

Mapping the density of giant trees in the Amazon

Robson Borges de Lima¹ , Diego Armando Silva da Silva² , Matheus Henrique Nunes³ , Paulo R. de Lima Bittencourt⁴ , Peter Groenendyk⁵ , Cinthia Pereira de Oliveira¹ , Daniela Granato-Souza⁶ , Rinaldo L. Caraciolo Ferreira⁷ , José A. Aleixo da Silva⁷ , Jesús Aguirre-Gutiérrez⁸ , Toby Jackson⁹ , João R. de Matos Filho¹⁰ , Perseu da Silva Aparício¹ , Joselane P. Gomes da Silva¹ , José Julio de Toledo¹¹ , Marcelino Carneiro Guedes¹² , Danilo R. Alves de Almeida¹³ , Niro Higuchi¹⁴ , Fabien H. Wagner¹⁵ , Jean Pierre Ometto¹⁶  and Eric Bastos Görgens¹⁷ 

¹Laboratório de Manejo Florestal, Universidade do Estado do Amapá, 68901-262, Macapá, AP, Brazil; ²Instituto Federal de Educação Ciência e Tecnologia do Amapá, 68.920-000, Laranjal do Jari, AP, Brazil; ³Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA; ⁴School of Earth and Environment Sciences, Cardiff University, CF10 3AT, Cardiff, UK; ⁵Department of Plant Biology, Institute of Biology, University of Campinas, PO Box 6109, 13083-970, Campinas, SP, Brazil; ⁶Department of Natural Resources and Environmental Sciences College of Agricultural, Alabama A & M University, Huntsville 35762 AL, USA; ⁷Departamento de Ciência Florestal, Universidade Federal Rural de Pernambuco, 52171-900, Recife, PE, Brazil; ⁸Environmental Change Institute, University of Oxford, Oxford, OX13QY, UK; ⁹School of Biological Sciences, University of Bristol, Bristol, BS8 1QU, UK; ¹⁰Promotória de Justiça do Meio Ambiente, Conflitos Agrários, Ministério Público, 68903-883, Macapá, AP, Brazil; ¹¹Universidade Federal do Amapá, 68903-419, Macapá, AP, Brazil; ¹²Empresa Brasileira de Pesquisa Agropecuária, 68903-419, Macapá, AP, Brazil; ¹³Department of Forest Sciences, 'Luiz de Queiroz' College of Agriculture, University of São Paulo, 13418-900, Piracicaba, SP, Brazil; ¹⁴Instituto Nacional de Pesquisas da Amazônia, Coordenação de Pesquisas em Silvicultura Tropical, 69080-971, Manaus, AM, Brazil; ¹⁵Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove, Pasadena, CA 91109, USA; ¹⁶Instituto Nacional de Pesquisas Espaciais, 12227-010, São José dos Campos, SP, Brazil; ¹⁷Departamento de Engenharia Florestal, Universidade Federal dos Vales do Jequitinhonha e Mucuri, 39100-000, Diamantina, MG, Brazil

Summary

Author for correspondence:
Robson Borges de Lima
Email: rbl_florestal@yahoo.com.br,
robson.lima@ueap.edu.br

Received: 5 March 2025
Accepted: 22 September 2025

New Phytologist (2025)
doi: 10.1111/nph.70634

Key words: airborne LiDAR, biogeographic provinces, environmental factors, spatial modeling, tallest trees.

- Tall trees (height ≥ 60 m) are keystone elements of tropical forests, strongly influencing biodiversity, carbon storage, and ecosystem resilience. Yet, their density and spatial distribution remain poorly quantified, especially in remote Amazonian regions, limiting our understanding of their ecological roles and contribution to forest–climate interactions.
- We combined airborne LiDAR data from 900 transects across the Brazilian Amazon with environmental predictors to model tall-tree density. Spatial extrapolations allowed us to generate regional distribution estimates and assess associations with climate, topography, and disturbance regimes.
- Our model predicts that tall trees are unevenly distributed, with c. 14% of the estimated density concentrated in c. 1% of the Amazon and c. 50% within c. 11%. The highest densities occur in Roraima and the Guiana Shield provinces, where water availability is high and lightning or storm incidence is low. Modeled density strongly correlates with aboveground biomass, highlighting the disproportionate contribution of tall trees to carbon stocks. We estimate c. 55.5 million tall trees across the Brazilian Amazon.
- These findings demonstrate that tall-tree distribution is a crucial but underused predictor for biomass models. Understanding their ecological and spatial dynamics is vital for forest conservation and climate-resilience strategies under increasing anthropogenic pressures.

Introduction

Understanding the density and distribution of larger trees in the Amazon is vital for predicting the carbon balance of Amazonian ecosystems with global environmental change (Bastin *et al.*, 2018; Lutz *et al.*, 2018; Enquist *et al.*, 2020). Particularly, the local and regional variations in the density of large trees are strongly linked to spatial variations in the aboveground biomass (AGB) of tropical forests (Slik *et al.*, 2013; Birdsey *et al.*, 2023)

and regulate the microclimate, water availability, light intensity, and understory species diversity (Lindenmayer, 2017; Lindenmayer & Laurance, 2017; Brando, 2018; Pinho *et al.*, 2020; Draper *et al.*, 2021). Around 390 billion trees are estimated to inhabit the Amazon (Crowther *et al.*, 2015). However, large uncertainties remain regarding how many tall trees reach the uppermost stratum of Amazonian forest canopies and how their survival is locally and regionally distributed. These large trees take centuries to reach such sizes, and many of the species may be

unable to regenerate on timescales compatible with the planet's rapidly changing climate (Larjavaara, 2014; Bennett *et al.*, 2015; Ali & Wang, 2021). Although there is still limited empirical evidence quantifying the recovery dynamics of giant trees across tropical forests, their exceptional longevity – combined with increasing forest degradation and intensifying climate-related disturbances (e.g. drought, wind, lightning) – suggests that they are being lost at a faster rate than they are being replaced (Trenberth *et al.*, 2014; McDowell *et al.*, 2018; Gora *et al.*, 2025). This imbalance makes them particularly vulnerable under current environmental pressures (Piovesan & Biondi, 2021). Monitoring the density of giant trees is therefore essential to understanding forest ecosystem stability, resilience, and long-term carbon storage potential.

The density of tall trees is largely influenced by variables that determine their growth and survival rates, encompassing both genetic factors and environmental conditions, such as climate, soil, and topography, which shape their structural characteristics (Jucker *et al.*, 2018; Caron *et al.*, 2021; Mills *et al.*, 2023; Ter Steege *et al.*, 2023). Previous studies, such as Slik *et al.* (2013), have demonstrated that large-diameter trees – often strongly correlated with extreme height – play a dominant role in explaining AGB variation across tropical forests. Moderate temperatures and increased light availability (i.e. providing more energy for growth), alongside low water stress, are factors that can favor a large density of tall trees (Stephenson, 1990; Madrigal-González *et al.*, 2023). However, the future survival of large trees may be threatened by climate change and land-use changes. For example, large climatic and atmospheric oscillations may lead to physiological and mechanical instability in the structure of tall trees (Gora & Esquivel-Muelbert, 2021; Jackson *et al.*, 2021). The efficient vascular systems of tall trees are particularly vulnerable to prolonged droughts, which can lead to the collapse of water and nutrient transport (Barros *et al.*, 2019; Araújo *et al.*, 2024), and high turbulence caused by strong winds and a high incidence of lightning kills tall trees disproportionately (Gora *et al.*, 2021; Feng *et al.*, 2023). Therefore, the ability of Amazonian forests to support tall trees likely depends on a balance between environmental factors associated with resource availability that promote tree growth to its potential maximum size, alongside mechanisms that promote survival under vegetation disturbances (Gorgens *et al.*, 2021). Analyzing the density of tall trees as a function of environmental determinants is crucial to understanding the Amazon's stability under increasing environmental pressures and climate change.

Field inventories have traditionally played a crucial role in mapping forest patterns and assessing tree diversity in the Amazon, offering essential insights into species composition and distribution (ter Steege *et al.*, 2013, 2023; ForestPlots.net *et al.*, 2021; De Lima *et al.*, 2023). However, conventional approaches based on forest inventory plots provide limited spatial coverage for evaluating the density of tall trees. This limitation is largely due to the logistical and technical challenges of detecting these rare and spatially scattered individuals, which often require large sampling areas in remote and difficult-to-access regions (Harris *et al.*, 2021; Carvalho *et al.*, 2023). Although

field inventories remain fundamental for understanding tree-environment interactions and ecological mechanisms, they frequently fall short in capturing the broad-scale spatial variability and true density of tall trees across the Amazon.

