Research paper draft

Structure metrics to generalize biomass estimation from lidar across forest types from different continents

Nikolai Knapp [a], Rico Fischer [a], Victor Cazcarra-Bes [b], Andreas Huth [a,c,d]

a) Department of Ecological Modeling, Helmholtz Centre for Environmental Research (UFZ), 04318 Leipzig, Germany; nikolai.knapp@ufz.de, rico.fischer@ufz.de, andreas.huth@ufz.de

b) Microwaves and Radar Institute, German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany; victor.cazcarrabes@dlr.de

c) Institute for Environmental Systems Research, Department of Mathematics / Computer Science, University of Osnabrück, 49076 Osnabrück, Germany

d) German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, 04103 Leipzig, Germany

Corresponding author: Nikolai Knapp (nikolai.knapp@ufz.de)

Abstract

Forest aboveground biomass is a key variable in remote sensing based forest monitoring. Active sensor systems, such as lidar, can generate detailed canopy height products. Relationships between canopy height and biomass are commonly established via regression analysis using information from ground-truth plots. In this way, many site-specific height-biomass relationships have been proposed in the literature and applied for mapping in regional contexts. However, such relationships are only valid within the specific forest type for which they were calibrated. A generalized relationship would facilitate biomass estimation across forest types and regions. In this study, a combination of lidar-derived and ancillary structural descriptors is proposed as an approach for generalization between forest types. Each descriptor is supposed to quantify a different aspect of forest structure, i.e., mean canopy height, maximum canopy height, maximum stand density, vertical heterogeneity and wood density. Airborne discrete return lidar data covering 194 ha of forest inventory plots from five different sites including temperate and tropical forests from Africa, Europe, North, Central and South America was used. Biomass predictions using the best general model (nRMSE = 12.4%, R^2 = 0.74) were found to be almost as accurate as predictions using five site-specific models (nRMSE = 11.6%, R^2 = 0.78). The results further allow interpretation about the importance of the employed structure descriptors in the biomass estimation and the mechanisms behind the relationships. Understanding the relationship between canopy structure and aboveground biomass and being able to generalize it across forest types are important steps towards consistent large scale biomass mapping and monitoring using airborne and potentially also spaceborne platforms.

Keywords: forest structure; aboveground biomass; canopy height; lidar; generalization

1. Introduction

Quantifying global carbon stocks of forests as well as their changes over time requires spatially explicit measurements and monitoring (Harris et al., 2012). The primary variable of interest hereby is forest aboveground biomass (AGB). Thus, there is a growing amount of literature about estimating forest biomass from canopy height metrics derived from light detection and ranging (lidar), synthetic aperture radar or photogrammetry (Asner and Mascaro, 2014; Goetz and Dubayah, 2011; Lu et al., 2014; Treuhaft et al., 2015; Zolkos et al., 2013). The majority of these studies investigated data from specific forest sites with the goal to find the best prediction model, i.e., maximizing explained variability (e.g., R^2), minimizing prediction error (e.g., RMSE) and minimizing systematic bias, while using the most parsimonious set of predictor variables (Zolkos et al., 2013). Different statistical approaches have been used including multiple linear regression models and machine learning methods (Fassnacht et al., 2014). As a result various site-specific and forest type-specific relations for biomass estimation have been proposed and applied successfully for biomass mapping at regional scale.

Comparatively few studies have tried to seek for generalization in the estimation approaches. In those studies conducted in the temperate and boreal biomes, field plots ranging from 500 to 1000 m^2 in size were used and reported estimation uncertainties ranged from 2.5 to 23% (Bouvier et al., 2015; Lefsky et al., 2002; Magnussen et al., 2012). In similar studies conducted in the tropical biome, field plots were commonly in the size range from 0.1 to 1 ha and reported estimation uncertainties were around 10% at the 1-ha scale (Gregory P. Asner et al., 2012; Asner and Mascaro, 2014; Vincent et al., 2014, 2012) For consistent global mapping of forest biomass, however, it would be desirable to have more generic relationships which are applicable across different forest types and biogeographic regions. Such an approach would contribute to a better understanding of how structural attributes differ between forest types and how they are related to biomass. It would also facilitate biomass mapping across the globe. Given the variety of metrics that can be derived from lidar data (Næsset, 2002), it would further be desirable to have a minimum set of meaningful metrics, describing different aspects of forest structure, to avoid problems with multicollinearity and extensive model selection procedures (Bouvier et al., 2015).

A widely used general approach is the one proposed by Asner et al. (2012) and modified by Asner & Mascaro (2014) for pan-tropical application. The function is inspired by individual tree allometry, where tree AGB can be modeled as a multiplicative power law of tree diameter at breast height (DBH; or tree basal area BA), tree height and species-specific wood density (Chave et al., 2014). Hence, as a stand level equivalent for area-based AGB estimation they used a power law of stand BA sum, mean top-of-canopy height (TCH) and average wood density. The approach further assumes a linear relationship between BA and TCH, which may differ between regions, and regional differences in average wood density can be considered. This approach has been established using data from different tropical regions (Hawaii, Panama, Peru, Madagascar) and has been applied successfully in other tropical regions, e.g., Colombia (G. P. Asner et al., 2012), Malaysia (Coomes et al., 2017), Tanzania (Getzin et al., 2017), with uncertainties between 12 and 15% at 1-ha scale.

Bouvier et al. (2015) suggested a different model for generalized AGB estimation. They used several a priori defined lidar-based metrics that captured different aspects of forest structure. Their model was able to produce accurate AGB estimations for different forest types in France. However, site-specific coefficients led to higher prediction accuracies for each site, compared to using only one set of coefficients across all sites.

In this study, we attempted to find a model that is generally applicable throughout different forest types and even different biomes by including structural information on forest stands. Tropical forests in Panama, French Guiana and Gabon were analyzed along with temperate forests in the United States of America and Germany. A generalized approach should also perform well across the gradients from intact to disturbed and natural to managed forests without requiring any stratification. Therefore, our analysis incorporated plots of primary and managed forests. We tried to estimate BA and AGB at the 1-ha scale. Usually, the two are closely correlated. However, BA is a simple inventory-derived metric, which is easily comparable among sites and studies (Vincent et al., 2012). Inventory-based AGB, on the other hand, is more complex to compute. Assumptions about allometric relationships and wood density values are required to derive single tree AGB. It can lead to considerable differences in stand AGB if different assumptions are chosen for the same stand (Duncanson et al., 2017). It has been argued that BA (or its two components stem number and quadratic mean stem diameter) is the essential variable that has to be derived from remote sensing and AGB could then be estimated based on predicted BA and the regional tree diameter-height allometries and wood densities (Vincent et al., 2014, 2012). Here, we chose to analyze both variables – BA for its robustness and comparability and AGB as the major variable of interest in forest carbon mapping efforts.

We hypothesized that the following structural forest attributes may contribute to explaining stand BA and AGB: a) mean canopy height, b) maximum possible stand density c) maximum possible tree height, d) vertical canopy heterogeneity, and for AGB, additionally, e) average wood density. Most of these structural attributes can be quantified in several alternative ways. Thus, for one attribute there may be a set of several candidate metrics. Some of the metrics can be derived from lidar, while others require ancillary information. In this analysis, data from 194 ha of temperate and tropical forest from five megaplot sites were combined with the following goals: 1) to find a generic approach for BA and AGB estimation that can be applied across all sites without causing prediction bias at any individual site and 2) to investigate the contributions of the different structural attributes.

2. Material & Methods

2.1. Study Sites

Data from five forest sites covering different forest types and biogeographical zones were used (Tab. 1). Four study sites are part of the ForestGEO megaplot network (Anderson-Teixeira et al., 2015) and thus they have been inventoried according to a standard protocol. The data structure for the fifth site, Paracou, is similar to the data structure of the ForestGEO sites. For each tree, the diameter at breast height (DBH), spatial position and species identity were recorded. In this study, only trees with a DBH ≥ 10 cm were considered, because 10 cm was the minimum DBH recorded at Paracou (Blanc et al., 2009), and all given numbers refer to trees above this size threshold. In the following, each of the five sites is briefly described.

1) Barro Colorado Island (BCI), Panama, is a Central American lowland tropical moist forest site with an annual precipitation of 2580 mm and an average temperature of 27.1 °C. The census on the 50-ha plot was conducted in 2010 (Condit, 1998; Condit et al., 2012; Hubbell et al., 1999) and comprised 22,084 trees, which belonged to 223 species. Trees of the six families Malvaceae (17%), Fabaceae (13%), Moraceae (7%), Euphorbiaceae (6%), Rubiaceae (6%) and Meliaceae (6%) account for 56% of AGB with *Quararibea asterolepis* being the single species with the highest contribution (6%).

2) Paracou, French Guiana, is a South American lowland tropical rainforest with an annual precipitation of 3040 mm and an average temperature of 26 °C. There are 16 large plots with plots 1 to 15 having an extent of 250 m × 250 m each and plot 16 having an extent of 500 m × 500 m. In 1986 and 1987, selective logging with different treatment intensities (timber logging, fuelwood logging, thinning) was conducted on some of the plots, while others have served as control plots (Hérault and Piponiot, 2018). Since our analysis was conducted for 100 m × 100 m units, only subareas of 200 × 200 m measured from the south-western corners of plots 1 to 15 were used. In total, 85 ha from Paracou were analyzed. Censuses were conducted in 2015 and comprised 53,501 trees of 713 species. Trees of the three families Fabaceae (25%),

Lecythidaceae (15%) and Chrysobalanaceae (11%) account for 51% of AGB with *Eperua falcata* being the single species with the highest contribution (7%).

3) Rabi, Gabon, is a Central African lowland tropical rainforest site with an annual precipitation of 2300 mm and an average temperature of 26 °C. The census on the 25-ha plot was conducted from 2010 to 2012 (Labrière et al., 2018) and comprised 12,019 trees, which belonged to 235 species. Trees of the three families Fabaceae (38%), Ochnaceae (9%) and Simaroubaceae (7%) account for 54% of AGB with *Lophira alata* being the single species with the highest contribution (9%).

4) The Smithsonian Environmental Research Center (SERC) plot, United States of America, is a North American deciduous broadleaved temperate forest site with an annual precipitation of 1070 mm and an average temperature of 13.2 °C. The census on the 16-ha plot was conducted in 2014 (Král et al., 2016; McMahon and Parker, 2015) and comprised 4,719 trees, which belonged to 39 species. Trees of the six species *Liriodendron tulipifera* (27%), *Liquidamber styraciflua* (16%), *Fagus grandifolia* (11%), *Quercus alba* (8%), *Carya alba* (7%) and *Fraxinus pennsylvanica* (6%) together account for 76% of the AGB.

5) Traunstein, Germany, is a Central European managed mixed temperate forest site, which includes conifer and broadleaf plantations. It has an annual precipitation of 1240 mm and an average temperature of 7.6 °C. The census on the 25-ha plot was conducted from 2015 to 2016. Due to the shape of the plot a rectangular 18-ha subarea was selected for the analysis. It comprised 7,182 trees, which belonged to 25 species. Trees of the four species *Picea abies* (39%), *Fagus sylvatica* (22%), *Acer pseudoplatanus* (19%) and *Abies alba* (15%) together account for 94% of the AGB.

