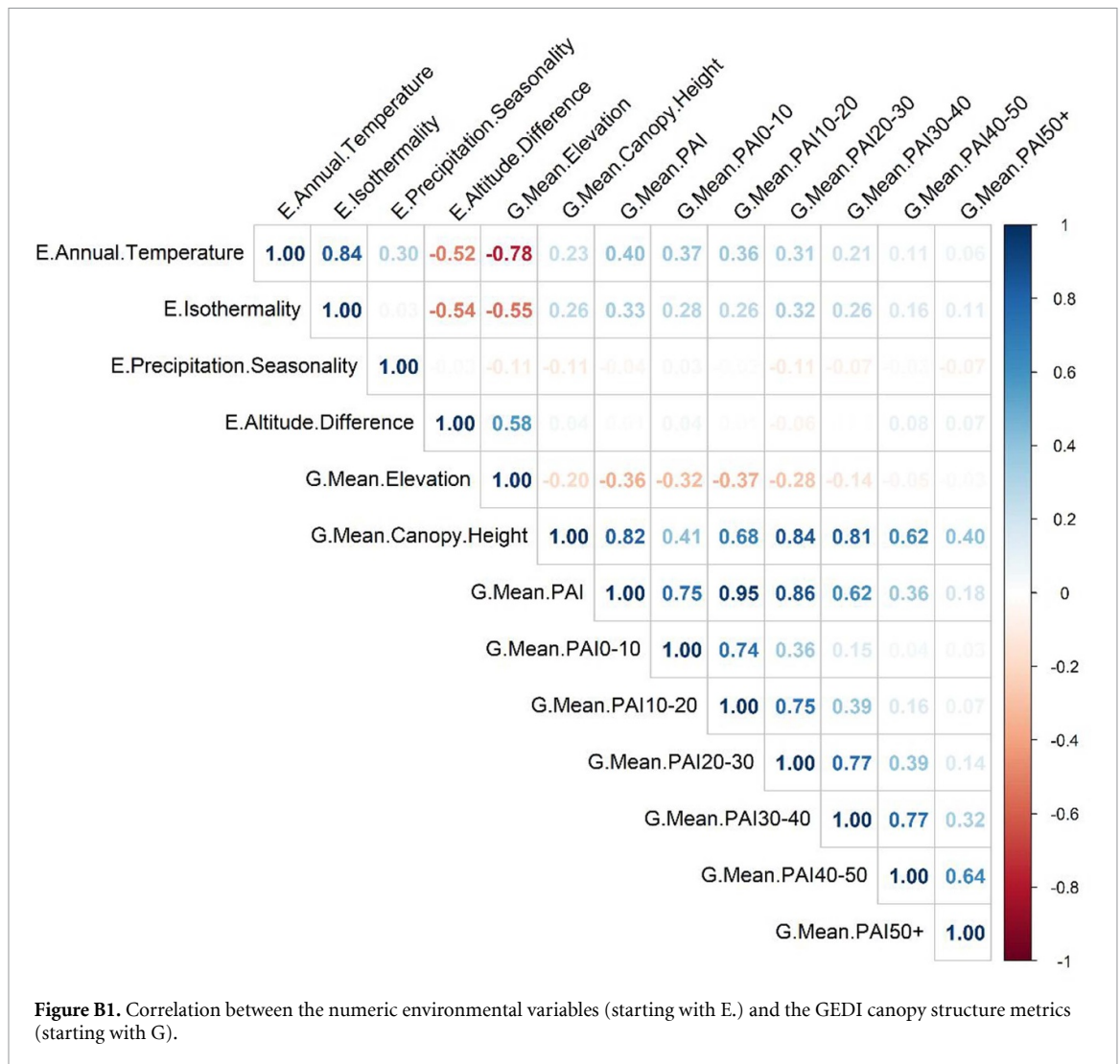
Figure B1. Correlation between the numeric environmental variables (starting with E.) and the GEDI canopy structure metrics (starting with G).

Appendix C. Relation of GEDI canopy metrics at different buffer zones

We included various buffer zones because of two main reasons: (a) it is not known what size buffer zone provides the most accurate modelling results and (b) the number of included plots would be higher with a larger buffer zone. The results show no real difference in model performance using canopy structure from the 1, 4 and 16 km² buffer zones. The estimates of canopy structure from the buffer zones are unbiased (figure C1). The canopy structure at the exact plot location may differ from the canopy structure in the buffer zones, however, the canopy structure at the

16 km² buffer zone is not biased, or in other words it does not describe a different canopy than the smaller buffer zones. This means that the canopy structure in each of the buffer zones could be regarded as the general forest structure at the plot location but does not necessarily constitute the small-scale variation in canopy structure at the plot location.