The advent of Light Detection and Ranging (LiDAR) has transformed forest assessments by providing high-resolution, three-dimensional data on forest structure over large spatial scales. This technology can be used for rapid and accurate estimates of canopy height and AGB in remote regions (Asner *et al.*, 2010; Dubayah *et al.*, 2010; Coomes *et al.*, 2017). The 'Biomass Estimation in the Amazon' (EBA) program, for example, mapped 900 transects, covering 375 ha each, which represents a remarkable increase of > 100-fold in sampling capacity compared to traditional forest inventory plot methods (Fig. 1). This technological breakthrough has, for the first time in the Amazon, enabled more precise and accurate mapping of biomass stocks (Ometto *et al.*, 2023) and reshaped our understanding of forest structure and the distribution of giant trees in the Amazon (Gorgens *et al.*, 2019, 2021). In particular, Gorgens *et al.* (2019) identified the tallest tree ever recorded in the Amazon – a *Dinizia excelsa* reaching 88.5 m height in the eastern portion of the Roraima biogeographical province in Brazil. This discovery underscored the ecological importance of these towering giants and highlighted the critical role remote sensing techniques have in uncovering previously unknown aspects of tropical forest ecosystems. Despite these advances, there is still a dearth of LiDAR-based estimates of large tree density across the tropics. Better estimates of large-tree density are crucial to improving our understanding of the ecological processes and mechanisms driving tree-environment interactions, which are vital for predicting the carbon cycle dynamics and species vulnerability of Amazonian forests.

Here, we generated the first map of the spatial distribution of the giant tree density (height ≥ 60 m) in the Amazon basin, integrating spatially explicit climate, topography, atmosphere, and soil information. We fit a Random Forest (RF) machine learning model to predict the density of giant trees across the Brazilian Amazon, based on environmental and climatic variables. Then, we assessed the underlying environmental factors and identified the principal components that explain most of the variance in tall-tree density in these biogeographic provinces. We also analyze the broader implications of these results for biomass estimates. These comprehensive analyses provide new insights into how different environmental variables influence tall-tree density and provide crucial information for forest conservation and management policies in the Amazon.

Materials and Methods

LiDAR data collection and standardization

The point cloud from airborne LiDAR was acquired between 2016 and 2018 for 900 transects distributed throughout the Brazilian Amazon biome (Fig. 1, Supporting Information Fig. S1a, b). The data collection was part of the EBA project developed by National Institute for Space Research (Gorgens *et al.*, 2021;

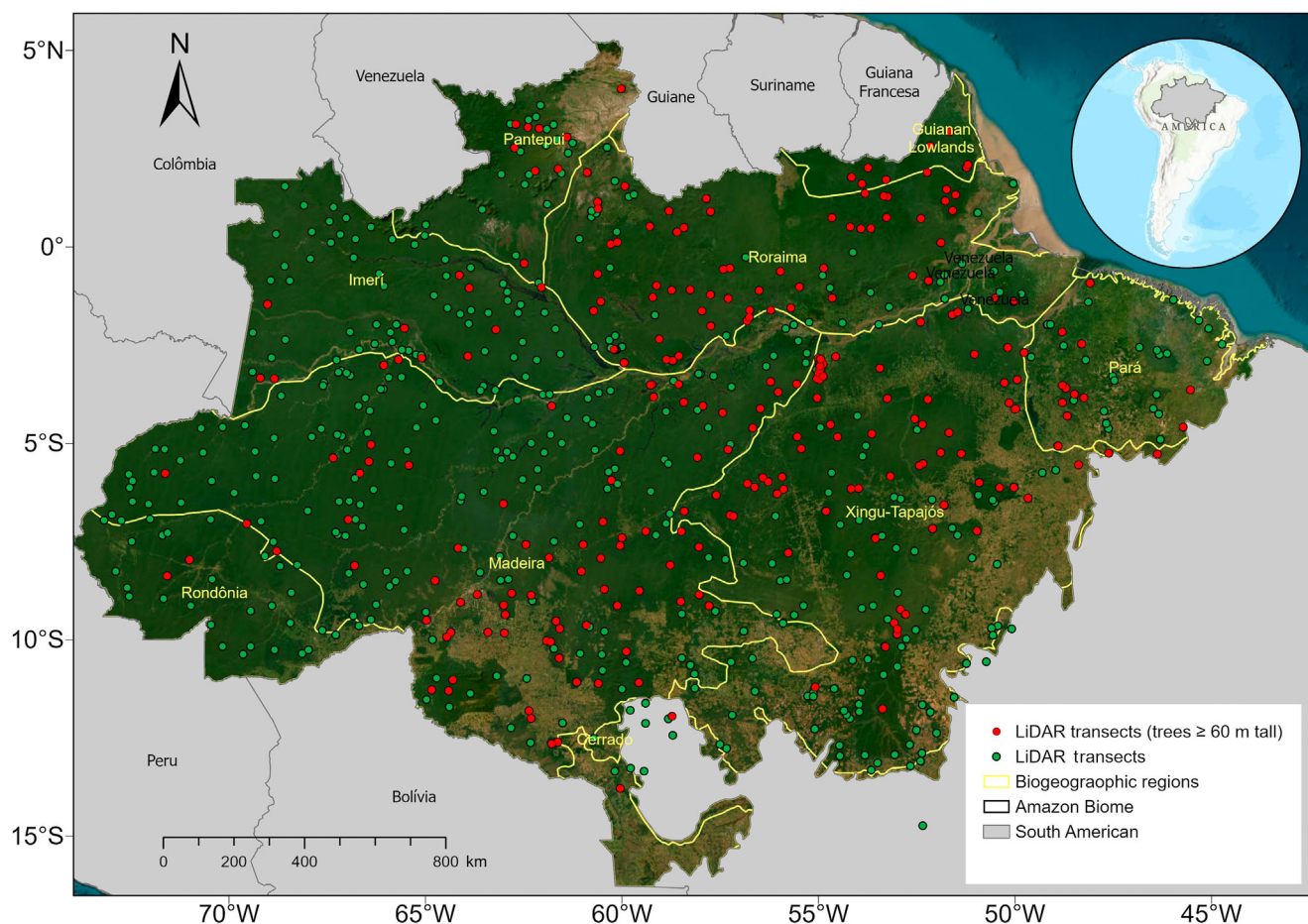


Fig. 1 Distribution of transects mapped by airborne Light Detection and Ranging (LiDAR) across the Amazon Basin. The figure shows 900 randomly distributed transects covering the eight biogeographic provinces in the region. Red points represent transects where trees taller than 60 m were identified, whereas green points indicate transects without trees ≥ 60 m. Yellow lines delimit major biogeographic regions, including Imeri, Roraima, Guiana Shield, Pantepui, Madeira, Rondônia, Xingu-Tapajós, and Pará. The LiDAR data transects are available at <https://zenodo.org/records/4091222#.YDe7l2hKjIV>.

Ometto *et al.*, 2023). The sensor used was the LiDAR HARRIER 68i, coupled to a CESSNA model 206 aircraft. The scanning angle was 45° , and the flight altitude was *c.* 600 m. The point cloud is formed by 4 returns m^{-2} . Each transect mapped 3.75 km^2 (representing $12.5 \times 0.3 \text{ km}$ strips). See more details at <https://zenodo.org/records/4091222#.YDe7l2hKjIV>.

For each LiDAR transect, tall trees were identified using a local maxima detection algorithm applied to a canopy height model (CHM) derived from a normalized digital surface model and a digital terrain model (Silva *et al.*, 2022). The CHM was generated at 1-m resolution, and local maxima were identified using a 100-m circular moving window, which helps to suppress noise and avoid the over-detection of treetops in dense canopy areas. Each treetop was assigned a geographic coordinate and its maximum height extracted from the CHM. To facilitate analysis, detected treetops were aggregated by transect and categorized into three height strata: < 40 , ≥ 40 and < 60 , and ≥ 60 m, with the latter defined as ‘giant trees’ for the purposes of this study (Fig. S1b). A unique string for each transect assigned as ID (transect number); the transect central coordinates (latitude

(LAT) and longitude (LON)) and total area (in hectares) formed the initial dataframe for subsequent analyses (Dataset S1). To ensure consistency and maximum accuracy in the density data per transect and the final density maps, we standardized the area size to square kilometers (km^2). Considering only tall trees (height ≥ 60 m), we detected their presence in 313 out of the 900 transects (*c.* 35%) across the Amazon biome. However, all 900 transects, including those with zero density of tall trees ($n = 587$), were used in the spatial modeling (RF) and principal component analysis (PCA). This approach ensured a comprehensive and representative analysis of the full gradient of environmental conditions and tall tree occurrence across the biome.

All transects in the final tree density data matrix cover eight biogeographic provinces proposed by Morrone (2014). This biogeographic definition seeks a universal classification in provinces with similar macroecological characteristics of biodiversity. To understand the optimal environmental conditions for the occurrence of the tallest trees, the density data km^{-2} were linked to spatially explicit environmental factors. The wide distribution of our sampling points (in number and distribution of LiDAR

transects) provides a substantially representative sampling effort for this vegetation stratum. It ensures that any uncertainty in transect locations or minor changes in forest area in the Amazon region are unlikely to alter mean values or estimates of giant's tree density.