Tab. 1: Information about the study sites

Site	Region	Forest type	Size [ha]	Location	Year of inventory	Year of lidar scan	Basal area range [m^2 ha^{-1}]
BCI	Central Panama	Neotropical moist	50	9.15° N, 79.85° W	2010	2009	17.3 – 38.5
Paracou	Northern French Guiana	Neotropical wet	85	5.27° N, 52.92° W	2015	2015	24.8 – 38.7
Rabi	Western Gabon	Afrotropical wet	25	1.92° S, 9.88° E	2010-2012	2015	20.8 – 36.7
SERC	Eastern USA	Nearctic temperate broadleaf	16	38.89° N, 76.56° W	2014	2017	26.2 – 43.5
Traunstein	Southern Germany	Palearctic temperate mixed	18	47.94° N, 12.67° E	2015-2016	2016	7 – 44.6

2.2. Inventory Data

The inventory data was processed to calculate AGB of each tree. Based on species, wood density values were assigned to each tree using the ForestGEO wood density database (http://ctfs.si.edu/Public/Datasets/CTFSWoodDensity/) and, in the case of Paracou, the Global Wood Density Database (Chave et al., 2009; Zanne et al., 2009). Wood density values for 82% of all trees were available at species level. For the remaining trees, median wood densities at genus (14.1%) or family level (3.7%) or the overall median at the site (0.2%) were assigned, respectively. From DBH, the height of each tree was calculated using site specific asymptotic allometric relationships. These relationships were derived by fitting regression models of the Michaelis-Menten type (Equation 1) to the diameter height dataset from Jucker et al. (2017), grouped by biogeographical region and forest type.

Equation 1: $$H = \frac{h_{max} \cdot D}{d_{1/2} + D}$$

This equation describes tree height H [m] as a function of DBH D [m] with two parameters: 1) h_{max} [m], which is the asymptotic maximal possible tree height, and 2) $d_{1/2}$ [m], which is the DBH of a tree that has reached a height of half of h_{max}. A verification of whether the derived models describe the DBH-height relations at each site reasonably well, was done by plotting the curves together with the maximal observed DBH and maximal (lidar-derived) height of each hectare (Fig. S1). This showed that the relationships match the observed values for the BCI, Paracou, Rabi and Traunstein plot, but strongly underestimate the tree heights at SERC. Thus, for SERC we discarded the diameter-height relationship obtained from the dataset and instead obtained parameters by directly fitting a regression model to the data points in Fig. S1, representing the lidar-derived maximal heights on each hectare. All height allometry parameters used are listed in Tab. 2.

Tab. 2: Parameters for the different diameter-height relationships modeled with a Michaelis-Menten equation.

Site	Species group	h_{max} [m]	$d_{1/2}$ [m]
BCI	All	57.4	0.43
Paracou	All	57.4	0.43
Rabi	All	59.9	0.48
SERC (discarded)	All	37	0.22
SERC (used instead)	All	54.7	0.27
Traunstein	Broadleafs	48.8	0.25
Traunstein	Conifers	68.9	0.5

Aboveground biomass (AGB [t]) of each tree was calculated according to the general allometric equation suggested by (Chave et al., 2014) ("Model 5") (Equation 2) with DBH D [m], height H [m] and wood density WD [t m^{-3}].

Equation 2: $$AGB = 0.559 \cdot D^2 \cdot H \cdot WD$$

2.3. Lidar Data

Small footprint discrete return lidar data was collected in BCI in August 2009 using an Optech ALTM Gemini sensor with a mean point density of 21 m^{-2} (Lobo and Dalling, 2014), in Paracou in October 2015 using an Riegl LMS Q 780 with a mean point density of 54.6 m^{-2}, in Rabi in 2015 using a Riegl VQ-480i sensor with a mean point density of 2.5 m^{-2} (Labrière et al., 2018), at SERC in July 2017 using a Riegl VQ-480i sensor with a mean point density of 54.1 m^{-2} (Cook et al., 2013) and in Traunstein in August 2016 using a Riegl LMS Q 680i sensor with a mean point density of 20.8 m^{-2}. The lidar point clouds were terrain-normalized using LAStools (Isenburg, 2011) and rasterized to canopy height models (CHM) with 1-m resolution, by taking the height of the highest return in each 1-m^2-cell (Fig. 1). No interpolation was used and cells with no return were filled with value zero (ground height).

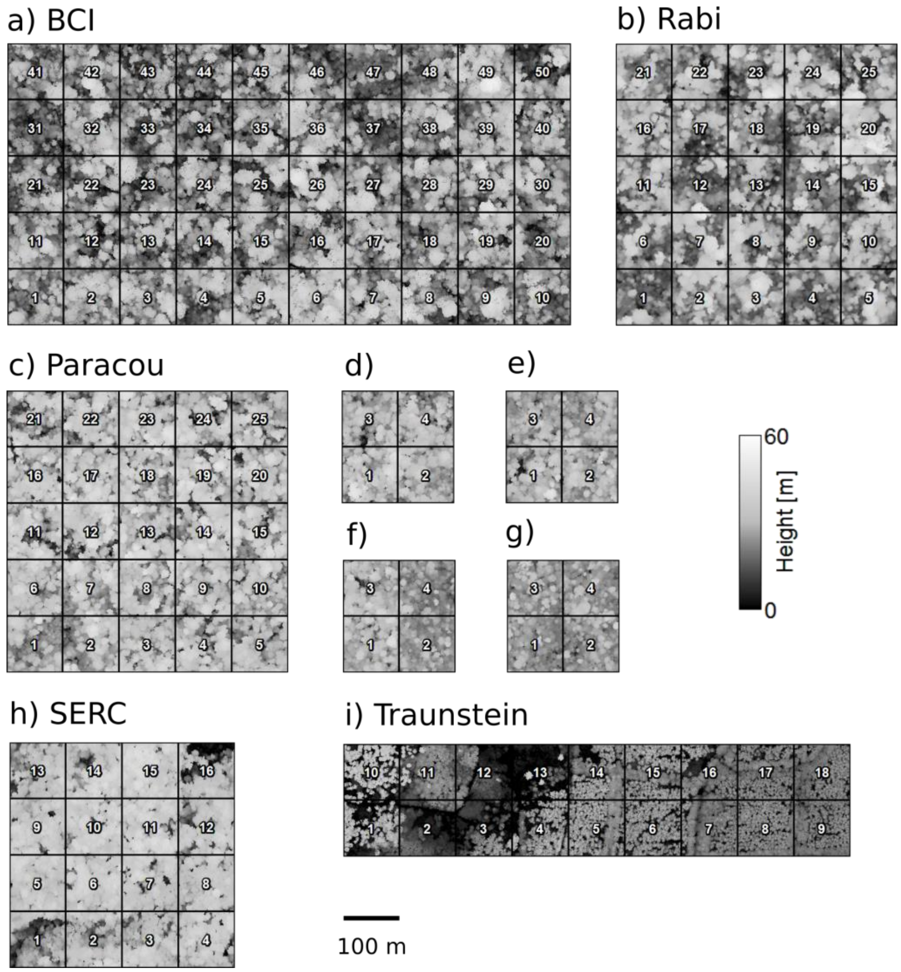

Fig. 1: Canopy height models of the five study sites: a) Barro Colorado Island, b) Rabi, c) Paracou Plot 16 (biodiversity plot), d) Paracou Plot 1 (control plot), e) Paracou Plot 2 (selective logging), f) Paracou Plot 3 (selective logging and timber stand improvement), g) Paracou Plot 4 (selective logging, timber stand improvement and fuelwood collection), h) Smithsonian Environmental Research Center and i) Traunstein. The black grids and numbers represent the 1-ha subplots with each ha representing one record in the analysis.

2.4. Forest Structure Metrics

All inventory plots were divided into square-shaped subplots of 1-ha size each. At 1-ha scale a variety of structural metrics was calculated from 1) the inventory data and 2) the lidar data. Inventory-based metrics included basal area sum (BA), number of stems per ha (N), quadratic mean tree diameter (at breast height, D_g), maximum DBH per ha (D_{max}), mean wood density weighted for tree basal area (WD_{BA}) or weighted for tree aboveground volume (WD_{AGV}) and stand density index (SDI, Equation 3), which is a standardized metric for stocking (Reineke, 1933).

Equation 3:
$$SDI = N \cdot \left(\frac{25}{D_g}\right)^{-1.605}$$

Additionally to those metrics derived at 1-ha-level, we derived a set of metrics at site-level. Those site-level metrics included maximum basal area sum (BA_{smax}) and maximum SDI (SDI_{smax}) of all the 1-ha plots at each site s and basal area-weighted mean wood density (WD_{sBA}) and aboveground volume-weighted mean wood density (WD_{sAGV}) across all trees at each site s.

Lidar-based metrics were maximum canopy height per 1-ha plot H_{max} and per site H_{smax}, mean top-of-canopy height from CHMs of two different resolutions (1-m and 10-m pixels called TCH_1 and TCH_{10}), standard deviation of the 1-m CHM (SD_{CHM}), coefficient of variation of the 1-m CHM (CV_{CHM}) and Gini index of the 1-m CHM ($Gini_{CHM}$). The vertical foliage profile (VFP) was derived from the vertical profile of the 1-m CHM following the approach described by Harding et al (2001) (Equation 4). Despite being originally developed for large footprint waveforms, the method can also be applied to vertical profiles of CHMs, since the latter have been shown to closely match coinciding waveforms (Blair and Hofton, 1999).

Equation 4:
$$VFP(h_i) = \frac{1}{k \cdot \Delta h} \cdot \ln\left(\frac{GP(h_i)}{GP(h_{i+1})}\right)$$

with k being the light extinction coefficient, Δh the width of one height bin (here 1 m) and $GP(h_i)$ the gap probability (value of the cumulative CHM profile) in height bin h_i (Ni-Meister et al., 2001). All pixels below 5 m height were regarded as ground and k was set to 0.3 for all sites. The parameter k can be described as the quotient of a projection coefficient G, which is 0.5 for a random leaf angle distribution, and a clumping index C, which on average is 1.58 for different forest types (Tang et al., 2012). A value of k = 0.3 has been shown to result in good LAI estimations (Getzin et al., 2017). The same vertical distribution metrics as for the CHMs were derived from the VFP, namely SD_{VFP}, CV_{VFP} and $Gini_{VFP}$. All metrics used for statistical modeling are listed in Tab. 3 and equations to calculate them are detailed in Fischer et al. (2019).

Tab. 3: List of metrics used in the statistical analysis (CHM = canopy height model).