The heterogeneity of the various canopy structure measurements does increase with buffer zone (table C1), this can be explained by the larger sample area (area of the buffer zone), describing a larger piece of forest, this is to be expected given that forest structure decorrelates within just a couple hundreds of meters (Réjou-Méchain *et al* 2014).
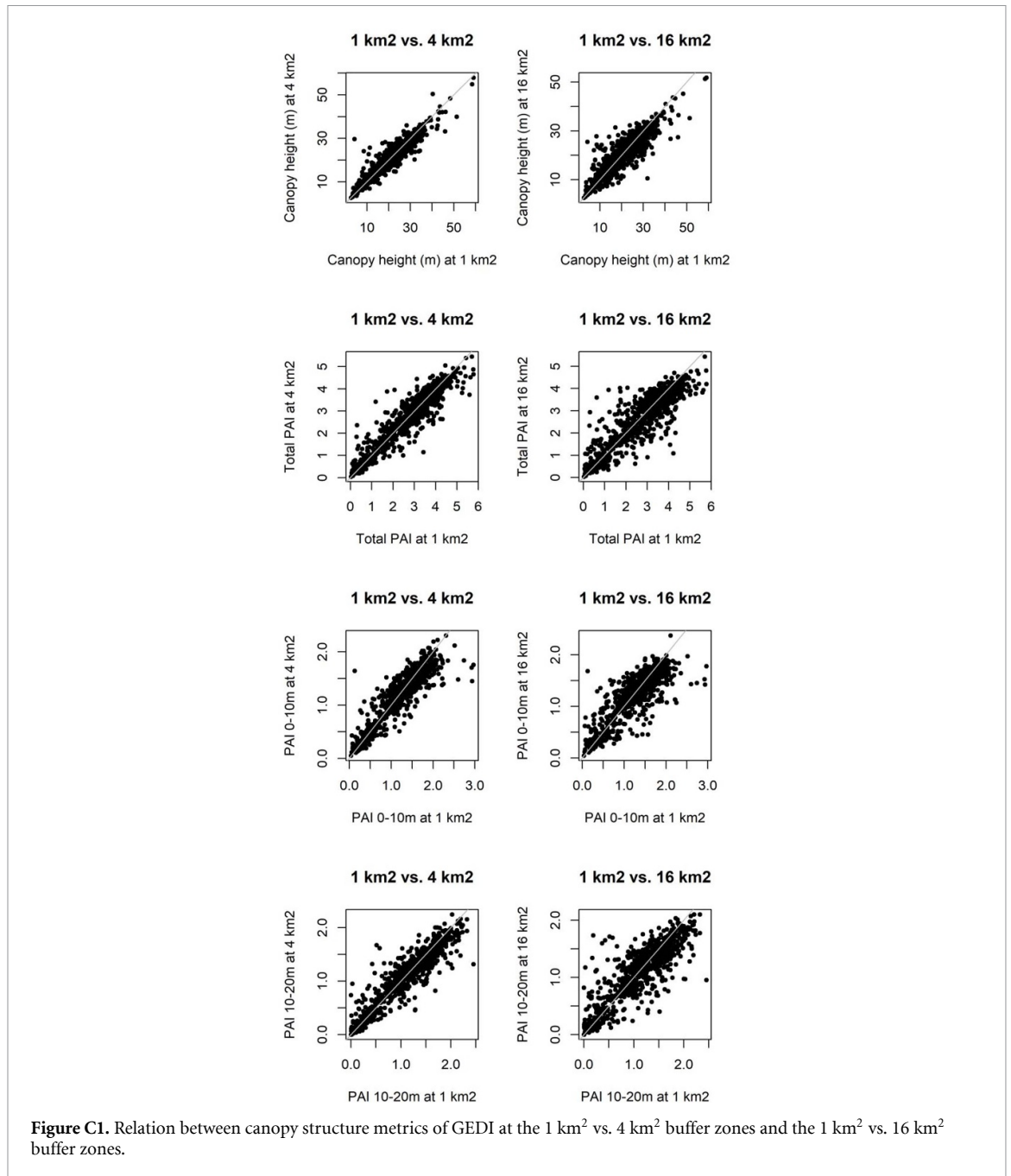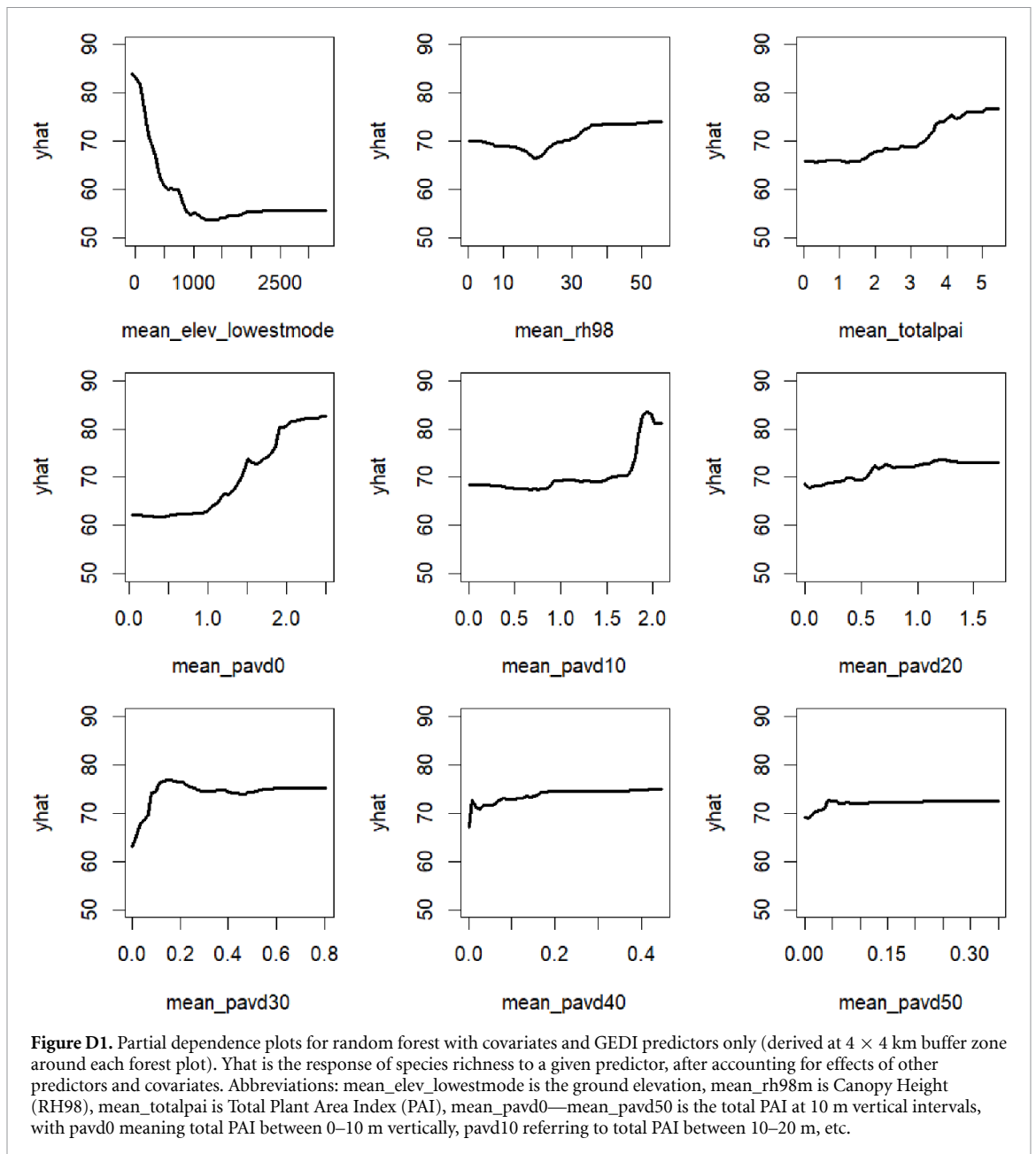
Figure C1. Relation between canopy structure metrics of GEDI at the 1 km^2 vs. 4 km^2 buffer zones and the 1 km^2 vs. 16 km^2 buffer zones.