Environmental factors

Our research initially involved 16 spatially explicit environmental predictor variables representing topography, climate, and soil, as detailed in Table 1 and Fig. S2. The data were then carefully cropped to fit the geographic boundaries of the Brazilian Amazon biome and, when necessary, resampled to a spatial resolution of 30 arc seconds (*c.* 1 km).

Information on temperature and precipitation derived from 19 bioclimatic variables was obtained from WORLDCLIM v.2 (Fick & Hijmans, 2017). Precipitation seasonality (BIO15) represents the coefficient of variation of monthly precipitation values and is used as an indicator of the intensity of the dry season. Temperature seasonality (BIO4) reflects the SD of monthly mean temperatures ($\times 100$). Higher values for both variables indicate greater climatic variability across the year. The average number of cloudless days throughout the year was obtained using surface reflectance products from the MODIS (Moderate Resolution Imaging Spectroradiometer) sensor. We used the Terra MOD09GA v.6 product, which estimates the MODIS surface spectral reflectance duly corrected for atmospheric conditions.

The annual average number of days with precipitation > 20 mm was calculated from the precipitation time series of the

Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) dataset (Funk *et al.*, 2015). Potential evapotranspiration was derived from data provided by TerraClimate, which combines WorldClim climatological normals, Climatic Research Unit (CRU) Ts4.0, the 55-yr Japanese Reanalysis (JRA-55) data, and the Penman-Monteith methodology. The fraction of absorbed photosynthetically active radiation (FAPAR) was obtained from the calibrated and corrected land surface reflectance product of the Advanced Very High-Resolution Radiometer (AVHRR) of the National Oceanic and Atmospheric Administration (NOAA), providing information on the photosynthetic activity of plants (Baret *et al.*, 2013).

Lightning frequency, associated with weather events and tree mortality (Gora *et al.*, 2020), was obtained from the Lightning Imaging Sensor (LIS) instrument aboard the Tropical Rainfall Measurement Mission, provided by NASA's Earth Observing System's Global Hydrological Resources Center (EOSDIS). Lightning activity was represented by average annual flash rate (flashes yr^{-1}), derived from LIS/TRMM data at *c.* 0.1° spatial resolution. The variable expresses the total number of lightning flashes per year per grid cell and is not normalized by area, as each value is associated with a fixed spatial extent defined by the sensor resolution. To represent long-term wind disturbance patterns, we used 5-yr (2014–2018) averages of the daily maximum wind speed components (*zonal-uspeed* and *meridional-uspeed*), calculated from hourly wind data derived from the ERA-Interim reanalysis (ECM-RWF). For each grid cell, the maximum wind speed for each day was identified from hourly records, and these daily maxima were then averaged across the 5-yr period to

Table 1 List of the main environmental variables selected for this study.

Environmental variable					
Category	Sub-category	Name and unit	Abbreviation	Spatial resolution (period)	Source
Topographic Climatic	Elevation	Elevation above sea level (m)	elev	30 m	SRTM
		Mean annual temperature (°C)	tannual	30 arc seconds	WorldClim
	Temperature	Maximum temperature (°C)	tmax	30 arc seconds	WorldClim
		Temperature seasonality (%)	tseason	30 arc seconds	WorldClim
		Average annual precipitation (mm)	pannual	30 arc seconds	WorldClim
	Precipitation	Precipitation seasonality (%)	pseason	30 arc seconds	WorldClim
		Precipitation of the wettest month	pwettest	30 arc seconds	WorldClim
		Number of clear days per year (days)	clearDays	500 m (2014–2018)	MODIS
	Physiologic influence	Days with precipitation > 20 mm (days)	days20	0.05° (2014–2018)	CHIRPS
		Potential evapotranspiration (mm yr^{-1})	pet	2.5 arc minutes (1990–2016)	TerraClimate
Edaphic	Stressors	Fraction of absorbed photosynthetically active radiation (%)	fapar	0.05° (2016–2018)	NOAA AVHRR
		Lightning rate (flashes rate yr^{-1})	lightning	0.1°	LIS TRMM
		Meridional speed (N–S) (m s^{-1})	vspeed	0.25° (2014–2018)	ECM-RWF
		Zonal speed (W–E) (m s^{-1})	uspeed	0.25° (2014–2018)	ECM-RWF
		Fraction of clay content	clayContent	250 m	SoilGrids
	Fraction of water content	Soil structure physical properties water availability (%)	waterContent	250 m	SoilGrids
		Soil structure physical properties water availability (%)			
	Fraction of clay content	Soil structure physical properties water availability (%)			
		Soil structure physical properties water availability (%)			
		Soil structure physical properties water availability (%)			

Variable categories, subcategories, names and their corresponding units and abbreviations are shown. AVHRR, Advanced Very High-Resolution Radiometer; CHIRPS, Climate Hazards Group InfraRed Precipitation with Station; LIS TRMM, Lightning Imaging Sensor instrument aboard the Tropical Rainfall Measurement Mission; MODIS, Moderate Resolution Imaging Spectroradiometer; NOAA, National Oceanic and Atmospheric Administration; SRTM, Shuttle Radar Topography Mission; WorldClim, global climate data.

produce a robust indicator of prevailing peak wind conditions. Previous studies indicate that winds are correlated with disturbances that result in tree mortality in the Amazon (Marra *et al.*, 2014; Rifai *et al.*, 2016).

Soil variables were obtained from the SoilGrids database (Hengl *et al.*, 2017), which applies machine learning to a global compilation of soil profiles. From this dataset, we selected two variables relevant to forest structure: clay content (% of fine particles < 2 µm) and soil water content (% volume at field capacity at 30 cm depth). Both layers were used at a spatial resolution of 250 m.

All geospatial data were processed in ArcMap 10.1 software. We extracted all geospatial covariate values from raster datasets for transect location points using the *raster::extract* function from the R RASTER package (Hijmans *et al.*, 2025) to construct a standardized plot-level dataframe.

Principal component analyses

To explore associations between environmental factors and understand the patterns governing giant tree density, we applied PCA to the normalized dataset of predictors. PCA was conducted separately for the entire Amazon biome and for each biogeographic province, allowing us to identify the main environmental gradients at both regional and local scales. The input variables were standardized to zero mean and unit variance before analysis to account for differences in measurement scales. Biplots of the first two principal components were generated using the FACTOEXTRA R package (Kassambara & Mundt, 2020), providing a visual representation of how environmental variables covary and how these patterns may influence tall tree density in different parts of the biome.

Spatial modeling

A RF model was used to model the density of tall trees from the environmental variables (Cutler & Wiener, 2022). This machine learning method detects global trends present in data using an ensemble strategy of decision trees to predict tree density in the uppermost strata of the Amazon rainforest canopy using the 16 environmental covariates. The RF algorithm applies the general bootstrap aggregation technique (bagging) with a modified tree learning algorithm that selects a random subset of the features at each candidate split in the learning process (Liang *et al.*, 2022). Since a random subset of variables is chosen for each tree, the RF algorithm based on trained tree ensembles avoids overfitting (Calhoun *et al.*, 2020). It circumvents potential multicollinearity issues (Genauer & Poggi, 2020) between the predictor variables.

For rigorous evaluation of the RF model, we employed the *k*-fold cross-validation method and spatial cross-validation. First, the 900 sampled transects were randomly divided into *k* subsets (or *folds*) of approximately equal size. The randomized cross-validation was of the *k*-fold type, randomly dividing into *k* groups. In this procedure, *k* is defined as 15. For each *k* subset, the RF model is trained using *k*–1 subsets as the training set and the remaining subset as the test (or validation) set. This process

was repeated 100 times with sample replacement to examine the accuracy of the estimated tree density values. After each round of training and validation, we calculated model performance metrics, such as root mean square error (RMSE) and coefficient of determination (R^2). RMSE is a measure of accuracy that reflects the average magnitude of prediction errors, while R^2 represents the proportion of variance explained by the model. Both provide a comprehensive assessment of model accuracy. These metrics are stored for each fold. At the end of the process, the metrics obtained in each *k* validation round are aggregated, usually by the average, to provide an overall estimate of the model performance. This average provides a more stable and reliable measure of model accuracy than a simple single split between training and testing.

A spatial cross-validation methodology was employed to assess the predictive ability and spatial uncertainty of an RF model in estimating the density of tall trees (height ≥ 60 m) based on environmental variables. The spatial cross-validation accounts for the spatial autocorrelation among sampling points, thereby minimizing the overestimation bias of model accuracy metrics due to the spatial proximity between training and test data. Initially, the data were organized in a spatial structure using the *sf* library (Pebesma *et al.*, 2024), allowing for the manipulation and visualization of sample points with geographic coordinates. Ten folds were used for spatial cross-validation, each representing a unique data partition. This division was conducted randomly and in a balanced manner, ensuring that each sample point was used exactly once as a test set. By contrast, the others were used for training, thus ensuring the generalization of the results and control over prediction variability. We acknowledge that there is ongoing debate regarding the balance between spatial independence and representativeness in training/testing splits, particularly in ecological studies involving complex environmental gradients. To address this, we examined residual spatial patterns separately to identify areas of potential model underperformance.