Acronym	Explanation of metric	Structural aspect	Predictor (P_x)
AGB	Aboveground biomass [t ha^{-1}]	Target variable	-
BA	Basal area [m^2 ha^{-1}]	Target variable	-
TCH_1	Mean top-of-canopy height from 1-m pixel CHM [m]	Mean canopy height	P_h
TCH_{10}	Mean top-of-canopy height from 10-m pixel CHM [m]	Mean canopy height	P_h
BA_{smax}	Maximal basal area at site s [m^2 ha^{-1}]	Maximum stand density	P_d
SDI_{smax}	Maximal stand density index at site s [ha^{-1}]	Maximum stand density	P_d
H_{smax}	Maximal canopy height at site s [m]	Maximum canopy height	P_m
SD_{CHM}	Standard deviation of the 1-m pixel CHM [m]	Vertical heterogeneity	P_v
CV_{CHM}	Coefficient of variation of the 1-m pixel CHM [m]	Vertical heterogeneity	P_v
$Gini_{CHM}$	Gini index of the 1-m pixel CHM [m]	Vertical heterogeneity	P_x
SD_{VFP}	Standard deviation of the vertical foliage profile [m]	Vertical heterogeneity	P_v
CV_{VFP}	Coefficient of variation of the vertical foliage profile [m]	Vertical heterogeneity	P_v
$Gini_{VFP}$	Gini index of the vertical foliage profile [m]	Vertical heterogeneity	P_v
WD_{sAGV}	Mean aboveground volume-weighted wood density at site s [t m^{-3}]	Wood density	P_w
WD_{sBA}	Mean basal area-weighted wood density at site s [t m^{-3}]	Wood density	P_w

2.5. Multivariate Regression Analysis

Regression analysis was conducted to find the best relationship and set of predictor variables for BA and AGB estimation with the main objective to minimize overall root mean squared error (RMSE) across sites. The regression models had the functional form of multivariate power laws. Each predictor was supposed to capture a different structural aspect of the forest. Several candidate metrics were grouped into sets of potential predictors and tested in different combinations. By categorizing metrics into different structural aspects the number of possible metric combinations for testing was reduced. Combinations of redundant metrics, i.e., metrics that capture the same structural aspect, were not tested. For BA, a four predictor equation was used (Equation 5). For AGB, a five predictor equation was used (Equation 6).

Equation 5: $$BA = a_0 \cdot {P_h}^{a_h} \cdot {P_d}^{a_d} \cdot {P_m}^{a_m} \cdot {P_v}^{a_v}$$

Equation 6: $$AGB = b_0 \cdot {P_h}^{b_h} \cdot {P_d}^{b_d} \cdot {P_m}^{b_m} \cdot {P_v}^{b_v} \cdot {P_w}^{b_w}$$

Each P_x represents a predictor for a certain structural aspect x with a_x being the coefficients for BA estimation and b_x being the coefficients for AGB estimation. For each P_x one single metric from a set of possible predictors was used in each regression model. The predictor sets were defined as follows: The predictor for mean canopy height P_h was either TCH_1 or TCH_{10}. The maximum possible canopy height P_m was exclusively represented by H_{smax}. The predictor for maximum density (stocking) P_d was either BA_{smax} or SDI_{smax}. For vertical heterogeneity of the canopy P_v, SD_{CHM}, $CV_{CHM,}$ $Gini_{CHM}$, SD_{VFP}, CV_{VFP} and $Gini_{VFP}$ have been explored. Average wood density P_w was only included in the AGB estimation and was either WD_{sBA} or WD_{sAGV}.

Correlations among predictors were highest within predictor groups P_x (Fig. S2, e.g., within P_h, P_v and P_w). Thus, by only combining predictors from different groups multicollinearity was reduced. Between groups, the highest correlations were observed between H_{smax} and either of the two density predictors BA_{smax} and SDI_{smax}, while the two among each other were less correlated. We do not expect that maximum possible canopy height and maximum stocking density of forests are generally related. We rather suspect that correlations between these site-level metrics occurred by chance due to the small sample size of five sites. Maximum likelihood parameter estimation in R (R Development Core Team, 2014) was used to derive the coefficients for Equations 5 and 6. All possible metric combinations were tested including all possible subsets discarding one or several predictors P_x. The goodness-of-fit was evaluated based on linear regression of predictions against ground-based observations of the dependent variable to quantify R^2, RMSE and nRMSE (normalized RMSE by dividing it by the mean observed value). Wilcoxon tests were performed to check whether the mean prediction residual at each site deviated significantly from zero, with the goal of identifying unbiased prediction models. For each predictor combination, 1000 bootstrapping replicates were performed by resampling the dataset randomly with replacement. Site-level metrics BA_{smax}, SDI_{smax}, H_{smax} and WD_{sAGV} were recalculated based on the resampled dataset, i.e., if, e.g., the plot with the largest H_{max} of site s was not in the resampled dataset, H_{smax} was set to the largest H_{max} of any plots from site s present in the resampled dataset. Mean bootstrapped statistics ($RMSE_b$, $nRMSE_b$ and R^2_b) served to evaluate the different models. The best predictor combinations for BA and AGB were finally tested in a leave-one-site-out cross validation, in which five different models were fit on data from four sites, respectively, and the fifth site was set aside for testing.

2.6. Site-specific Reference Regression Models

To assess the performance of the derived general, site-independent, structure-based multi predictor regression models site-specific reference models were required. For this purpose, single predictor regression models were fit. As predictors for these reference models P_h, i.e., TCH_1 or TCH_{10}, were used. These metrics have been most widely used for this purpose (Asner and Mascaro, 2014; Knapp et al., 2018) and showed the highest individual correlations with the target variables BA and AGB (Fig. S2). The models given by Equations 7 and 8 were fit by splitting the dataset into five subsets with each subset containing only records from one site and using the same fitting procedure as described above.

Equation 7: $$BA = a_{0,site} \cdot {P_h}^{a_{h,site}}$$

Equation 8: $$AGB = b_{0,site} \cdot {P_h}^{b_{h,site}}$$

As an alternative, also site-specific multi predictor models are possible. However, such models make comparisons more complicated, because the set of best predictors may vary from site to site. In our analysis the predictors P_d, P_m and P_w were all at site level. Thus, the only relevant predictor for site-specific multi predictor models, apart from P_h, was P_v. The best model using a combination of P_h and P_v for each site are listed in the supplements (Tab. S2).

3. Results

3.1. Forest Structure at Different Sites

The different structure attributes at 1-ha scale varied within and among the five sites (Fig. 2 and Fig. S3). This section gives an overview on the distributions of the most important ground- and lidar-based structure metrics. BA values ranged from 7 to 44.6 m^2 ha^{-1} with a mean of 29.8 m^2 ha^{-1}. AGB values ranged from 76 to 638 t ha^{-1} with a mean of 354 t ha^{-1}. Mean top-of-canopy height ranged from 5.6 to 38 m when calculated from 1-m × 1-m pixels (TCH_1) and from 15.8 to 41.8 m when calculated from 10-m × 10-m pixels (TCH_{10}). In both cases, the distributions for the tropical sites were similar while TCH were on average higher at SERC and lower at Traunstein. Mean wood densities per hectare were calculated on a BA-weighted and on an AGV-weighted basis. Both were similar in their distributions and mean wood densities calculated for each site across the entire megaplot weighted by either BA or AGV were almost identical with the largest difference being 0.02 t m^{-3} in the case of BCI (WD_{sBA}: BCI: 0.51 t m^{-3}, Paracou: 0.69 t m^{-3}, Rabi: 0.66 t m^{-3}, SERC: 0.48 t m^{-3}, Traunstein: 0.5 t m^{-3}; WD_{sAGV}: BCI: 0.49 t m^{-3}, Paracou: 0.69 t m^{-3}, Rabi: 0.66 t m^{-3}, SERC: 0.47 t m^{-3}, Traunstein: 0.5 t m^{-3}). Due to this similarity, only AGV-weighted wood density at site level WD_{sAGV} was considered in the further analysis. Mean wood densities at Paracou and Rabi exceeded the values from all other sites strongly with WD_{AGV} at the 1-ha scale ranging from 0.6 to 0.74 t m^{-3} at Paracou and Rabi and from 0.42 to 0.55 t m^{-3} at all other sites. Stand density index values ranged from 138 to 778 with the lowest values occurring in recently managed parts of Traunstein. The maximal SDIs are proxies for the highest possible stocking density in the different forest types (BCI: 683, Paracou: 749, Rabi: 703, SERC: 708, Traunstein: 778). The maximum canopy height covered a wide range from 27.1 m to 54.7 m. H_{max} at SERC only covered a very narrow range falling inside the range of the tropical sites, while H_{max} at Traunstein were much lower. The maximum canopy heights per site H_{smax} were 54.7 m at BCI, 50 m at Paracou, 52.6 m at Rabi, 46.2 m at SERC and 40.3 m at Traunstein.

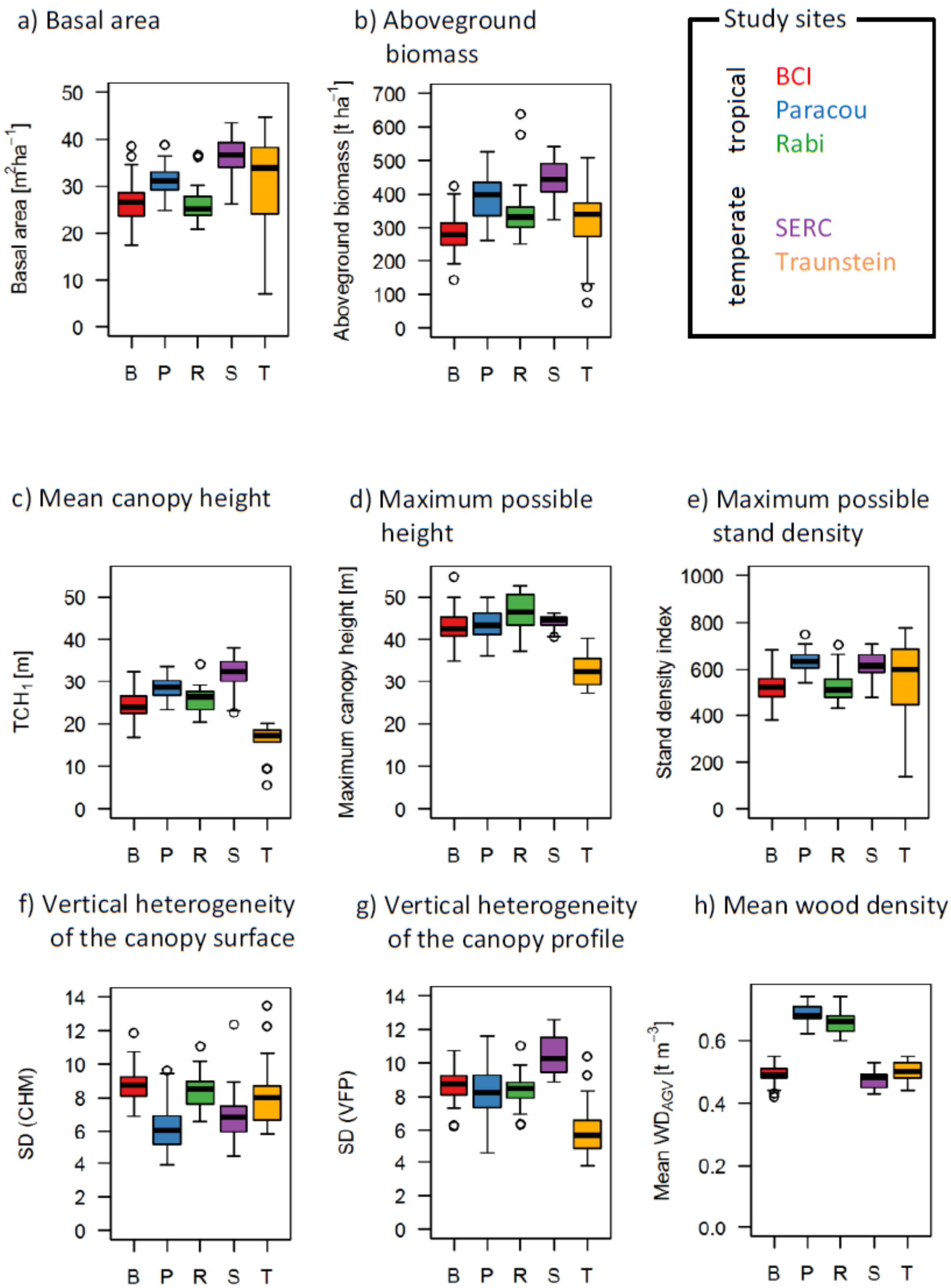

Fig. 2: Boxplots of the distributions of a selection of forest structure metrics across the five study sites Barro Colorado Island (B), Paracou (P), Rabi (R), Smithsonian Environmental Research Center (S) and Traunstein (T). Graphics (a) and (b) depict the two target variables basal area and aboveground biomass. Graphics (c) to (h) depict six possible predictor variables. Variables are partly lidar-derived (c, d, f, g) and partly field-derived (a, b, e, h).