Table C1. The heterogeneity of the canopy structure slightly increases with buffer zones size. The table indicates for four important canopy structure variables the average standard deviation of e.g. canopy height for all plots at that specific buffer zone. This value is computed as the average of all buffer zones, using the standard deviation of all GEDI footprints within a buffer zone for the specified metric.

	1 km^2	4 km^2	16 km^2
Canopy height (m)	7.00	7.55	8.14
Total PAI	1.30	1.37	1.42
PAI between 0 and 10 m	0.75	0.78	0.80
PAI between 10 and 20 m	0.60	0.63	0.65

Figure D1. Partial dependence plots for random forest with covariates and GEDI predictors only (derived at 4 × 4 km buffer zone around each forest plot). Yhat is the response of species richness to a given predictor, after accounting for effects of other predictors and covariates. Abbreviations: mean_elev_lowestmode is the ground elevation, mean_rh98m is Canopy Height (RH98), mean_totalpai is Total Plant Area Index (PAI), mean_pavd0—mean_pavd50 is the total PAI at 10 m vertical intervals, with pavd0 meaning total PAI between 0–10 m vertically, pavd10 referring to total PAI between 10–20 m, etc.

Appendix D. Partial dependence plots

The partial dependence plots show the response of species richness to a given predictor, after accounting for effects of other predictors and covariates (figure D1: GEDI predictors, figure D2: Environmental predictors). For example, mean total PAI has a positive relationship with species richness.
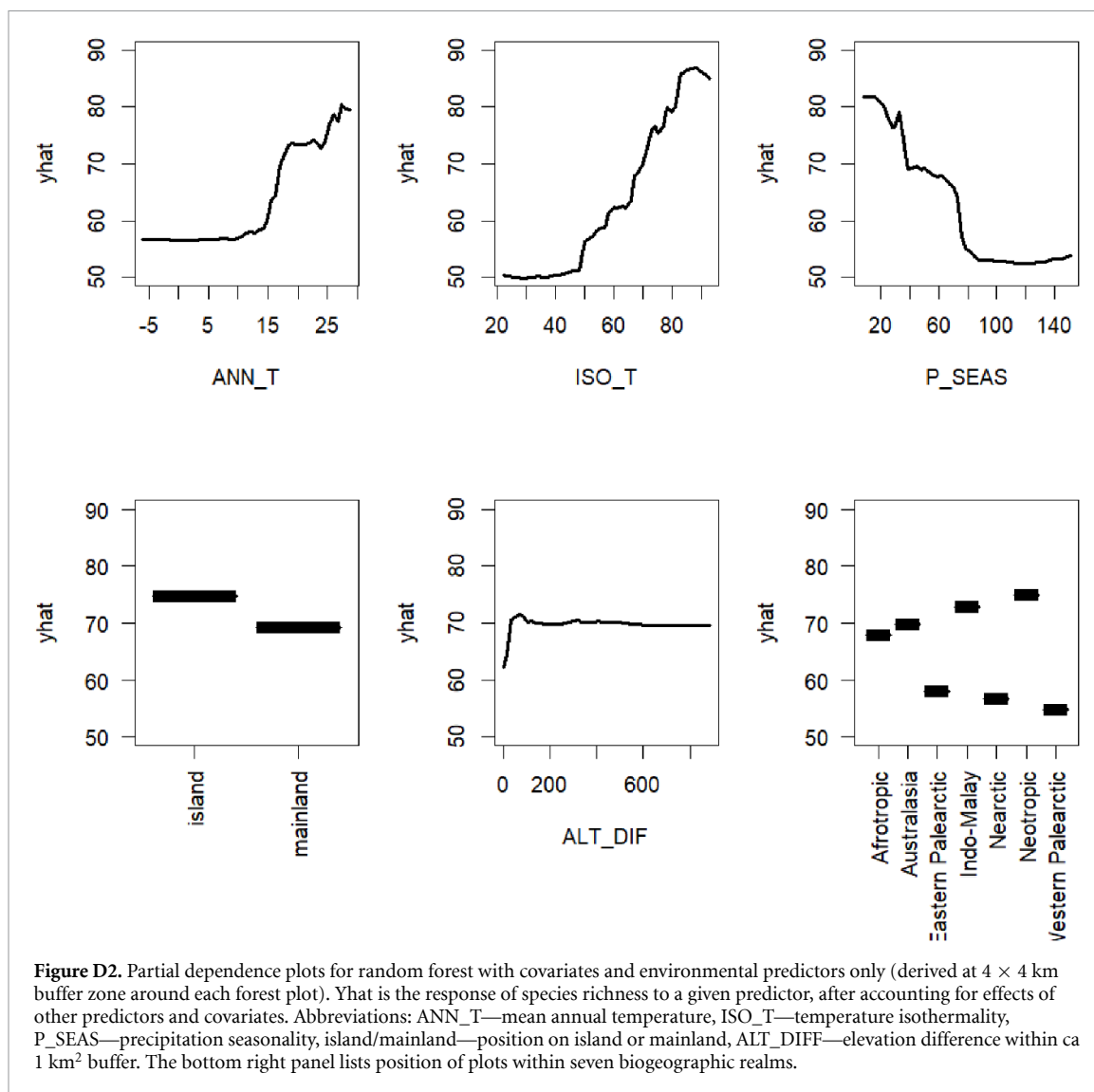
Figure D2. Partial dependence plots for random forest with covariates and environmental predictors only (derived at 4 × 4 km buffer zone around each forest plot). Yhat is the response of species richness to a given predictor, after accounting for effects of other predictors and covariates. Abbreviations: ANN_T—mean annual temperature, ISO_T—temperature isothermality, P_SEAS—precipitation seasonality, island/mainland—position on island or mainland, ALT_DIFF—elevation difference within ca 1 km^2 buffer. The bottom right panel lists position of plots within seven biogeographic realms.