For each fold, the model was fitted using the *ranger* package (Wright *et al.*, 2017), an efficient implementation of RF configured with 500 trees. The response variable was the density of tall trees (*#N_trees_h60_km2*). The explanatory variables included climatic, topographic, and edaphic environmental factors, carefully selected to capture spatial variability in environmental conditions. To avoid missing data issues (NA), the complete cases function was applied to the training data, ensuring that only complete records were used to fit the model. After fitting the model in each fold, predictions were made for the corresponding test set. These predictions were stored in a matrix, where each column represented the predictions of a specific fold. The uncertainty associated with the predictions was estimated by calculating the SD across folds for each sample point, creating an uncertainty map. This SD measure reflects the variability in predictions across different folds, indicating areas with greater or lesser confidence in the model estimates. The statistical metrics used to evaluate model performance included the RMSE and the coefficient of determination (R^2) of the tall tree density calculated for each fold. The mean bias was calculated as the mean difference

between the predicted and observed values of tall tree density across all transects:

$$\text{Mean bias} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i),$$

where $\hat{y}_i$ is the predicted value, and y_i is the observed value for transect i , and n is the total number of transects ($n = 900$). This metric provides a straightforward indication of whether the model tends to systematically overestimate or underestimate tall tree density.

After cross-validation, the final model is fitted using the entire available dataset (all 900 transects) and is used to predict giant tree density across the study area, considering environmental factors as predictor variables. The final RF model refers to the structure (i.e. hyperparameters and predictor variables) that achieved the best performance during spatial cross-validation. The importance of environmental variables was analyzed using marginal plots, holding the other variables constant at a mean value. This approach consisted of a sensitivity analysis in which the importance of variables is measured by permuting variables in the model and measuring the increase or decrease in tree density.

Finally, the parameters of the final RF model were applied to stacked environmental layers at the pixel level across the entire Amazon biome using map algebra, producing spatially explicit density maps for the uppermost strata of the forest canopy, with emphasis on giant trees (≥ 60 m) (Fig. S1). All statistical and spatial modeling and analysis procedures were conducted in the R ENVIRONMENT v.4.2.1 (R Core Team, 2024), using the MASS (Ripley *et al.*, 2022) and RANDOMFOREST (Cutler & Wiener, 2022) packages.

Assessing biomass estimates

To estimate AGB as a function of tall tree density in the Amazon biome, we used LiDAR transect biomass reference data developed by Ometto *et al.* (2023). To ensure consistency and methodological independence from interpolated values, we recalculated AGB per transect based on observed biomass estimates. Each 3.75 km² transect was subdivided into square pixels, and biomass was estimated per pixel using LiDAR-derived structural metrics calibrated with field inventory data. These pixel-level values were then aggregated using the PRODES forest mask (250 m $\times$ 250 m resolution), retaining only those pixels classified as forest. For each transect, total AGB was calculated by summing the biomass values of all intersecting PRODES-classified pixels. Finally, AGB values were converted from megagrams per hectare (Mg ha⁻¹) to gigatons per square kilometer (Gt km⁻²), enabling a standardized interpretation of biomass stocks at broader spatial scales. This harmonization of spatial resolution and biomass estimation ensured coherence between the tall tree density and AGB datasets while minimizing redundancy in data derivation.

To assess the relationship between tall tree density and biomass stocks across the Amazon biome, we initially fit a linear model (AGB $\sim$ density_of_tall_trees) to quantify the strength and direction of the correlation, reporting the coefficient of determination

and significance level. However, for better visual interpretation and to account for nonlinear trends – especially at higher density levels – a locally weighted smoothing function (LOESS) was applied to the scatterplot. This methodological approach enables a robust analysis of the role of tall trees in biomass stocks in the Amazon, contributing to the understanding of the spatial distribution of carbon stored in tropical forests and the factors that influence this dynamic at different geographic scales.

Results

Density of tall trees and underlying environmental factors

LiDAR mapping detected 5522 736 trees, with an approximate average of 767 trees km⁻² in c. 7.2 km² of overflow area. The maximum height found was 88.5 m, validating this information in the field as the tallest known tree in South America (Gorgens *et al.*, 2019). The density of tall trees showed weak to moderate positive correlations with the FAPAR ($R = 0.13$), the number of clear days ($R = 0.11$), and soil clay content ($R = 0.12$), while displaying negative associations with wind speed ($R = -0.27$), precipitation seasonality ($R = -0.02$), and lightning frequency ($R = -0.34$), indicating potential limiting effects of wind and disturbance on the occurrence of giant trees (Fig. S3). The observed density of trees ≥ 60 m per transect ranged from 0 to 227 trees km⁻² (see Dataset S1; Fig. S4a,b). The correlation structure revealed moderate to strong associations among several climatic variables. Variables, tannual, tmax, and tseason were strongly positively correlated ($R > 0.7$), as were pannual and pwettest. These patterns reflect expected coherence between related environmental drivers (e.g. different dimensions of temperature or precipitation). Conversely, wind variables (uspeed and vspeed) showed weak to moderate correlations with most other predictors.

PCA revealed the dominant environmental gradients across the Amazon (Fig. 2) and within provinces. The first two components explained 48.7% of the variance (PC1 = 32.1%, PC2 = 16.6%). PC1 summarized broad climatic–disturbance contrasts and water balance, whereas PC2 captured edaphic/topographic variation. These axes separated provinces along moisture–disturbance and soil–terrain gradients, respectively; variable loadings and biplots are shown in Fig. 2.

PC1 was primarily structured by climatic variables, including precipitation seasonality (pseason) and potential evapotranspiration (pet), whereas PC2 reflected a combination of disturbance- and soil-related factors, including wind speeds (uspeed, vspeed), clay content, and lightning frequency. The spread of data points in PCA space showed partial separation of biogeographic provinces, indicating regional differences in environmental conditions.

While PCA does not predict tall tree density directly, the ordination helps contextualize environmental heterogeneity. For instance, provinces such as Guiana Shield and Roraima were positioned in regions of PCA space associated with lower wind speed, higher clay content, and fewer clear days – conditions that may indirectly favor the persistence of tall trees. By contrast, provinces like Pará and Pantepui clustered in areas associated with

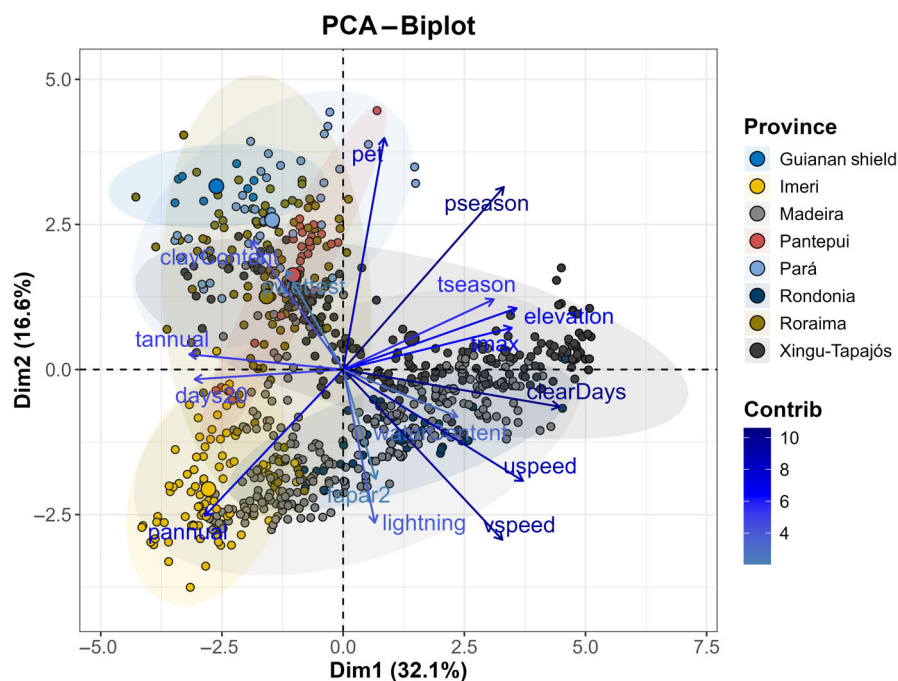


Fig. 2 The biplot shows the first two principal components of the principal component analysis (PCA), together explaining 48.7% of the data variance. The length of the arrows indicates the degree of contribution (influence) of the environmental variables to the principal axes. Variables with longer arrows (such as pseason, pet, and lightning) significantly impact discriminating biogeographic provinces. The direction of the arrows indicates correlation: vectors that are close or aligned indicate positive correlation. Vectors that are opposite ($> 90^\circ$) indicate a negative correlation. For example, pseason positively correlates with Dim1, whereas uspeed and vspeed correlate negatively. The arrow colors reflect each variable's relative contribution to the model, with a gradient from light blue (low contribution) to dark blue (high contribution). Variables with high contribution best explain the observed variation between provinces, such as pet and lightning. Each point represents a sampling unit (site) within a specific biogeographic province. Point sizes in the biplot are scaled using a transformation ($1 + \text{tree density}$) to enhance visual clarity while retaining zero-density information. Ellipses group provinces based on their environmental similarities. Example: Guiana Shield (blue) is associated with specific environmental features such as elevation. Pantepui (red) is strongly associated with clayContent and waterContent. The size of the dots reflects the density of tall trees (contribution of giant species) in each biogeographic province. Areas with larger dots indicate a higher density of tall trees, possibly associated with variables such as lightning and clearDays.

stronger wind regimes, higher climatic seasonality, or more edaphically constrained conditions, aligning with their generally lower density of giant trees (see also Fig. S2). Province-level PCAs further revealed differences in the correlation structure of environmental variables. These local-scale analyses complement the biome-wide PCA by illustrating how the relative influence and collinearity of predictors may vary across ecological regions, supporting the interpretation of region-specific environmental constraints on tall tree distributions.