The vertical heterogeneity was measured in several different ways using standard deviation, coefficient of variation and Gini index of the canopy height model and the vertical foliage profile, respectively. CHM- and VFP-based vertical structure metrics showed quite different distribution patterns. For the CHM-based metrics, Paracou and SERC showed the lowest values, due to a homogenous canopy surface, BCI and Rabi showed intermediate values, due to their rough canopy surface with large trees alternating with gaps and Traunstein showed (at least for CV and Gini index) the highest values, due to its heterogeneous structure composed of young and old stands interrupted by forest roads. For the VFP-based vertical structure metrics, Traunstein showed the lowest values, which is in accordance with the fact that large parts of the plot are single-layered stands of different age, while the other sites showed higher values, indicating a more complex, multi-layered canopy.

3.2. Basal Area Estimation

3.2.1. Site-specific Basal Area Estimation

Here, basal area was estimated from lidar using a single structural descriptor of stand height P_h. Mean top-of-canopy height at 1- and 10-m pixel resolution (TCH_1 and TCH_{10}) were tested as P_h to derive site-specific power law coefficients ($a_{0,site}$ and $a_{h,site}$). Coefficients for each site are listed in Tab. S1 and a scatterplot with site-specific curves is displayed in Fig. 3. The following goodness-of-fit statistics were derived across all sites using the site-specific relationships: The TCH_1-based basal area predictions resulted in RMSE = 2.5 m^2 ha^{-1} (8.3%) and R^2 = 0.79. The TCH_{10}-based basal area predictions resulted in RMSE = 2.8 m^2 ha^{-1} (9.5%) and R^2 = 0.73. In both cases, the mean residuals were not significantly different from zero at any site (Wilcoxon tests with the smallest p-value of all sites being p = 0.54). Using multi predictor models (different metrics of P_h and P_v), site-specific predictions with RMSE = 2.3 m^2 ha^{-1} (7.6%) and R^2 = 0.83 were achieved. However, the set of best metrics varied from site to site (Tab. S2).

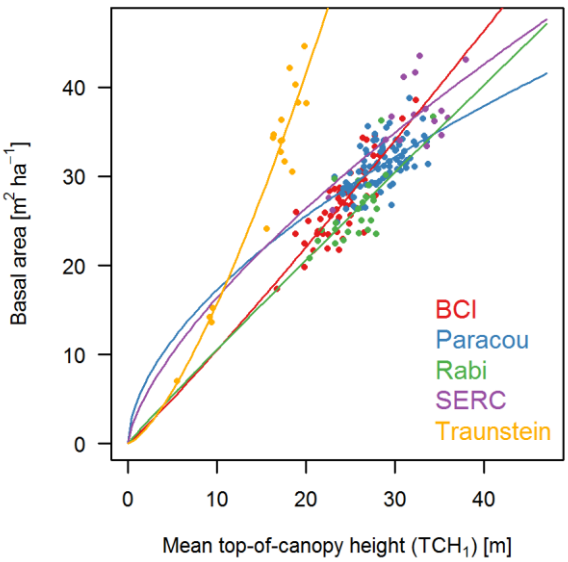

Fig. 3: Site-specific relationships (power laws) between basal area and TCH_1 with each point representing 1 ha.

3.2.2. Generalized Basal Area Estimation

Here, basal area was estimated using several structural descriptors from lidar, which were supposed to capture different aspects of forest structure (Equation 5). In total, 125 regression models consisting of different descriptors and metrics were analyzed. The best models found are listed in Tab. 4. The models are ranked according to increasing mean bootstrapped $RMSE_b$. The listed models represent the best overall model and the best models with certain structural descriptors P_x removed. For the different P_x in most cases the same metrics were selected. For P_h mostly TCH_{10} was selected and TCH_1 only occurred once. For P_d always the site specific maximal basal area BA_{smax} was selected. For P_m the site specific maximal tree height H_{smax} was the only available metric. For P_v, however, four different metrics appear in the list of best models, namely SD_{VFP}, $Gini_{VFP}$., CV_{VFP} and CV_{CHM}.

Tab. 4: The best basal area estimation models for different predictor combinations ranked for increasing mean bootstrapping root mean squared error ($RMSE_b$). For explanation of the variable names please refer to the main text.

Mean canopy height P_h	Maximal stand density P_d	Maximal canopy height P_m	Vertical heterogeneity P_v	RMSE	nRMSE	R^2	$RMSE_b$	$nRMSE_b$	R^2_b	Bias
TCH_{10}	BA_{smax}	H_{smax}	SD_{VFP}	2.9	9.8%	0.71	2.9	9.6%	0.72	yes
TCH_{10}	BA_{smax}	-	$Gini_{VFP}$	3.0	10.0%	0.70	3.1	10.4%	0.67	no
TCH_1	-	H_{smax}	-	3.3	10.9%	0.64	3.1	10.4%	0.67	yes
TCH_{10}	BA_{smax}	-	-	3.4	11.5%	0.60	3.6	12.0%	0.56	yes
TCH_{10}	-	-	CV_{VFP}	3.6	12.2%	0.55	3.6	12.0%	0.56	yes
-	BA_{smax}	-	CV_{CHM}	4.2	14.2%	0.40	4.2	14.1%	0.39	no
TCH_{10}	-	-	-	4.4	14.6%	0.36	4.3	14.5%	0.36	yes

Equation 9: $$BA = 9.2 \cdot {TCH_{10}}^{1.3} \cdot {BA_{smax}}^{0.359} \cdot {H_{smax}}^{-1.03} \cdot {SD_{VFP}}^{-0.305}$$

The overall best model was one using all four structural descriptors (nRMSE = 9.8%, Equation 9). The goodness-of-fit decreased only marginally if information on maximal possible height (P_m) was excluded from the predictors (nRMSE = 10%). The third best model was a two-predictor model using only current and maximal possible canopy height (P_h and P_m, nRMSE = 10.9%). Hence, there was no other three-predictor model that could exceed this two predictor model in accuracy. It was followed by a model using mean canopy height and maximal possible stand density (P_h and P_d, nRMSE = 11.5%), and one using mean canopy height and vertical heterogeneity (P_h and P_v, nRMSE = 12.2%). At the lower end, the best model making no use of mean canopy height (no P_h) was somewhat better (nRMSE = 14.2%) than the one using exclusively mean canopy height (P_h = TCH_{10}, nRMSE = 14.6%). Thus, adding any structural descriptor decreased nRMSE by at least 2.4% compared to a model purely based on canopy height.

The additional goal was finding a relationship that is unbiased across all sites. According to the Wilcoxon tests, predictions of the best found model were slightly but significantly biased for BCI ($p = 0.0036$) and Paracou ($p = 0.05$), whereas the second best model did not show any significant bias for any of the sites (Tab. S3). Fig. 4 shows the 1:1 plots for site-specific TCH_1-based predictions (a), the generalized TCH_{10}-based predictions (c) and the predictions using the best model based on structural descriptors (e). Fig. 4 also shows the residual distributions resulting from each of the three predictions for the different sites (b, d, f). In the leave-one-site-out cross validation of the best predictor combination, the overall nRMSE increased from 9.8% to 12.3% and moderate but significant biases were observed for three sites (Fig. S4).

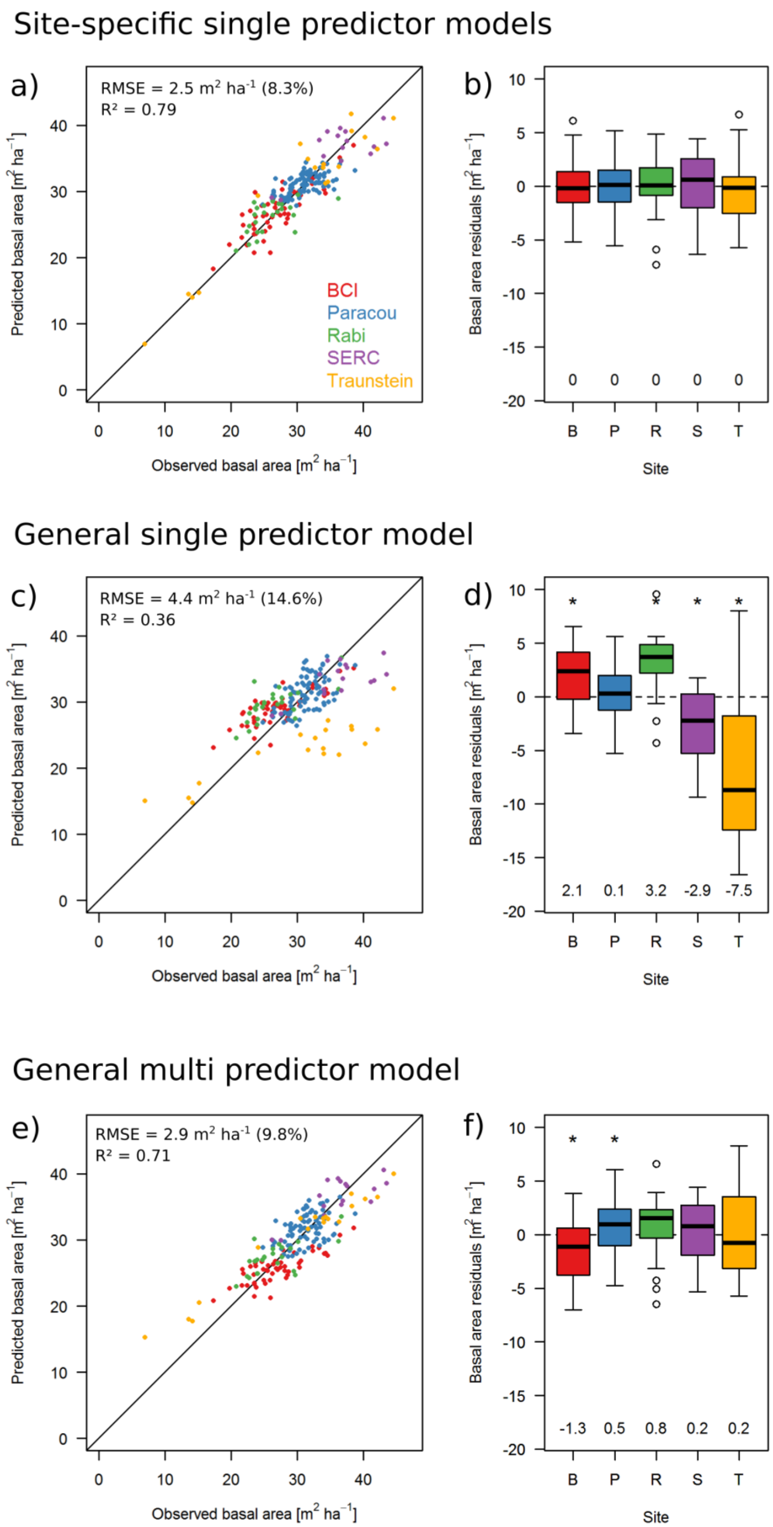

Fig. 4: Scatterplots of predicted basal area against observed basal area using a) site-specific single predictor models, c) a general single predictor model (based on TCH_{10}) and e) the best general multi predictor model (Equation 9). The boxplots on the right hand side show the distribution (quartiles) of prediction residuals at each site with numbers below displaying the mean residual value (bias) and asterisks above indicating whether the means deviate significantly from zero (b, d, f).