ORCID iD

Suzanne M Marselis ⓘ https://orcid.org/0000-0001-9187-8799

References

Abbott I 1984 Comparisons of spatial pattern, structure, and tree composition between virgin and cut-over jarrah forest in Western Australia *For. Ecol. Manage.* **9** 101–26

Adam J H 2001 Changes in forest community structures of tropical montane rain forest on the slope of Mt. Trus Madi in Sabah, Malaysia *J. Trop. For. Sci.* **13** 76–92

Addo-Fordjour P, Obeng S, Anning A and Addo M 2009 Floristic composition, structure and natural regeneration in a moist semi-deciduous forest following anthropogenic disturbances and plant invasion *Int. J. Biodivers. Conserv.* **1** 21–37

Adekunle V A J 2006 Conservation of tree species diversity in tropical rainforest ecosystem of South-West Nigeria *J. Trop. For. Sci.* **18** 91–101

Algar A C, Kharouba H M, Young E R and Kerr J T 2009 Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods *Ecography* **32** 22–33

Anderson-Teixeira K J *et al* 2015 CTFS-ForestGEO: a worldwide network monitoring forests in an era of global change *Glob. Change Biol.* **21** 528–49

Ansley S J-A and Battles J J 1998 Forest composition, structure, and change in an old-growth mixed conifer forest in the northern Sierra Nevada *J. Torrey Bot. Soc.* **125** 297–308

Ayyappan N and Parthasarathy N 1999 Biodiversity inventory of trees in a large scale permanent plot of tropical evergreen forest at Varagaliar. Anamalais, Western Ghats, India *Biodivers. Conserv.* **8** 1533–54

Beals E W 1965 The remnant cedar forests of Lebanon *J. Ecol.* **53** 679–94

Bonino E E and Araujo P 2005 Structural differences between a primary and a secondary forest in the Argentine Dry Chaco and management implications *For. Ecol. Manage.* **206** 407–12

Borcard D, Legendre P and Drapeau P 1992 Partialling out the spatial component of ecological variation *Ecology* **73** 1045–55

Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32

Cairns M A, Olmsted I, Granados J and Argaez J 2003 Composition and aboveground tree biomass of a dry semi-evergreen forest on Mexico's Yucatan Peninsula *For. Ecol. Manage.* **186** 125–32

Cao M and Zhang J 1997 Tree species diversity of tropical forest vegetation in Xishuangbanna, SW China *Biodivers. Conserv.* **6** 995–1006

Chase J M and Leibold M A 2003 *Ecological Niches: Linking Classical and Contemporary Approaches* (Chicago, IL: University of Chicago)

Cheng-Yang Z, Zeng-Li L I U and Jing-Yun F 2004 Tree species diversity along latitudinal gradient on southeastern and northwestern slopes of Mt. Huanggang, Wuyi Mountains, Fujian, China *Biodivers. Sci.* **12** 63–74

Clinebell R R, Phillips O L, Gentry A H, Stark N and Zuuring H 1995 Prediction of neotropical tree and liana species richness from soil and climatic data *Biodivers. Conserv.* **4** 56–90

Coelho de Souza F *et al* 2016 Evolutionary heritage influences Amazon tree ecology *Proc. R. Soc.* B **283** 20161587

Craven D *et al* 2020 A cross-scale assessment of productivity–diversity relationships *Glob. Ecol. Biogeogr.* **29** 1940–55

Danielson J J and Gesch D B 2011 *Global Multi-Resolution Terrain Elevation Data 2010 (GMTED2010)* (US Department of the Interior, US Geological Survey)

Davis M A, Curran C, Tietmeyer A and Miller A 2005 Dynamic tree aggregation patterns in a species-poor temperate woodland disturbed by fire *J. Veg. Sci.* **16** 167–74