Spatial modeling

Our analysis confirmed, with a high level of accuracy, that the fitted RF model (k -fold cross-validation: $\text{RMSE} = 4.8 \text{ trees km}^{-2}$) indicates a general spatial trend in the density of tall trees throughout the Amazon basin. The spatial model explained 79% of the variation in tall tree density and presented a significant correlation between the observed and estimated values ($R = 0.96$; $\text{RMSE} = 10.45$, Fig. S4a). The distribution of predictions closely matched observed values (Fig. S4b), although the model tended to slightly underestimate higher densities. Residual analysis revealed systematic overestimation at low densities and

underestimation at higher values, indicating some nonlinearity not fully captured by the model.

The results from the spatial cross-validation demonstrated satisfactory performance of the RF model in estimating the density of tall trees. The RMSE was $10.5 \text{ trees km}^{-2}$, indicating reasonable accuracy in the predictions. By contrast, the mean coefficient of determination (R^2) of 0.6 shows that the model could explain a substantial portion of the variability in the observed data. Additionally, the mean bias of 0.4 reveals a low systematic deviation, suggesting that the model showed good overall accuracy without tendencies toward underestimation or overestimation. Spatially, most transects exhibited low residuals, but localized deviations in northeastern Amazonia suggest areas with greater environmental heterogeneity or reduced predictor performance (Fig. 4c,d). Transects with exceptionally high-observed densities of giant trees tended to have their values underestimated by the model. This pattern likely reflects the rarity of such high-density transects in the training data and highlights a common limitation of machine learning models in capturing extreme values.

The spatial model prediction of large tree density shows differences among biogeographic provinces, Fig. 3, with markedly

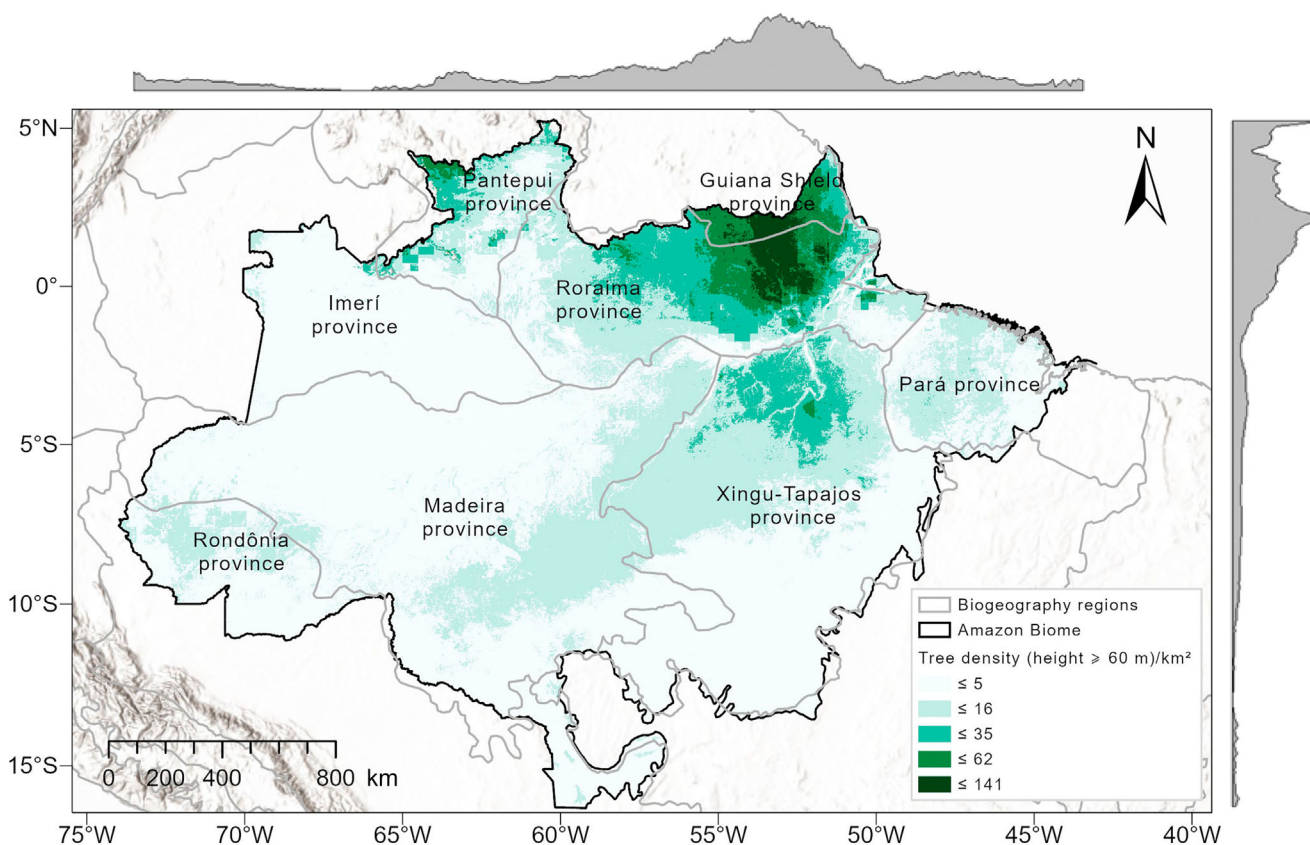


Fig. 3 Potential map generated by the random forest spatial model to define ideal zones for the occurrence of high density of giant trees in the Amazon. The areas are color coded according to different density ranges, where lighter shades indicate lower tree density (≤ 5 trees km^{-2}), darker shades indicate higher density (up to 141 trees km^{-2}). Gray lines delimit major biogeographic provinces within the Amazon biome (such as the Guiana Shield, Xingu-Tapajós, and Roraima), showing distinct density patterns across the Amazon region. Gray shading on the top and side of the map represents latitudinal and longitudinal distribution of density. The highest densities of giant trees are concentrated in the Guiana Shield province and northern Roraima province. By contrast, potential areas with near-zero density of giant trees include parts of the Madeira and Rondônia provinces as well as extensive regions in the southwestern Amazon. The map is available at <https://doi.org/10.5281/zenodo.13850976>.

higher values in the northeastern portion of the Amazon, where there is a predominance of *terra firme* forests in the northern Guiana Shield and the eastern portion of the Roraima biogeographic provinces (darker area on the map, $c. -58$ to -51° longitude, and -4° to -3° latitude). The density of tall trees in these areas may exceed 120 individuals km^{-2} , estimated by the spatial model (Figs S4b, 3). Note that these regions are characterized by a low incidence of lightning and lower wind speeds (Fig. S2). Densities close to zero can be seen in large portions of the south-central eastern Amazon (Pará, Xingu-Tapajós, and Madeira provinces; note that these biogeographic provinces coincide with the arc of deforestation). We also noted lower densities throughout the Imeri province and the entire northern portion of the Madeira province. This westernmost strip of the biome is characterized by the solid influence of the seasonality of the dry season (Fig. S2). Most regions, including the Madeira, Rondônia, and Xingu-Tapajós provinces, exhibited low to moderate uncertainty (≤ 6.86 trees km^{-2}), reflecting well-sampled areas with more predictable environmental conditions. By contrast, higher uncertainty (> 16.03 trees km^{-2}) was

concentrated in northeastern Amazonia, particularly in parts of the Guianan Lowlands, Pará, and Roraima provinces (Fig. 4). Prediction uncertainty is heteroscedastic: absolute errors and interval widths increase with the predicted mean density, particularly in the highest-density bins, whereas relative errors remain comparable across most of the range. This scale effect is as important as variation in sampling intensity or environmental predictability (Fig. S4d).