3.3. Aboveground Biomass Estimation

3.3.1. Site-specific Aboveground Biomass Estimation

Analogous to BA, AGB was modeled as a power law function of mean canopy height P_h, using either TCH_1 or TCH_{10} by fitting site-specific coefficients ($b_{0,site}$ and $b_{h,site}$; Tab. S1). Applying these site-specific relationships, the following goodness-of-fit statistics were derived across all sites: The TCH_1-based AGB predictions resulted in RMSE = 41 t ha^{-1} (11.6%) and R^2 = 0.78. The TCH_{10}-based AGB predictions resulted in RMSE = 41.8 t ha^{-1} (11.8%) and R^2 = 0.77. In both cases, the mean prediction residuals were not significantly different from zero at all sites (Wilcoxon tests with the smallest p-value of all sites being p = 0.39). Using multi predictor models (different metrics of P_h and P_v), site-specific predictions with RMSE = 38.1 t ha^{-1} (10.8%) and R^2 = 0.81 were achieved. However, the set of best metrics varied from site to site (Tab. S2).

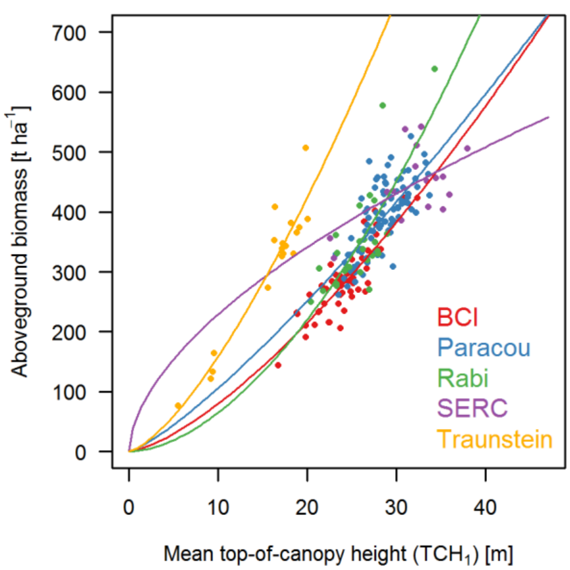

Fig. 5: Site-specific relationships (power laws) between aboveground biomass and TCH_1 with each point representing 1 ha.

3.3.2. Generalized Aboveground Biomass Estimation

To derive a generalized AGB estimation model the same structural descriptors as for basal area were used. Additionally, a fifth descriptor for the average wood density P_w was introduced, which resulted in 251 models in total. Tab. 5 lists the best models found for different combinations of structural descriptors. The models are ranked according to increasing $RMSE_b$ (derived from bootstrapping). The listed models represent the best overall model and the best models with a reduced number of structural descriptors P_x. Compared to the basal area estimation more descriptor combinations are possible due to the additional parameter P_w. TCH_{10} was selected in most cases for P_h. For P_d maximal basal area BA_{smax} and maximal stand density index SDI_{smax} per site do appear in the models. For P_v three different metrics have been selected: SD_{CHM}, SD_{VFP} and CV_{VFP}.

Tab. 5: The best aboveground biomass estimation models for different predictor combinations ranked for increasing mean bootstrapping root mean squared error ($RMSE_b$). For explanation of the variable names please refer to the main text.

Mean canopy height P_h	Maximal stand density P_d	Maximal canopy height P_m	Vertical heterogeneity P_v	Mean wood density P_w	RMSE	nRMSE	R^2	$RMSE_b$	$nRMSE_b$	R^2_b	Bias
TCH_1	SDI_{smax}	H_{smax}	SD_{CHM}	WD_{sAGV}	44.0	12.4%	0.74	46.1	13.0%	0.71	no
TCH_{10}	BA_{smax}	H_{smax}	SD_{VFP}	-	47.4	13.4%	0.70	47.9	13.5%	0.69	yes
TCH_{10}	SDI_{smax}	H_{smax}	-	WD_{sAGV}	45.6	12.9%	0.73	48.5	13.7%	0.69	no
TCH_{10}	SDI_{smax}	H_{smax}	-	-	46.0	13.0%	0.72	48.5	13.7%	0.69	yes
TCH_1	-	H_{smax}	-	-	47.7	13.5%	0.70	48.9	13.8%	0.68	yes
TCH_{10}	BA_{smax}	-	CV_{VFP}	WD_{sAGV}	51.2	14.5%	0.67	49.7	14.0%	0.67	yes
TCH_{10}	BA_{smax}	-	-	WD_{sAGV}	60.1	17.0%	0.55	50.6	14.3%	0.66	yes
TCH_{10}	SDI_{smax}	-	-	-	50.9	14.4%	0.66	50.7	14.3%	0.66	yes
TCH_{10}	-	-	CV_{VFP}	-	54.3	15.3%	0.61	53.9	15.2%	0.61	yes
TCH_{10}	-	-	-	WD_{sAGV}	57.2	16.1%	0.57	57.2	16.1%	0.57	yes
TCH_{10}	-	-	-	-	58.5	16.5%	0.55	58.3	16.4%	0.55	yes
-	BA_{smax}	-	SD_{VFP}	-	63.0	17.8%	0.48	62.2	17.5%	0.49	yes

In the case of aboveground biomass estimation the best model was the one using all five available structural descriptors (Equation 10; nRMSE = 12.4%).

Equation 10: $$AGB = 1.92 \cdot {TCH_1}^{1} \cdot {SDI_{smax}}^{0.979} \cdot {H_{smax}}^{-1.24} \cdot {SD_{CHM}}^{0.212} \cdot {WD_{sAGV}}^{0.0838}$$

Leaving either P_v, P_w or both aside increased the nRMSE by around 1%. The best model that did not rely on any site related ground-based information was the one using only P_h and P_m (nRMSE = 13.5%). Tab. 5 documents the results for other descriptor combinations. E.g., a single predictor model based on TCH_{10} only had nRMSE = 16.5%. P_d and P_m were more important than P_v (according to their presence in the best models), which is different from the basal area estimation where P_v was more important than P_d and P_m. The best model without P_h had an nRMSE = 17.8%.

With regard to the goal of finding a relationship that is unbiased across all sites, the Wilcoxon tests identified two models for which the mean of residuals at none of the single sites differed significantly from zero (Tab. S4): the models in lines 1 and 3 in Tab. 5. For all other models, predictions were biased for at least one site. Fig. 6 shows the 1:1 plots for site-specific TCH_1-based predictions (a), the generalized TCH_{10}-based predictions (c) and the structure guided predictions using the best model (e). Fig. 6 also shows the residual distributions resulting from each of the predictions for the different sites (b, d, f). In the leave-one-site-out cross validation of the best predictor combination, the overall nRMSE increased from 12.4% to 24.1% and significant overestimations were observed for all sites but Rabi (Fig. S4).

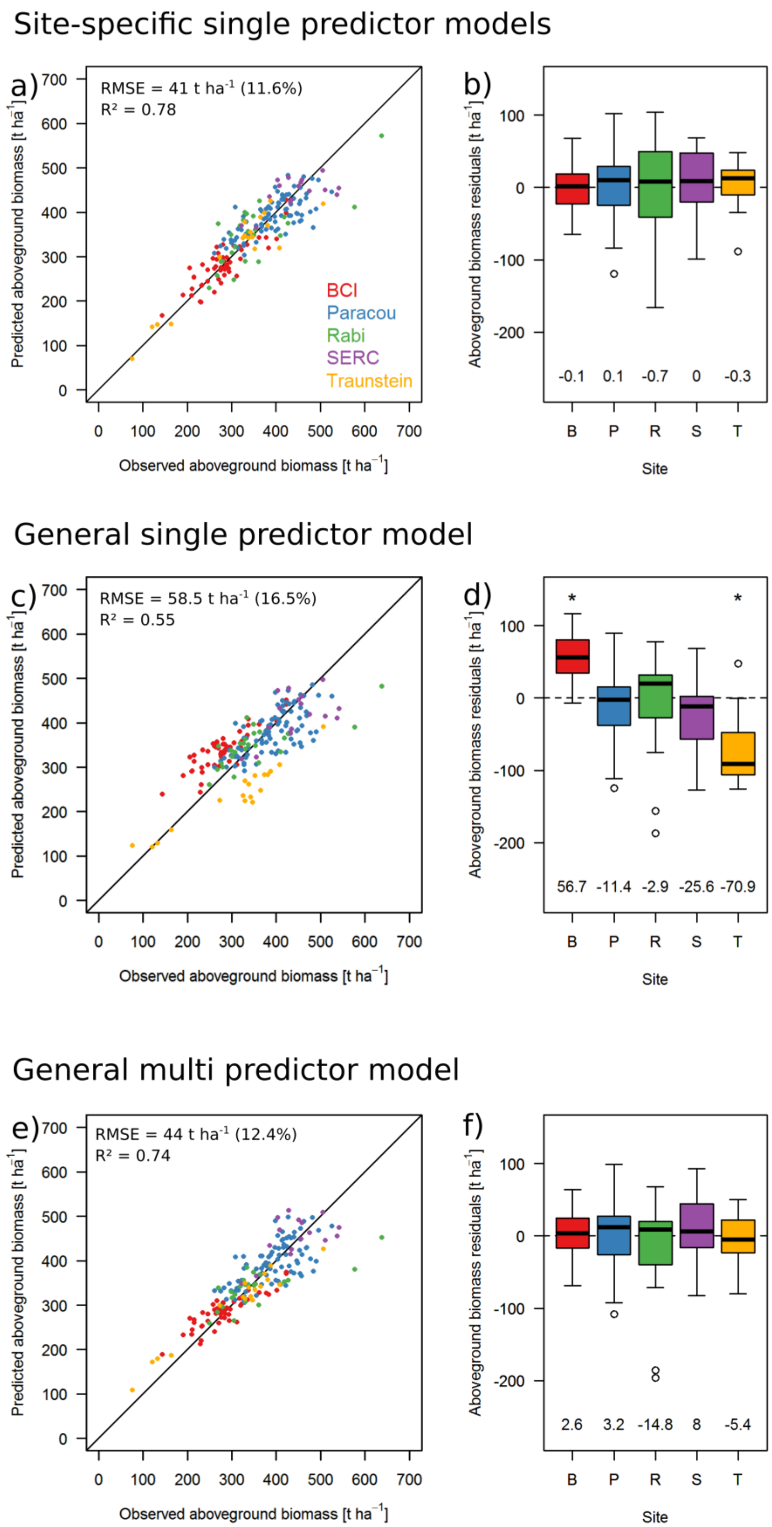

Fig. 6: Scatterplots of predicted aboveground biomass against observed aboveground biomass using a) a site specific single predictor model, c) a general single predictor model (based on TCH_{10}) and e) the best general multi predictor model (Equation 10). The boxplots on the right hand side show the distributions (quartiles) of prediction residuals at each site with numbers below displaying the mean residual value (bias) and asterisks above indicating whether the means deviate significantly from zero.