De Natale F, Floris A, Gasparini P, Scrinzi G, Tabacchi G and Tosi V 2005 *Inventario nazionale delle foreste e dei serbatoi forestali di carbonio* (Trento: Ispettorato Generale del Corpo Forestale dello Stato, CRA-ISAFA)

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** 100002

Dubayah R, Hofton M, Blair J, Armston J, Tang H and Luthcke S 2021a GEDI L2A elevation and height metrics data global footprint level V002 [Data set] (https://doi.org/10.5067/GEDI/GEDI02_A.002)

Dubayah R, Tang H, Armston J, Luthcke S, Hofton M and Blair J 2021b GEDI L2B canopy cover and vertical profile metrics data global footprint level V002 [Data set] (https://doi.org/10.5067/GEDI/GEDI02_B.002)

Eichhorn M 2010 Boreal forests of Kamchatka: structure and composition *Forests* **1** 154–76

Enoki T 2003 Microtopography and distribution of canopy trees in a subtropical evergreen broad-leaved forest in the northern part of Okinawa Island, Japan *Ecol. Res.* **18** 103–13

Eshete A, Sterck F and Bongers F 2011 Diversity and production of Ethiopian dry woodlands explained by climate- and soil-stress gradients *For. Ecol. Manage.* **261** 1499–509

Fashing P J, Forrestel A, Scully C and Cords M 2004 Long-term tree population dynamics and their implications for the conservation of the Kakamega Forest, Kenya *Biodivers. Conserv.* **13** 753–71

Francis A P and Currie D J 1998 Global patterns of tree species richness in moist forests: another look *Oikos* **81** 598

Frolking S, Palace M W, Clark D B, Chambers J Q, Shugart H H and Hurtt G C 2009 Forest disturbance and recovery: a general review in the context of spaceborne remote sensing of impacts on aboveground biomass and canopy structure *J. Geophys. Res. Biogeosci.* **114** n/a-n/a

Ganzhorn J U, Malcomber S, Andrianantoanina O and Goodman S M 1997 Habitat characteristics and lemur species richness in Madagascar1 *Biotropica* **29** 331–43

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

Graham A W 2006 *The CSIRO Rainforest Permanent Plots of North Queensland—Site, Structural, Floristic and Edaphic Descriptions* (Cairns: CSIRO and the Cooperative Research Centre for Tropical Rainforest Ecology and Management. Rainforest CRC)

Hakkenberg C R and Goetz S J 2021 Climate mediates the relationship between plant biodiversity and forest structure across the United States *Glob. Ecol. Biogeogr.* **30** geb.13380

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

Hirayama K and Sakimoto M 2003 Spatial distribution of canopy and subcanopy species along a sloping topography in a cool-temperate conifer-hardwood forest in the snowy region of Japan *Ecol. Res.* **18** 443–54

Huang Y *et al* 2018 Impacts of species richness on productivity in a large-scale subtropical forest experiment *Science* **362** 80–83

IGN 2017 French National Forest Inventory (FNFI) (available at: http://inventaire-forestier.ign.fr/)

Iriondo J M, Albert M J and Escudero A 2003 Structural equation modelling: an alternative for assessing causal relationships in threatened plant populations *Biol. Conserv.* **113** 367–77

Jing-Yun F, Yi-De L I, Biao Z H U, Guo-Hua L I U and Guang-Yi Z 2004 Community structures and species richness in the montane rain forest of Jianfengling, Hainan Island, China *Biodivers. Sci.* **12** 29–43

Jucker T, Bouriaud O and Coomes D A 2015 Crown plasticity enables trees to optimize canopy packing in mixed-species forests *Funct. Ecol.* **29** 1078–86

Kay R F, Madden R H, Van Schaik C and Higdon D 1997 Primate species richness is determined by plant productivity: implications for conservation *Proc. Natl Acad. Sci.* **94** 13023–7

Keil 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

Kohira M, Ninomiya I, Ibrahim A Z and Latiff A 2001 Diversity, diameter structure and spatial pattern of trees in semi-evergreen rain forest of Langkawi island, Malaysia *J. Trop. For. Sci.* **13** 460–76

Kohyama T 1986 Tree size structure of stands and each species in primary warm-temperate rain forests of Southern Japan *Bot. Mag. Tokyo* **99** 267–79

Krishnamurthy Y L, Prakasha H M, Nanda A, Krishnappa M, Dattaraja H S and Suresh H S 2010 Vegetation structure and floristic composition of a tropical dry deciduous forest in Bhadra Wildlife Sanctuary, Karnataka, India *Trop. Ecol.* **51** 235–46