Across the 4.2 million km^2 Brazilian Amazon biome, the spatial model estimates $c. 55.5$ million trees taller than 60 m (0.0001% of the estimated ≈ 390 billion trees for the Brazilian Amazon biome), with an average density of $13.5 (\pm 4.5)$ tall trees km^{-2} . However, we found a sizeable spatial aggregation of these tall trees, with $c. 14\%$ of tall tree density occurring in 1.28% of the total area of the Amazon (latitudes -2° to 3° S and 53° to 58° W) and 50% occurring in 11.2% (latitudes -5° to 5° S and 51° to 60° W).

The essential variables according to %IncMSE for the model were pannual (13.01%), tannual (11.89%), uspeed (11.65%), and clayContent (10.98%) (Fig. S5b). Meridional wind speed

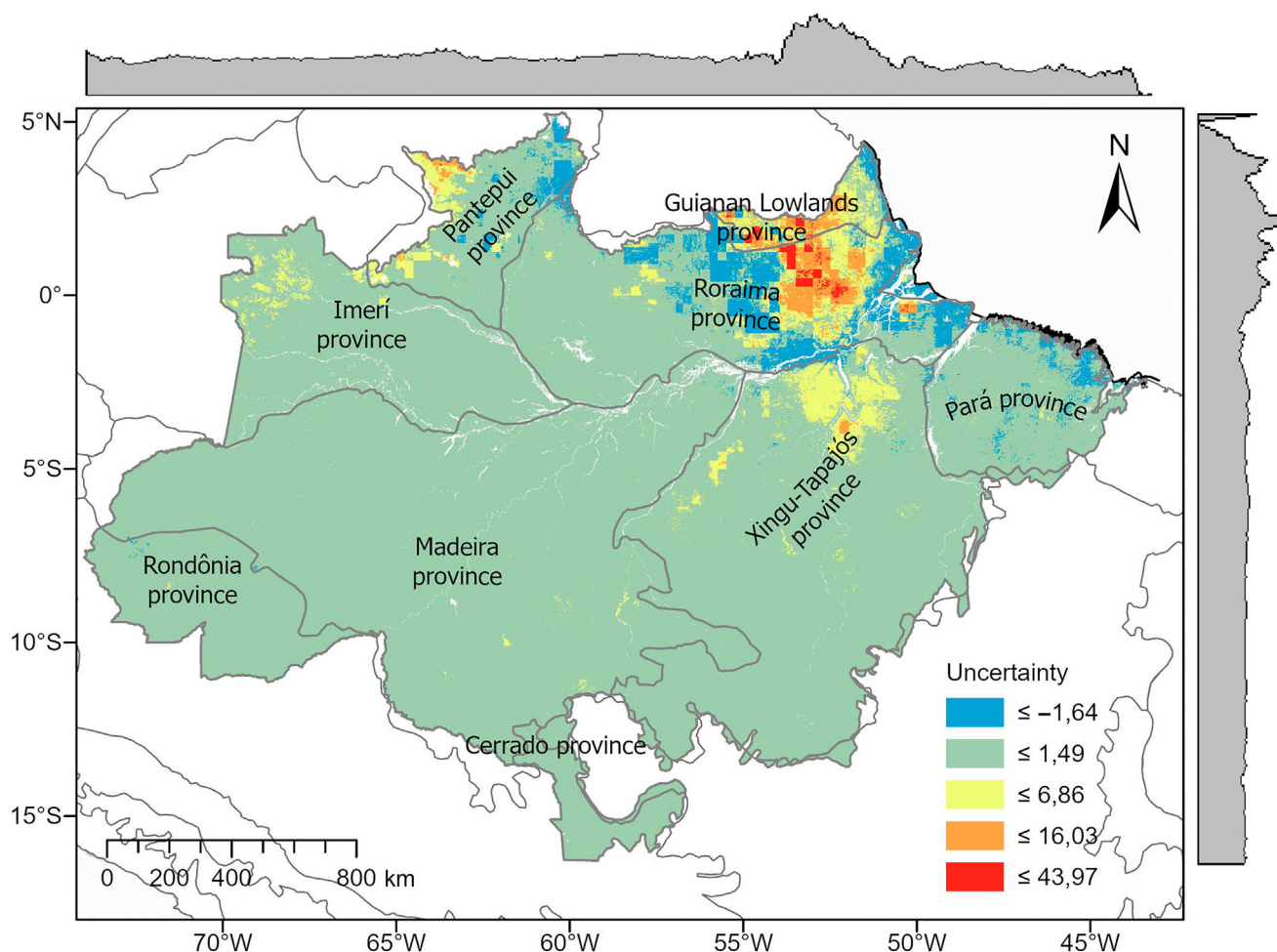


Fig. 4 Relative prediction uncertainty for tall-tree density (height ≥ 60 m) across the Brazilian Amazon. Uncertainty is expressed as the standardized width of the prediction interval (z score of pixel-level PI width from spatial cross-validation/predictive distribution). Thus, negative values indicate below-average uncertainty, and positive values indicate above-average uncertainty. Legend colors denote nonoverlapping value ranges (classes), with warmer colors indicating higher uncertainty and cooler colors indicating lower uncertainty. Biogeographic provinces are delineated for reference. Uncertainty is generally low across most of the biome but is concentrated in the northeast, particularly in the Guianan Lowlands, Pará, and Roraima provinces. Gray marginal bars at the top and right depict the latitudinal and longitudinal distributions of modeled tall-tree density.

(N–S, m s^{-1}) (vspeed) and lightning rate, claycontent, and number of clear days per year (clearDays) were found to be the primary environmental factors associated with decreases in spatial model residual variance (IncNodePurity) and contributed significantly to separating the observations into more homogeneous groups along the trees in the RF model (Fig. S5a). High mean and maximum temperatures also negatively affected the density of tall trees. Tree density increased with photosynthetically active radiation (fapar) by $> 70\%$ but decreased with the number of clear days linked to direct radiation (Fig. S6). Mean annual precipitation (pannual) and precipitation regimes (pwettest, days20), as well as soil clay content (Claycontet), elevation above sea level (m) (elev), and temperature seasonality (tseason), indicate increases and stability in spatial model-estimated density values. Although water availability is a fundamental driver of tall tree density, we observed that biome-wide relationships between climate-based indicators – such as precipitation seasonality

(pseason) and potential evapotranspiration (pet) – and tall tree density were generally weak or inconsistent. This suggests that large-scale climatic seasonality alone does not fully explain the spatial variation in giant tree occurrence. One possible explanation is that local-scale ecological processes, including topographic heterogeneity, soil buffering capacity, or historical disturbance regimes, may mask or override broader climatic signals in shaping density patterns (Fig. S7). By contrast, the effect of soil water content showed a clearer pattern: a generally positive or unimodal response, where tall tree density declined under very low water content (indicative of stress), but increased once a certain threshold of soil moisture was surpassed (Fig. S6).

Implications for biomass estimates

At a regional and biogeographic scale, we compared our tall tree density estimates with the potential AGB stock map produced by