3.4. Comparison of Results

Overall, achieved relative errors in BA estimation were somewhat lower than the ones for AGB estimation. The exclusion of different structural descriptors led to an increase in estimation errors. Fig. 7 shows the obtained nRMSE for different sets of P_x in comparison to the nRMSE of site-specific estimations. For BA estimation, the best unbiased generic model required four coefficients and resulted in an nRMSE of 10.4%, which is 2.1% higher than the nRMSE of 8.3% obtained from five site-specific models, requiring ten (two per site) coefficients. For AGB estimation, the best unbiased generic model required six coefficients and resulted in an nRMSE of 13%, which is 1.4% higher than the nRMSE of 11.6% obtained from five site-specific models, requiring ten coefficients.

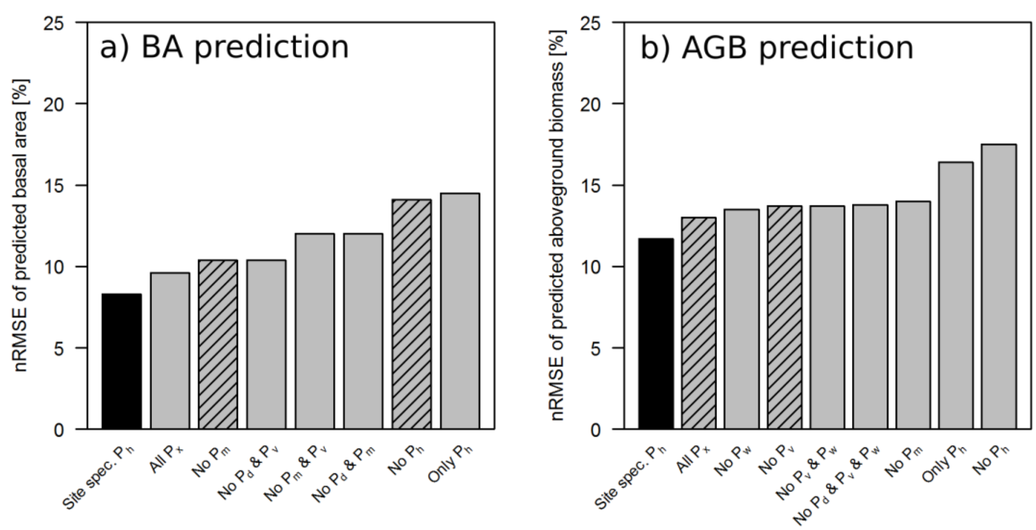

Fig. 7: Summary of how the exclusion of certain structural descriptors P_x influence the normalized root mean squared error (nRMSE) of basal area (a) and aboveground biomass estimation (b). The black bar represents the site-specific reference model. The grey bars represent mean bootstrapping nRMSE of the different generic models. The striped bars mark the models which produce unbiased predictions at all sites. For the meaning of the indices of the predictors please refer to the main text.

4. Discussion

The goal of this study was to determine a set of forest structure metrics that can be used for BA and AGB estimation from CHMs at very distinct forest sites, which belong to different biomes. It could be shown that a combination of four metrics capturing mean canopy height, maximal stand density, maximal canopy height and vertical heterogeneity could estimate BA using a generic model across all sites with a high accuracy, which was almost as good as the accuracy achieved by site-specific models. The accuracy for AGB estimation was slightly weaker than the one for BA estimation.

4.1. The Role of Mean Canopy Height

It was found that the mean canopy height (P_h), represented here by TCH_1 and TCH_{10}, was the most important predictor variable, which is in support of its wide use in previous studies (Asner and Mascaro, 2014; Duncanson et al., 2015; Lefsky et al., 2002). It was important in BA and AGB estimation, with accuracies decreasing considerably when P_h was dropped from the models. Despite the mathematical simplicity of TCH (the mean height of all CHM pixels) it is a quite comprehensive metric capturing much of the forest structure in a single number. It is influenced by the heights and crown sizes of the trees (which contribute to the CHM) and therefore closely related to Lorey's height (BA-weighted mean tree height) (Asner and Mascaro, 2014). However, TCH also provides information on horizontal vegetation density, if ground pixels, e.g., in canopy gaps, are included in its computation (Lu et al., 2014). There have been studies that tried to separate the "height" and "density" aspect of TCH by calculating mean canopy height only from canopy pixels (and excluding ground pixels) and capturing horizontal vegetation density as fractional canopy cover, i.e. the relative proportion of canopy pixels above an arbitrary height threshold, or its inverse, the gap fraction (Bouvier et al., 2015). It has also been shown that fractional canopy cover alone can predict AGB in tropical forests quite well over a range of canopy height thresholds (Meyer et al., 2018).

It was found that TCH_{10} derived from a rough 10-m-pixel CHM often performed better than TCH_1 derived from 1-m pixels. This has been observed also in an earlier study at BCI using TCH in single predictor models (Knapp et al., 2018). It might be explained by the ability of TCH_{10} to capture the canopy structure of the large trees, which also contribute most to BA and AGB, and the larger gaps where such trees are missing. TCH_1 includes more detail and is influenced by the structure of individual tree crowns and small gaps within and between crowns, which might not be relevant or even counterproductive for estimating BA and AGB. In particular, in the context of generalization between different forest types it might be beneficial that TCH_{10} "ignores" differences in crown shapes.

4.2. The Role of Stand Density

Maximal stand density per site (P_d) was of high importance for AGB and BA estimation. Only one in the best eight AGB models did not contain P_d. Asner & Mascaro (2014) pointed out that for many sites BA shows a linear relationship with TCH, but with considerable differences in the slopes, which was therefore an important term in their AGB estimation model. Differences in this relationship can be expected because at different sites different tree species may occur, which have different geometries, in particular regarding the relationships between DBH and height and DBH and crown diameter. Of two stands with the same canopy height, one may contain trees with slender crowns and has a much higher stocking than the other one containing trees with wide crowns. We tried to reduce the necessary information about density as much as possible by only using the maximum observed value per site. As this parameter is not derived from remote sensing data, either inventory data or expert knowledge on the maximum possible density of the forest type would be required. As metrics for P_d, BA_{smax} and SDI_{smax} were used. The two are independent from each other: The highest SDI identifies the stand with the highest stocking according to the self-thinning rule (Reineke, 1933), which is not necessarily the stand with the highest current BA (Fig. S5). Among the five sites investigated, the tropical sites had lower BA_{smax} (38.5, 38.7, and 36.7 m^2 ha^-

[1]) than the temperate sites (43.5 and 44.6 m^2 ha^{-1}). SDI_{smax}, however, was similar at BCI, Rabi and SERC (683, 703, 708) and somewhat higher at Paracou (749) and Traunstein (778). As shown, either of the two metrics could improve the AGB estimation in comparison to the case of missing P_d.

4.3. The Role of Maximum Height

An inclusion of maximum possible height (P_m) was expected to improve estimation models. The reason behind is the same as for stand density, namely the possibility of regionally different DBH-height relationships of trees, that lead to differences in the maximum possible canopy height. P_m can be easily extracted from the remote sensing data (in contrast to P_d). Here, the maximum observed canopy height H_{smax} (in the CHM) was used under the assumption that the plots are large enough to be representative for the maximum possible tree height in the respective forest types. Maximum height showed no relevance in BA estimation, but prediction errors for AGB increased from 11.8% to 14.8% if it was dropped from the model. H_{smax} represents the maximum value that TCH could possibly reach, if the whole area would be fully occupied by trees that all have reached their maximum potential height. Hence, H_{smax} might act as a standardization for TCH. Site-specific relationships between TCH and AGB (Fig. 5) show that, e.g., the forest at Traunstein reaches high AGB values at much lower mean canopy heights than other sites. By additionally providing the information that also H_{smax} at Traunstein is lower than elsewhere, the TCH values are put into the perspective of how high is the forest now and how high could it possibly become. This standardization role of H_{smax} is supported by the fact that all selected models have negative coefficients (a_m, b_m), commonly close to -1 for H_{smax}, and positive coefficients (a_h, b_h), commonly close to +1 for TCH, i.e., the ratio TCH / H_{smax} is used in the predictions. Models based on TCH_1 and H_{smax} only were the best two-predictor models in BA and AGB estimation, respectively.

4.4. The Role of Vertical Heterogeneity

Vertical heterogeneity (P_v) was after mean canopy height (P_h) the second descriptor derived at individual plot-level rather than site-level. It was included in the best BA and AGB models, however, dropping it increased the nRMSE by less than 1%. From the six candidate metrics for P_v, various were chosen in different models.

The calculation of the vertical metrics was either based on the canopy height model or on the vertical foliage profile. As visible in Fig. S3, the distributions of SD, CV and Gini index differed strongly depending on whether they were CHM- or VFP-based. CHM-based variability metrics describe the heterogeneity of the canopy surface, including ground pixels, i.e., canopy gaps. VFP-based variability metrics describe the vertical layering of the reconstructed foliage profile, which does not contain any ground component, but up-weights profile parts in the lower heights to compensate for the occlusion by high trees. Hence, their contributions to BA and AGB estimation might be different: CHM-based metrics rather characterize forests in the spectrum from smooth canopy surfaces, as observed for young, dense stands, to rough canopy surfaces, as observed for old or disturbed stands. VFP-based metrics rather account for the overseen trees in the lower canopy. Other studies have also identified the vertical heterogeneity as a component in prediction models. Magnussen et al. (2012) proposed a two-predictor model based on 1) TCH and 2) the variance of the CHM divided by TCH, which is closely related to CV_{CHM} used here. Bouvier et al. (2015) considered two vertical metrics in their four-predictor model: 1) variance of the CHM and 2) CV of the leaf area density in the VFP. In our analysis, vertical heterogeneity was also able to improve site-specific estimation models compared to solely TCH-based models (about 1% decrease in nRMSE) with different vertical metrics being chosen at different sites. To conclude, there is a wide variety of metrics that characterize vertical heterogeneity and they may in fact capture quite different aspects of forest structure. They do contribute in the improvement and generalization of BA and AGB estimation. Future analyses should try to achieve a better understanding of how the different metrics are related to ground-based metrics of forest structure, and whether a combination of several of them could further improve estimation results.

4.5. The Role of Wood Density

Regional differences in average wood density have been suspected to be a main reason behind differences in the height-to-biomass relationship of forests (Asner and Mascaro, 2014; Meyer et al., 2018; Vincent et al., 2014). In our analysis, however, dropping the wood density parameter (P_w) led only to a slight increase in nRMSE of less than 1% for AGB estimations. The values of WD_{sAGV} were very similar for BCI, SERC and Traunstein, but considerably higher for Paracou and Rabi. If region-specific estimates on average wood density are available, they should definitely be considered in AGB estimation models. Nevertheless, our results suggest that compared to other parameters wood density is of minor importance for a generalized AGB estimation. With regard to how average wood density should be calculated, Vincent et al. (2014) argued to use AGV instead of BA as a weighting variable, as AGV of the trees in the ground-truth plots has to be calculated anyway to derive tree AGB, and AGV is the structurally more appropriate weighting variable compared to BA. In this study, for all five study sites WD_{sAGV} and WD_{sBA} were found to be very similar. Thus, only WD_{sAGV} was further used in the analyses.