Lalfakawma S, Roy U K, Vanlalhriatpuia K and Vanalalhluna P C 2010 Community composition and tree population structure in undisturbed and disturbed tropical semi-evergreen forest stands of North-East India *Appl. Ecol. Environ. Res.* **7** 303–18

Lam T Y and Maguire D A 2012 Structural equation modeling: theory and applications in forest management *Int. J. For. Res.* **2012** 1–16

Linder P, Elfving B and Zackrisson O 1997 Stand structure and successional trends in virgin boreal forest reserves in Sweden *For. Ecol. Manage.* **98** 17–33

Lopes C G R, Ferraz E M N and de Araújo E L 2008 Physiognomic-structural characterization of dry- and humid-forest fragments (Atlantic Coastal Forest) in Pernambuco State, NE Brazil *Plant Ecol.* **198** 1–18

Lü X-T, Yin J and Tang J-W 2010 Structure, tree pecies diversity and composition of tropical seasonal rainforests in Xishuangbanna, South-West China *J. Trop. For. Sci.* **22** 260–70

Malizia A and Grau R 2006 Liana–host tree associations in a subtropical montane forest of north-western Argentina *J. Trop. Ecol.* **22** 331–9

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** 094013

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

Marselis S M and Keil P 2022 The use of GEDI canopy structure for explaining variation in tree species richness in natural forests [Data] Figshare (https://doi.org/10.6084/m9.figshare.16810864)

Maycock F P, Guzik J, Jankovic J, Shevera M and Carleton J T 2000 Composition, structure and ecological aspects of mesic old growth Carpathian deciduous forests of Slovakia, southern

Poland and the western Ukraine *Fragm. Florist. Geobot.* **45** 281–321

Nagel T A, Svoboda M, Rugani T and Diaci J 2010 Gap regeneration and replacement patterns in an old-growth Fagus–Abies forest of Bosnia–Herzegovina *Plant Ecol.* **208** 307–18

Namikawa K, Matsui T, Kobayashi M, Goto R and Kuramoto S 2010 Initial establishment and regeneration processes of an outlying isolated Fagus crenata Blume forest stand in the northernmost boundary of its range in Hokkaido, northern Japan *Plant Ecol.* **207** 161–74

Pfeifer M *et al* 2018 Tropical forest canopies and their relationships with climate and disturbance: results from a global dataset of consistent field-based measurements *For. Ecosyst.* **5** 7

Phillips O L, Martínez R V, Vargas P N, Monteagudo A L, Zans M E, Sánchez W G, Cruz A P, Timaná M, Yli-Halla M and Rose S 2003 Efficient plot-based floristic assessment of tropical forests *J. Trop. Ecol.* **19** 629–45

Phillips O L and Miller J S 2002 *Global Patterns of Plant Diversity* (San Luis, MO: Missouri Bot. Gard. Press)

Poorter L *et al* 2015 Diversity enhances carbon storage in tropical forests *Glob. Ecol. Biogeogr.* **24** 1314–28

Popradit A, Srisatit T, Kiratiprayoon S, Yoshimura J, Ishida A, Shiyomi M, Murayama T, Chantaranothai P, Outtaranakorn S and Phromma I 2015 Anthropogenic effects on a tropical forest according to the distance from human settlements *Sci. Rep.* **5** 14689

Potapov P *et al* 2021 Mapping global forest canopy height through integration of GEDI and Landsat data *Remote Sens. Environ.* **253** 112165

Qi W, Saarela S, Armston J, Ståhl G and Dubayah R 2019 Forest biomass estimation over three distinct forest types using TanDEM-X InSAR data and simulated GEDI lidar data *Remote Sens. Environ.* **232** 111283

Ramesh B R, Swaminath M H, Patil S V, Pélissier R, Venugopal P D, Aravajy S, Elouard C and Ramalingam S 2010 Forest stand structure and composition in 96 sites along environmental gradients in the central Western Ghats of India *Ecology* **91** 3118

Réjou-Méchain M *et al* 2014 Local spatial structure of forest biomass and its consequences for remote sensing of carbon stocks *Biogeosciences* **11** 6827–40

Ricklefs R E and He F 2016 Region effects influence local tree species diversity *Proc. Natl Acad. Sci.* **113** 674–9

Round P, Pierce A, Sankamethawee W and Gale G 2011 The Mo Singto forest dynamics plot, Khao Yai National Park, Thailand *Nat. Hist. Bull. Siam Soc.* **57** 57–80