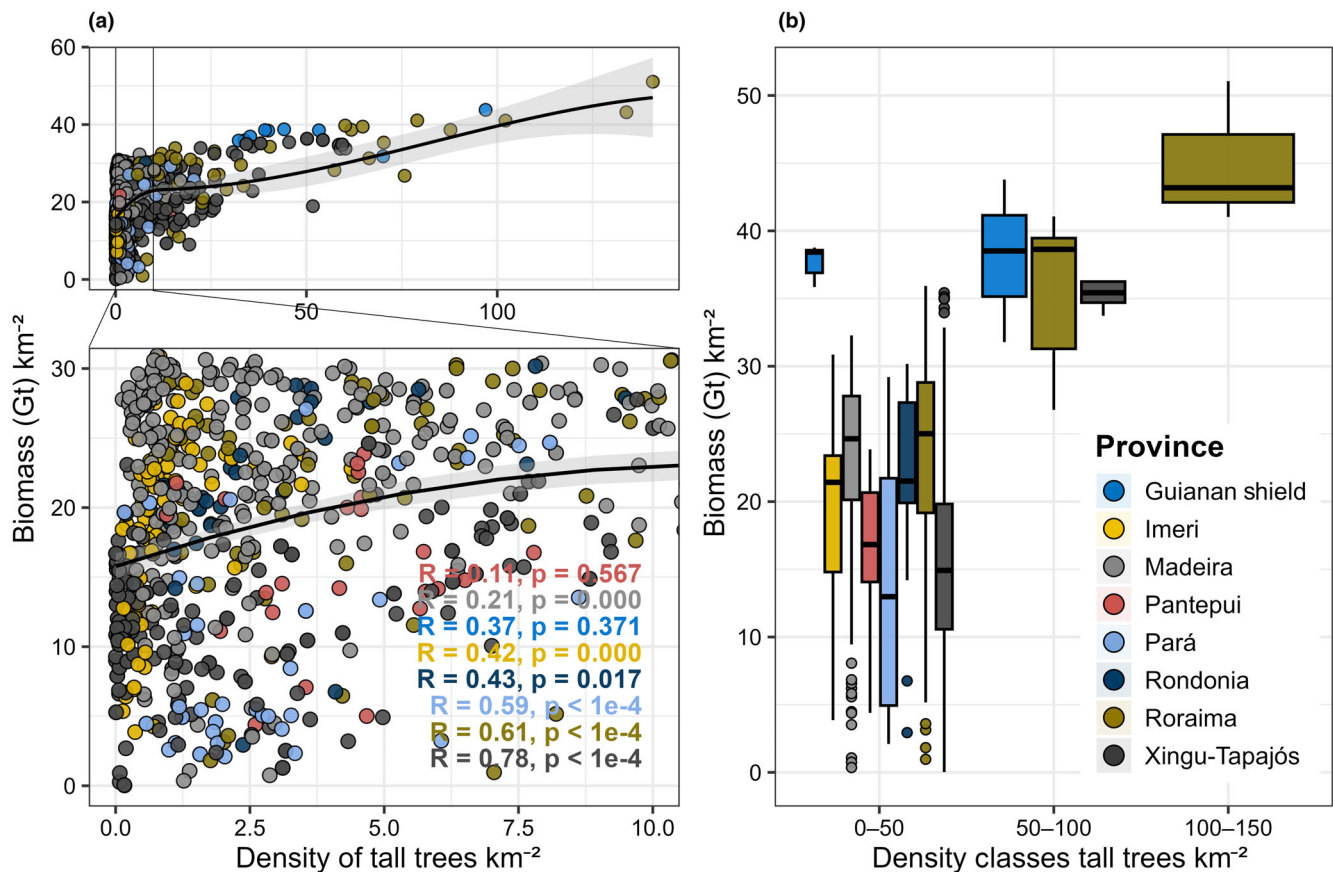


Fig. 5 Relationship between aboveground biomass (AGB) and the density of tall trees (i.e. $\geq 60 \text{ m}$) for the 900 LiDAR plots studied in the Amazon biome. (a) Scatterplots colored by biogeographic province, showing AGB (Gt km^{-2}) as a function of tall tree density (trees km^{-2}). The black curve is a LOESS smoother to aid visual interpretation. The inset colored annotations report province-specific linear model statistics (R and P) used to quantify association strength; linear regression lines are not plotted to avoid over-plotting and misleading linearity when the relationship is non-linear at higher densities (gray shading representing confidence intervals ($\text{CI} = 95\%$)). The top panel highlights density values $< 10 \text{ trees km}^{-2}$, while the bottom panel shows the full distribution. (b) Boxplots of biomass values across three tall-tree density classes (0–50, 50–100, and 100–150 trees km^{-2}). The horizontal lines in the boxplots represent the first, second (median), and third quartiles of the biomass data distribution. The vertical lines indicate the interquartile range of the data.

the same airborne LiDAR campaign as this study (Ometto *et al.*, 2023). As expected, a higher density of tall trees is strongly associated with higher AGB ($R = 0.5$, $P < 2.2e-16$) (Fig. 5a). These relationships are more pronounced where large tree density is higher, specifically in the provinces of Roraima and Guiana Shield, which also correspond to the largest AGB densities across the biome. Interestingly, Xingu-Tapajós and Madeira have relatively high AGB even in areas with a low density of tall trees.

However, in areas with higher densities of tall trees, the relationship is not linear and may indicate saturation at extreme density rates. For example, areas with low density of tall trees may still have significant biomass levels due to the contribution of smaller trees. The relationship between giant tree density and AGB was also analyzed in different classes of tree density (0–50, 50–100, and 100–150 trees km^{-2}) (Fig. 5b). We observed that, across the biome, spatial patterns agree well and that our density estimates are predictive of biomass even in biogeographic provinces with densities of up to 50 tall trees km^{-2} . Notably, the

relationship between biomass and density is most significant in areas with > 100 tall trees km^{-2} (Fig. 5b). The results reveal a consistent increase in biomass with increasing tall tree density, mainly in Roraima and the Guiana Shield. The classes with the highest density (100–150 trees km^{-2}) presented the highest median biomass values (40.0–50.0 Gt km^{-2}) and the lowest variability, while the lower density classes (0–50 and 50–100 trees km^{-2}) exhibited more visible dispersion in biomass values, reflecting structural heterogeneity.

Discussion

Patterns of tall tree density and environmental drivers

The density of tall trees is a valuable input for modeling large-scale biological and biogeochemical processes and a prominent component of ecosystem structure, governing elementary processes such as carbon, water, and nutrient retention, competitive

dynamics, and habitat suitability for many plant and animal species (Slik *et al.*, 2013; Bastin *et al.*, 2018; Pinho *et al.*, 2020). Our results show that a complex combination of environmental factors influences the density of tall trees in the Amazon. Clay-rich soils, moderate precipitation and temperature, and low incidence of lightning and strong winds create stable conditions for a high density of tall trees. These findings are consistent with the existing literature highlighting the importance of water availability and stable climatic conditions for the growth and maintenance of large trees (Venter *et al.*, 2017; Gorgens *et al.*, 2021). Areas located in the northeastern portion of the biome, between the provinces of Roraima and the Guiana Shield (*c.* -58° to -51° longitude, and -4° to -3° latitude), have the highest local density of tall trees across the Amazon. We estimate up to 140 individuals per km^2 in these areas, where well-drained hilltops *terra firme* forests predominate. These patterns confirm previous results at the same scale for biomass (Ometto *et al.*, 2023), maximum height (Gorgens *et al.*, 2021), and diversity of large species (De Lima *et al.*, 2023). Furthermore, these areas are among the most remote and least explored in the Amazon (Carvalho *et al.*, 2023), suggesting a possible role of a low current human population density and the absence of roads or large infrastructures as factors controlling the occurrence of tall trees. Being isolated areas with low anthropic pressure enhances the conservation value of these areas.

Lightning and strong winds associated with convective storms and cold fronts in the southern and western Amazon can cause significant mortality of tall trees through damage, toppling, or structural failure (Arellano *et al.*, 2019; Cushman *et al.*, 2021; Ibanez *et al.*, 2024). The relative height and canopy exposure of giant trees makes them particularly vulnerable to these disturbances (Rifai *et al.*, 2016; Cushman *et al.*, 2021; Ibanez *et al.*, 2024). In regions where tall trees are scarce, this may reflect selective mortality pressure exerted by recurrent windthrow or lightning events, rather than an absence of disturbance. Such filtering processes may help explain regional variations in tall tree density across the biome. Recurrent increases in wind speed can significantly modify density patterns in many regions (Marra *et al.*, 2014). Wind and lightning act as disturbance constraints across tropical forests of varying canopy height. While exposure and crown architecture likely amplify risk for tall trees (≥ 60 m), empirical studies demonstrate that lightning is a leading cause of large-tree mortality even in forests with 25–45 m canopies (Gora *et al.*, 2020; Yanoviak *et al.*, 2020; Gora & Esquivel-Muelbert, 2021). Since most trees in the biome are not tall (average canopy height 30–45 m, e.g. Lang *et al.*, 2023), lightning and wind are much less critical for forests with lower canopies (Zoletto *et al.*, 2023). In our maps, provinces with higher wind speed and lightning frequency show lower modeled densities of giant trees, consistent with these agents as limiting factors rather than implying they are unimportant in shorter forests.

At local scales, topography and hydrological conditions play a critical role in shaping the density of tall trees. Our results support previous findings that giant trees are more frequently associated with elevated terrains, such as plateaus, which offer more stable edaphic and hydrological conditions (Jucker *et al.*, 2018;

Zuleta *et al.*, 2020). By contrast, lowland areas, particularly those < 50 m above sea level and close to river systems, experience frequent or seasonal flooding that imposes oxygen and nutrient limitations on root systems (Meinzer *et al.*, 2001; Koch *et al.*, 2004). These conditions limit vertical growth by favoring species adapted to flooding stress over those that invest in structural height (Schöngart *et al.*, 2002; Durgante *et al.*, 2023). Moreover, hydrological variability and sediment loads can lead to unstable nutrient availability, further constraining the development of tall trees (Koch *et al.*, 2004; Bittencourt *et al.*, 2020; Bartholomew *et al.*, 2022). Because tall trees have high energy and hydraulic demands, particularly for transporting water and nutrients over long vertical distances, any restriction in root oxygenation can compromise their growth and survival (Domec *et al.*, 2008; Couvreur *et al.*, 2018; Skiadasis *et al.*, 2021). Thus, lowland and seasonally flooded environments tend to support tree communities with shorter statures, better suited to mechanical and physiological stability under these challenging conditions.