4.6. Generalization and Outlook

With the identified structural variables and the fitted coefficients, we propose general prediction models for BA and AGB estimation, which are applicable on temperate and tropical forests under natural and managed conditions. Having such models and also understanding the contribution of different forest structural aspects is important for consistent large scale mapping of forest carbon stocks (Lefsky et al., 2002). This is particularly relevant for upcoming spaceborne missions such as GEDI (Hancock et al., 2019; Stavros et al., 2017), ICESat-2 (Narine et al., 2019), BIOMASS (Le Toan et al., 2011) or Tandem-L (Moreira et al., 2015) which will provide consistent forest height measurements across very different forest types, not all of which are represented sufficiently in ground-truth datasets.

As a next step, the proposed relationships need to be tested at other forest sites to either confirm or, if necessary, adapt them. The results of the leave-one-site-out cross validation suggest that the presented approach for BA estimation is more robust than the one for AGB estimation, with regard to reducing the number of forest types for fitting. The distributions of the various structure metrics have shown that all of the five sites differ in one or another aspect from all others, which apparently makes each of them essential in the calibration of the AGB prediction model. In future analyses, further datasets need to be taken into account. Achieving robust leave-one-site-out cross validation results will require a sufficient degree of structural redundancy among sites, i.e., several sites representing similar forest types.

The Traunstein site was the most distinct site concerning various structural aspects. It remains unclear to which degree this can be explained by ecological differences alone (e.g., only site with large proportion of conifers), and to which degree the intensive management there plays a role. Forest management may alter some of the relationships among structure metrics, compared to natural stands. Future research should try to identify such changes. In case they are significant, management would be an additional aspect which should be considered in generalization approaches. Remote sensing methods for estimating the management regimes and parameters could then complement the biomass estimation.

Furthermore, the influence of spatial scale needs to be investigated, as different sensors produce measurements at different scales (Knapp et al., 2018; Tello et al., 2018). Finally, methods need to be developed for acquiring more of the structural variables entirely from remote sensing and becoming independent from any ground-based input. Individual tree delineation from high resolution canopy height data can be applied to derive stand density information directly from remote sensing (Duncanson et al., 2015; Ferraz et al., 2016). Average wood density can be estimated based on forest type or even species classification using passive optical remote sensing (Fassnacht et al., 2016). These technologies have to be combined to derive very detailed estimates from airborne acquisitions at landscape scale. The estimates can then serve as training areas for wall-to-wall mapping using spaceborne products.

5. Conclusion

Data from temperate and tropical forest plots was combined to develop a general equation for biomass (and basal area) estimation based on a set of forest structure metrics from lidar remote sensing. The different structural predictors were a priori defined. The results provided insight in the relative importance of mean and maximal canopy height, stand density, vertical heterogeneity and wood density for biomass estimation. Not all of those forest attributes can be derived from lidar data. For maximal stand density and mean wood density field-based information is required at the site level. Alternatively, a model without those attributes can be chosen from the list of models, at the expense of slightly lower prediction accuracies. The found relationships should provide guidance towards a standardized workflow for estimating aboveground biomass for forest carbon mapping and monitoring from remote sensing.

Acknowledgments

This study would not have been possible without the work of a large number of people collecting data in the field and conducting airborne lidar campaigns. We thank all of them for providing access to their data. In particular we thank the Smithsonian Tropical Research Institute for providing the census data for BCI. The BCI forest dynamics research project was founded by S.P. Hubbell and R.B. Foster and is now managed by R. Condit, S. Lao, and R. Perez under the Center for Tropical Forest Science and the Smithsonian Tropical Research in Panama. Numerous organizations have provided funding, principally the U.S. National Science Foundation, and hundreds of field workers have contributed. We thank J. Dalling for providing the lidar data from BCI. We thank B. Hérault, G. Derroire and A. Dourdain for providing census data and G. Vincent for providing the lidar data from Paracou. We thank K. Scipal for providing the Rabi data from the AfriSAR project and A. Alonso and H. Memiaghe for providing the census data and S. Saatchi for providing the lidar data for AfriSAR. The Rabi plot is a collaborative project of CENAREST, the Center for Conservation and Sustainability (CCS) and the Smithsonian Institution Forest Global Earth Observatory (ForestGEO). We thank G. Parker, S. McMahon and S. Davis for providing census data from SERC. These data were gathered as part of forest ecology studies at the Smithsonian Environmental Research Center. SERC is a participant in the ForestGEO network. We thank B. Cook and D. Morton for providing the lidar data for SERC, which was collected during a G-LiGT campaign. We thank H. Pretzsch, P. Biber and M. Heym for providing the census data from Traunstein. Traunstein is a participant in the ForestGEO network. We thank K. Papathanassiou, F. Kugler, M. Tello and M. Pardini for providing the lidar data from Traunstein. This study was conducted with funding by the German Federal Ministry for Economic Affairs and Energy (BMWi) under the funding reference 50EE1416. R.F., V.C. and A.H. were supported by the HGF-Helmholtz Alliance "Remote Sensing and Earth System Dynamics" HA-310 under the funding reference RA37012.

References

Anderson-Teixeira, K.J., Davies, S.J., Bennett, A.C., Gonzalez-Akre, E.B., Muller-Landau, H.C., Joseph Wright, S., Abu Salim, K., Almeyda Zambrano, A.M., Alonso, A., Baltzer, J.L., Basset, Y., Bourg, N.A., Broadbent, E.N., Brockelman, W.Y., Bunyavejchewin, S., Burslem, D.F.R.P., Butt, N., Cao, M., Cardenas, D., Chuyong, G.B., Clay, K., Cordell, S., Dattaraja, H.S., Deng, X., Detto, M., Du, X., Duque, A., Erikson, D.L., Ewango, C.E.N., Fischer, G.A., Fletcher, C., Foster, R.B., Giardina, C.P., Gilbert, G.S., Gunatilleke, N., Gunatilleke, S., Hao, Z., Hargrove, W.W., Hart, T.B., Hau, B.C.H., He, F., Hoffman, F.M., Howe, R.W., Hubbell, S.P., Inman-Narahari, F.M., Jansen, P.A., Jiang, M., Johnson, D.J., Kanzaki, M., Kassim, A.R., Kenfack, D., Kibet, S., Kinnaird, M.F., Korte, L., Kral, K., Kumar, J., Larson, A.J., Li, Y., Li, X., Liu, S., Lum, S.K.Y., Lutz, J.A., Ma, K., Maddalena, D.M., Makana, J.R., Malhi, Y., Marthews, T., Mat Serudin, R., Mcmahon, S.M., McShea, W.J., Memiaghe, H.R., Mi, X., Mizuno, T., Morecroft, M., Myers, J.A., Novotny, V., de Oliveira, A.A., Ong, P.S., Orwig, D.A., Ostertag, R., den Ouden, J., Parker, G.G., Phillips, R.P., Sack, L., Sainge, M.N., Sang, W., Sri-ngernyuang, K., Sukumar, R., Sun, I.F., Sungpalee, W., Suresh, H.S., Tan, S., Thomas, S.C., Thomas, D.W., Thompson, J., Turner, B.L., Uriarte, M., Valencia, R., Vallejo, M.I., Vicentini, A., Vr??ka, T., Wang, X., Wang, X., Weiblen, G., Wolf, A., Xu, H., Yap, S., Zimmerman, J., 2015. CTFS-ForestGEO: A worldwide network monitoring forests in an era of global change. Glob. Chang. Biol. 21, 528–549. https://doi.org/10.1111/gcb.12712

Asner, G.P., Clark, J.K., Mascaro, J., Galindo García, G. a., Chadwick, K.D., Navarrete Encinales, D. a., Paez-Acosta, G., Cabrera Montenegro, E., Kennedy-Bowdoin, T., Duque, Á., Balaji, a., von Hildebrand, P., Maatoug, L., Phillips Bernal, J.F., Knapp, D.E., García Dávila, M.C., Jacobson, J., Ordóñez, M.F., 2012. High-resolution Mapping of Forest Carbon Stocks in the Colombian Amazon. Biogeosciences Discuss. 9, 2445–2479. https://doi.org/10.5194/bgd-9-2445-2012

Asner, G.P., Mascaro, J., 2014. Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric. Remote Sens. Environ. 140, 614–624. https://doi.org/10.1016/j.rse.2013.09.023

Asner, G.P., Mascaro, J., Muller-Landau, H.C., Vieilledent, G., Vaudry, R., Rasamoelina, M., Hall, J.S., van Breugel, M., 2012. A universal airborne LiDAR approach for tropical forest carbon mapping. Oecologia 168, 1147–1160. https://doi.org/10.1007/s00442-011-2165-z

Blair, J.B., Hofton, M.A., 1999. Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. Geophys. Res. Lett. 26, 2509–2512. https://doi.org/10.1029/1999GL010484

Blanc, L., Echard, M., Herault, B., Bonal, D., Marcon, E., Chave, J., Baraloto, C., 2009. Dynamics of aboveground carbon stocks in a selectively logged tropical forest. Ecol. Appl. 19, 1397–1404. https://doi.org/10.1890/08-1572.1

Bouvier, M., Durrieu, S., Fournier, R.A., Renaud, J.P., 2015. Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data. Remote Sens. Environ. 156, 322–334. https://doi.org/10.1016/j.rse.2014.10.004

Chave, J., Coomes, D., Jansen, S., Lewis, S.L., Swenson, N.G., Zanne, A.E., 2009. Towards a worldwide wood economics spectrum. Ecol. Lett. 12, 351–366. https://doi.org/10.1111/j.1461-0248.2009.01285.x

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B.C., Duque, A., Eid, T., Fearnside, P.M., Goodman, R.C., Henry, M., Martínez-Yrízar, A., Mugasha, W.A., Muller-Landau, H.C., Mencuccini, M., Nelson, B.W., Ngomanda, A., Nogueira, E.M., Ortiz-Malavassi, E., Pélissier, R., Ploton, P., Ryan, C.M., Saldarriaga, J.G., Vieilledent, G., 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. Glob. Chang. Biol. 20, 3177–3190. https://doi.org/10.1111/gcb.12629

Condit, R., 1998. Tropical forest census plots. Springer-Verlag and R. G. Landes Company, Berlin, Germany and George Town, Texas. https://doi.org/10.1007/978-3-662-03664-8

Condit, R., Lao, S., Pérez, R., Dolins, S.B., Foster, R.B., Hubbell, S.P., 2012. Barro Colorado Forest Census Plot Data. Cent. Trop. For. Sci. Databases. https://doi.org/10.5479/data.bci.20130603

Cook, B.D., Corp, L. a., Nelson, R.F., Middleton, E.M., Morton, D.C., McCorkel, J.T., Masek, J.G., Ranson, K.J., Ly, V., Montesano, P.M., 2013. NASA goddard's LiDAR, hyperspectral and thermal (G-LiHT) airborne imager. Remote Sens. 5, 4045–4066. https://doi.org/10.3390/rs5084045

Coomes, D.A., Dalponte, M., Jucker, T., Asner, G.P., Banin, L.F., Burslem, D.F.R.P., Lewis, S.L., Nilus, R., Phillips, O.L., Phua, M.H., Qie, L., 2017. Area-based vs tree-centric approaches to mapping forest carbon in Southeast Asian forests from airborne laser scanning data. Remote Sens. Environ. 194, 77–88. https://doi.org/10.1016/j.rse.2017.03.017

Duncanson, L., Huang, W., Johnson, K., Swatantran, A., McRoberts, R.E., Dubayah, R., 2017. Implications of allometric model selection for county-level biomass mapping. Carbon Balance Manag. 12. https://doi.org/10.1186/s13021-017-0086-9