Sanchez M, Pedroni F, Eisenlohr P V and Oliveira-Filho A T 2013 Changes in tree community composition and structure of Atlantic rain forest on a slope of the Serra do Mar range, southeastern Brazil, from near sea level to 1000m of altitude *Flora: Morphol. Distrib. Funct. Ecol. Plants* **208** 184–96

Sawada H, Ohkubo T, Kaji M and Oomura K 2005 Spatial distribution and topographic dependence of vegetation types and tree populations of natural forests in the Chichibu Mountains, central Japan *J. Jpn. For. Soc.* **87** 293–303

Schoener T W 1974 Resource partitioning in ecological communities *Science* **185** 27–39

Senf C, Mori A S, Müller J and Seidl R 2020 The response of canopy height diversity to natural disturbances in two temperate forest landscapes *Landsc. Ecol.* **35** 2101–12

Sheil D and Salim A 2004 Forest tree persistence, elephants, and stem scars *Biotropica* **36** 505–21

Shu-Qing Z, Jing-Yun F, Shi-Long P, Zhan-Jiang Z, Xiao-Pu W U and Tao G U 2004 Structure and species diversity of boreal forests in Mt. Baikalu, Huzhong area, Daxing'an Mountains, Northeast China *Biodivers. Sci.* **12** 182–9

Šímová I, Storch D, Keil P, Boyle B, Phillips O L and Enquist B J 2011 Global species-energy relationship in forest plots: role of abundance, temperature and species climatic tolerances *Glob. Ecol. Biogeogr.* **20** 842–56

Specht A and Specht R L 1994 Biodiversity of overstorey trees in relation to canopy productivity and stand density in the climatic gradient from warm temperate to tropical Australia *Biodivers. Lett.* **2** 39

Splechtna B E, Gratzer G and Black B A 2005 Disturbance history of a European old-growth mixed-species forest—a spatial dendro-ecological analysis *J. Veg. Sci.* **16** 511–22

Sullivan M J P *et al* 2017 Diversity and carbon storage across the tropical forest biome *Sci. Rep.* **7** 39102

Szwagrzyk J and Gazda A 2007 Above-ground standing biomass and tree species diversity in natural stands of Central Europe *J. Veg. Sci.* **18** 555–62

Taubert F, Fischer R, Groeneveld J, Lehmann S, Müller M S, Rödig E, Wiegand T and Huth A 2018 Global patterns of tropical forest fragmentation *Nature* **554** 519–22

Tropek R, Sedláček O, Beck J, Keil P, Musilova Z, Imova I and Storch D 2014 Comment on 'High-resolution global maps of 21st-century forest cover change' *Science* **344** 981

USDA 2016 *Forest Inventory and Analysis—Fiscal Year 2016 Bussiness Report* (US Department of Agriculture) (available at: www.fia.fs.fed.us/)

van Do T, Sato T, Saito S, Kozan O, Yamagawa H, Nagamatsu D, Nishimura N and Manabe T 2015 Effects of micro-topographies on stand structure and tree species diversity in an old-growth evergreen broad-leaved forest, southwestern Japan *Glob. Ecol. Conserv.* **4** 185–96

Večeřa M *et al* 2019 Alpha diversity of vascular plants in European forests *J. Biogeogr.* **46** 1919–35

Viana D S, Keil P and Jeliazkov A 2021 Disentangling spatial and environmental effects: flexible methods for community ecology and macroecology (bioRxiv:871251)

Weigelt P, Jetz W and Kreft H 2013 Bioclimatic and physical characterization of the world's islands *Proc. Natl Acad. Sci.* **110** 15307–12

Wu X-P, Zhu B and Zhao S-Q 2004 Comparison of community structure and species diversity of mixed forests of deciduous broad-leaved tree and Korean pine in Northeast China *Biodivers. Sci.* **12** 174–81

Wusheng X, Tao D, Shihong L and Li X 2015 A comparison of tree species diversity in two subtropical forests, Guangxi, Southwest China *J. Resour. Ecol.* **6** 208–16

Yamada I 1976 Forest ecological studies of the montane forest of Mt. Pangrango, West Java: II. Stratification and floristic composition of the forest vegetation of the higher part of Mt. Pangrango *South East Asian Stud.* **13** 513–34

Yasuoka H 2008 The variety of forest vegetations in south-eastern Cameroon, with special reference to the availability of wild yams for the forest hunter-gatherers *Afr. Study Monogr.* **30** 89–119