Duncanson, L.I., Dubayah, R.O., Cook, B.D., Rosette, J., Parker, G., 2015. The importance of spatial detail: Assessing the utility of individual crown information and scaling approaches for lidar-based biomass density estimation. Remote Sens. Environ. 168, 102–112. https://doi.org/10.1016/j.rse.2015.06.021

Fassnacht, F.E., Hartig, F., Latifi, H., Berger, C., Hernández, J., Corvalán, P., Koch, B., 2014. Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. Remote Sens. Environ. 154, 102–114. https://doi.org/10.1016/j.rse.2014.07.028

Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. Remote Sens. Environ. 186, 64–87. https://doi.org/10.1016/j.rse.2016.08.013

Ferraz, A., Saatchi, S., Mallet, C., Jacquemoud, S., Gonï¿½alves, G., Silva, C.A., Soares, P., Tomï¿½, M., Pereira, L., 2016. Airborne lidar estimation of aboveground forest biomass in the absence of field inventory. Remote Sens. 8, 1–18. https://doi.org/10.3390/rs8080653

Fischer, R., Knapp, N., Bohn, F., Shugart, H.H., Huth, A., 2019. The Relevance of Forest Structure for Biomass and Productivity in Temperate Forests: New Perspectives for Remote Sensing. Surv. Geophys. https://doi.org/10.1007/s10712-019-09519-x

Getzin, S., Fischer, R., Knapp, N., Huth, A., 2017. Using airborne LiDAR to assess spatial heterogeneity in forest structure on Mount Kilimanjaro. Landsc. Ecol. 32, 1881–1894. https://doi.org/10.1007/s10980-017-0550-7

Goetz, S., Dubayah, R., 2011. Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. Carbon Manag. 2, 231–244. https://doi.org/10.4155/cmt.11.18

Hancock, S., Armston, J., Hofton, M., Sun, X., Tang, H., Duncanson, L.I., Kellner, J.R., Dubayah, R., 2019. The GEDI Simulator: A Large-Footprint Waveform Lidar Simulator for Calibration and Validation of Spaceborne Missions. Earth Sp. Sci. 6, 294–310. https://doi.org/10.1029/2018EA000506

Harding, D.J., Lefsky, M.A., Parker, G.G., Blair, J.B., 2001. Laser altimeter canopy height profiles methods and validation for closed-canopy, broadleaf forests. Remote Sens. Environ. 76, 283–297. https://doi.org/10.1016/S0034-4257(00)00210-8

Harris, N.L., Brown, S., Hagen, S.C., Saatchi, S.S., Petrova, S., Salas, W., Hansen, M.C., Potapov, P. V., Lotsch, A., 2012. Baseline map of carbon emissions from deforestation in tropical regions. Science (80-.). 336, 1573–1576. https://doi.org/10.1126/science.1217962

Hérault, B., Piponiot, C., 2018. Key drivers of ecosystem recovery after disturbance in a neotropical forest. For. Ecosyst. 5, 2. https://doi.org/10.1186/s40663-017-0126-7

Hubbell, S., Foster, R., O'Brien, S., Harms, K., Condit, R., Wechsler, B., Wright, S., Loo de Lao, S., 1999. Light gap disturbances recruitment llimitation, and tree diversity in a neotropical forest. Science

(80-.). 283, 554–557. https://doi.org/10.1126/science.283.5401.554
Isenburg, M., 2011. LAStools - efficient tools for LiDAR processing.
Jucker, T., Caspersen, J., Chave, J., Antin, C., Barbier, N., Bongers, F., Dalponte, M., van Ewijk, K.Y., Forrester, D.I., Haeni, M., Higgins, S.I., Holdaway, R.J., Iida, Y., Lorimer, C., Marshall, P.L., Momo, S., Moncrieff, G.R., Ploton, P., Poorter, L., Rahman, K.A., Schlund, M., Sonké, B., Sterck, F.J., Trugman, A.T., Usoltsev, V.A., Vanderwel, M.C., Waldner, P., Wedeux, B.M.M., Wirth, C., Wöll, H., Woods, M., Xiang, W., Zimmermann, N.E., Coomes, D.A., 2017. Allometric equations for integrating remote sensing imagery into forest monitoring programmes. Glob. Chang. Biol. 23, 177–190. https://doi.org/10.1111/gcb.13388
Knapp, N., Fischer, R., Huth, A., 2018. Linking lidar and forest modeling to assess biomass estimation across scales and disturbance states. Remote Sens. Environ. 205, 199–209. https://doi.org/10.1016/j.rse.2017.11.018
Král, K., Shue, J., Vrška, T., Gonzalez-Akre, E.B., Parker, G.G., McShea, W.J., McMahon, S.M., 2016. Fine-scale patch mosaic of developmental stages in Northeast American secondary temperate forests: the European perspective. Eur. J. For. Res. 135, 981–996. https://doi.org/10.1007/s10342-016-0988-1
Labrière, N., Tao, S., Chave, J., Scipal, K., Toan, T. Le, Abernethy, K., Alonso, A., Barbier, N., Bissiengou, P., Casal, T., Davies, S.J., Ferraz, A., Hérault, B., Jaouen, G., Jeffery, K.J., Kenfack, D., Korte, L., Lewis, S.L., Malhi, Y., Memiaghe, H.R., Poulsen, J.R., Réjou-Méchain, M., Villard, L., Vincent, G., White, L.J.T., Saatchi, S., 2018. In Situ Reference Datasets From the TropiSAR and AfriSAR Campaigns in Support of Upcoming Spaceborne Biomass Missions. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 1–11. https://doi.org/10.1109/JSTARS.2018.2851606
Le Toan, T., Quegan, S., Davidson, M.W.J., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S., Shugart, H., Ulander, L., 2011. The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. Remote Sens. Environ. 115, 2850–2860. https://doi.org/10.1016/j.rse.2011.03.020
Lefsky, M.A., Cohen, W.B., Harding, D.J., Parker, G.G., Acker, S.A., Gower, S.T., 2002. Lidar remote sensing of above-ground biomass in three biomes. Glob. Ecol. Biogeogr. 11, 393–399. https://doi.org/10.1046/j.1466-822x.2002.00303.x
Lobo, E., Dalling, J.W., 2014. Spatial scale and sampling resolution affect measures of gap disturbance in a lowland tropical forest: Implications for understanding forest regeneration and carbon storage. Proc. R. Soc. B Biol. Sci. 281, 1–8. https://doi.org/10.1098/rspb.2013.3218
Lu, D., Chen, Q., Wang, G., Liu, L., Li, G., Moran, E., 2014. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. Int. J. Digit. Earth 1–43. https://doi.org/10.1080/17538947.2014.990526
Magnussen, S., Næsset, E., Gobakken, T., Frazer, G., 2012. A fine-scale model for area-based predictions of tree-size-related attributes derived from LiDAR canopy heights. Scand. J. For. Res. 27, 312–322. https://doi.org/10.1080/02827581.2011.624116
McMahon, S.M., Parker, G.G., 2015. A general model of intra-annual tree growth using dendrometer bands. Ecol. Evol. 5, 243–254. https://doi.org/10.1002/ece3.1117
Meyer, V., Saatchi, S., Clark, D.B., Keller, M., Vincent, G., Ferraz, A., Espírito-Santo, F., D'Oliveira, M.V.N., Kaki, D., Chave, J., 2018. Canopy area of large trees explains aboveground biomass variations across neotropical forest landscapes. Biogeosciences 15, 3377–3390. https://doi.org/10.5194/bg-15-3377-2018
Moreira, A., Krieger, G., Hajnsek, I., Papathanassiou, K., Younis, M., Lopez-Dekker, F., Huber, S., Villano, M., Pardini, M., 2015. Tandem-L: A highly innovative bistatic SAR mission for global observation of dynamic processes on the earth's surface. IEEE Geosci. Remote Sens. Mag. June, 8–23. https://doi.org/10.1109/MGRS.2015.2437353
Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical

two-stage procedure and field data. Remote Sens. Environ. 80, 88–99. https://doi.org/10.1016/S0034-4257(01)00290-5

Narine, L.L., Popescu, S., Neuenschwander, A., Zhou, T., Srinivasan, S., Harbeck, K., 2019. Estimating aboveground biomass and forest canopy cover with simulated ICESat-2 data. Remote Sens. Environ. 224, 1–11. https://doi.org/10.1016/j.rse.2019.01.037

Ni-Meister, W., Jupp, D.L.B., Dubayah, R., 2001. Modeling lidar waveforms in heterogeneous and discrete canopies. IEEE Trans. Geosci. Remote Sens. 39, 1943–1958. https://doi.org/10.1109/36.951085

R Development Core Team, 2014. R: A Language and Environment for Statistical Computing.

Reineke, L.H., 1933. Perfecting a stand-density index for even-aged forests. J. Agric. Res. 46, 627–638.

Stavros, E.N., Schimel, D., Pavlick, R., Serbin, S., Swann, A., Duncanson, L., Fisher, J.B., Fassnacht, F., Ustin, S., Dubayah, R., Schweiger, A., Wennberg, P., 2017. ISS observations offer insights into plant function. Nat. Ecol. Evol. 1, 1–4. https://doi.org/10.1038/s41559-017-0194

Tang, H., Dubayah, R., Swatantran, A., Hofton, M., Sheldon, S., Clark, D.B., Blair, B., 2012. Retrieval of vertical LAI profiles over tropical rain forests using waveform lidar at La Selva, Costa Rica. Remote Sens. Environ. 124, 242–250. https://doi.org/10.1016/j.rse.2012.05.005

Tello, M., Cazcarra-Bes, V., Fischer, R., Papathanassiou, K., 2018. Multiscale forest structure estimation from SAR tomography, in: Proceedings of the European Conference on Synthetic Aperture Radar, EUSAR. VDE, Aachen, Germany, pp. 600–603.

Treuhaft, R., Goncalves, F., Dos Santos, J.R., Keller, M., Palace, M., Madsen, S.N., Sullivan, F., Graca, P.M.L.A., 2015. Tropical-forest biomass estimation at X-band from the spaceborne TanDEM-X interferometer. IEEE Geosci. Remote Sens. Lett. 12, 239–243. https://doi.org/10.1109/LGRS.2014.2334140

Vincent, G., Sabatier, D., Blanc, L., Chave, J., Weissenbacher, E., Pélissier, R., Fonty, E., Molino, J.F., Couteron, P., 2012. Accuracy of small footprint airborne LiDAR in its predictions of tropical moist forest stand structure. Remote Sens. Environ. 125, 23–33. https://doi.org/10.1016/j.rse.2012.06.019

Vincent, G., Sabatier, D., Rutishauser, E., 2014. Revisiting a universal airborne light detection and ranging approach for tropical forest carbon mapping: Scaling-up from tree to stand to landscape. Oecologia 175, 439–443. https://doi.org/10.1007/s00442-014-2913-y

Zanne, A.E., Lopez-Gonzalez, G., Coomes, D.A., Ilic, J., Jansen, S., Lewis, S.L., Miller, R.B., Swenson, N.G., Wiemann, M.C., Chave, J., 2009. Data from: Towards a worldwide wood economics spectrum. Ecol. Lett. https://doi.org/10.5061/dryad.234

Zolkos, S.G., Goetz, S.J., Dubayah, R., 2013. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. Remote Sens. Environ. 128, 289–298. https://doi.org/10.1016/j.rse.2012.10.017