Our findings also indicate that clay and soil water content are critical for the occurrence and density of giant trees in the Amazon, as they directly influence nutrient availability and water stability, which are essential for the growth of these trees (Quesada *et al.*, 2011; Spanner *et al.*, 2022). Soils with high clay content have a higher cation exchange capacity, which allows the retention of essential nutrients such as calcium, magnesium, and potassium, favoring the sustained growth of trees. In addition, clay contributes to the formation of aggregates in the soil, improving its structure and increasing water retention. In Amazonian regions with clayey soils, giant trees' density and biomass are higher than in areas with sandy soils due to their water retention capacity (Durgante *et al.*, 2023). Soil water availability, especially during dry periods, is crucial for the survival of these trees, and clayey soils provide a stable source of moisture. However, other studies suggest that in the tropics, higher soil fertility is often associated with lower maximum canopy height (Muller-Landau *et al.*, 2021; Jucker & Ali, 2023). One proposed explanation is that faster-growing plants in fertile soils may prioritize growth over structural defenses or longevity, resulting in increased vulnerability to disturbances such as windthrow (Jackson *et al.*, 2021; Joswig *et al.*, 2022; Durgante *et al.*, 2023; Sanchez-Martinez *et al.*, 2025). While this growth–defense trade-off hypothesis offers a compelling framework, especially in explaining species-level differences in wood density and mortality risk, it remains challenging to test directly and should be interpreted as a conceptual model in need of further empirical validation.

In environments well drained with regular rainfall and defined seasons, there is a constant supply of water in the soil, which is essential for continued growth and for the maintenance of complex physiological processes that require high levels of water, such as transpiration and photosynthesis for many tall trees (Fig. S5b), and this is directly related to the number of days with clouds or days with few clouds ('clear days'). An increase in cloudless days indicates a greater incidence of sunlight and high temperatures, which can increase the rate of photosynthesis, favoring tree growth to a thermal tolerance threshold (Green *et al.*, 2020; Sullivan *et al.*, 2020).

Although atmospheric drought – driven by increased temperatures, clear-sky radiation, and reduced precipitation – can temporarily enhance photosynthetic activity when water is abundant (Aguirre-Gutiérrez *et al.*, 2019; Green *et al.*, 2020), prolonged drought conditions increase the risk of hydraulic failure, particularly in tall trees. Taller individuals are thought to be more vulnerable to xylem embolism due to longer water transport pathways and higher evaporative demand under exposure to intense solar radiation (Oliveira *et al.*, 2019; Garcia *et al.*, 2023; Mattos *et al.*, 2023). Indeed, Bennett *et al.* (2015) found a consistent negative effect of drought on the growth and survival of large trees at the global scale. However, it is important to note that this pattern is not universal—studies such as (Fauset *et al.*, 2012; Zuleta *et al.*, 2017) have shown that smaller trees can also exhibit elevated mortality under drought conditions, depending on forest structure, rooting depth, and competition for light and water. Therefore, while tree height is a key factor influencing drought vulnerability, drought-induced mortality patterns are context-dependent, and shaped by the interaction of multiple structural and environmental factors. On broader scales, climatic drivers such as the number of clear days and precipitation seasonality have also been linked to spatial patterns in tree density (Bruijnzeel *et al.*, 2011; Wagner *et al.*, 2016; Jiang *et al.*, 2017; Ehbrecht *et al.*, 2021).

Importance of giant trees density for biomass stocks

The positive relationship between tall tree density and biomass for the entire biome and biogeographic provinces highlights the importance of giant trees, which play an essential role in biogeochemical processes despite representing a small fraction of the total trees (Bastin *et al.*, 2018; Ali & Wang, 2021). For example, in the tropics, Sullivan *et al.* (2020) showed that only 1% of trees in tropical forests account for *c.* 50% of AGB. Slik *et al.* (2009, 2013) suggest that the impact of large trees (dbh ≥ 70 cm) on AGB is, on average, 25.1% in South America and more than two-thirds at the pantropical scale. These results are consistent with the meta-analysis by de Lima *et al.* (2022) in the Amazon biome, which highlighted that only one giant tree (dbh ≥ 70 cm) could accumulate 82% of all stored AGB Mg ha⁻¹. In general, tall tree density maps can potentially characterize habitat heterogeneity directly (Tuanmu & Jetz, 2015; Bastin *et al.*, 2018), so canopy height has been classified as a high-priority biodiversity variable to be observed from space (Lang *et al.*, 2023). Therefore, the density of tall trees can be used as a direct indicator of the potential for carbon storage in tropical forests.

Our comparison between giant-tree density and LiDAR-derived AGB should be interpreted as a relationship between two structural metrics derived from canopy height, rather than as an independent validation. The AGB surface is modeled from top-canopy height calibrated with field plots (Ometto *et al.*, 2023), whereas giant-tree density reflects the upper-tail exceedance of the height distribution (height ≥ 60 m). This shared dependence on LiDAR height introduces partial circularity, making a positive association unsurprising. We therefore avoid causal language and frame the result as co-variation within canopy structure: AGB

largely captures stand-level central tendency of height and biomass, while giant-tree density captures the frequency of extreme canopy elements. Future work will quantify the incremental information provided by the exceedance metric (e.g. partial correlations or SEM controlling for mean or P90 height) and, where possible, confront both layers with fully independent field-based biomass estimates.

Our study did not test the effect of local processes influencing the density of tall trees on AGB. However, a possible explanation for the high AGB in biogeographic provinces dominated by large trees may be related to the reproductive strategy of wind dispersal of some species, that is they need to be tall and emergent to effectively disperse their seeds and exceptionally maximize their geographic influence (Slik *et al.*, 2013, 2024). In the Amazon, the most abundant larger species (dbh ≥ 70 cm) were *Goupia glabra*, *D. excelsa*, *Aspidosperma excelsum*, *Couratari guianensis*, *Manilkara huberi*, *Dipteryx odorata*, *Tabebuia serratifolia*, *Bertholletia excelsa*, and *Caryocar villosum* (de Lima *et al.*, 2022). These species represented *c.* 20% of the individuals sampled in 240 plots inventoried in the eight biogeographic provinces in the biome. Therefore, although environmental controls influence maximum tree height and the diversity of large species at a regional scale (Gorgens *et al.*, 2021; De Lima *et al.*, 2023), this dominance is also driven by species-specific traits that provide the ability to reproduce and support large biomass stocks at a local scale (Loubota Panzou *et al.*, 2018). Different strategies in seed dispersal are linked to maximum tree height and aggregation patterns (Wunderle, 1997; Clark *et al.*, 2005; Slik *et al.*, 2024). In particular, wind-dispersed species such as *D. excelsa*, *A. excelsum*, and *T. serratifolia* show more aggregated patterns than animal-dispersed species such as *B. excelsa* and *M. huberi*, which may affect large tree density and AGB at the local scale, as animals are more effective at dispersing seeds in a more uniform spatial distribution (Wright *et al.*, 2003). This finding emphasizes the importance of including species traits and changes in species composition as explanations for AGB gradients in tropical forests (Slik *et al.*, 2009, 2013; Loubota Panzou *et al.*, 2018).

Another critical factor at the local scale is gap formation, which provides opportunities for seed germination and recruitment, especially in areas where light is a limiting resource. Higher densities of tall trees produce more significant gaps when they fall (Arasa-Gisbert *et al.*, 2021; Reis *et al.*, 2022), although the frequency and size of these gaps are also influenced by disturbances such as wind and lightning (Hubbell *et al.*, 1999; Cramer *et al.*, 2007; Cushman *et al.*, 2021). A relatively low proportion of significant gaps is observed in the provinces of Roraima and Guiana Shield, where we found the highest densities of tall trees and stored biomass (probably caused by the low mortality rate of large trees caused by strong winds and lightning). These results suggest that low gaps may limit the recruitment of tall trees, contributing to their aggregated patterns and more prominent AGB stock over time, as germination and recruitment may be limited to fewer areas with more significant gaps. For example, clusters of *D. excelsa* and *T. serratifolia* can be found in an aggregated form, likely due to a combination of anemochoric seed dispersal, a potential dependence on light for germination and seed

recruitment, as well as local topography (lowlands vs plateaus) that favor their aggregation patterns (Lewis *et al.*, 2017; Ali *et al.*, 2020; de Lima *et al.*, 2022).

The results presented here reinforce the hypothesis that preserving areas with a high density of giant trees is crucial to maintaining ecosystem services associated with carbon storage. Including tall tree density as a critical variable in biomass estimation models can substantially improve the accuracy of these estimates. These models need to better account for the role of these large trees, and airborne LiDAR surveys are practical tools to detect and quantify large trees and quantify their contributions to biomass stocks (Asner *et al.*, 2010; Coomes *et al.*, 2017). This is especially interesting in difficult-to-access regions with limited ground-based measurements. The observed relationship suggests that the loss of tall trees due to deforestation or forest degradation can significantly reduce biomass stocks. Inadequate management that does not consider the conservation of these trees can have drastic implications for the carbon budget, amplifying the impacts of glasshouse gas emissions.

Broader implications and concluding observations

The main findings of this study are significant for understanding the ecology and conservation of giant trees in the Amazon biome. We also detail some of the main limitations of the study that can be seen in notes S1. Mapping tall tree density provides valuable insights into forest structure and health, which are crucial for biodiversity conservation and climate change mitigation. The data generated can inform forest management and conservation strategies, helping to identify areas in need of protection or restoration. Furthermore, by linking tree density to environmental variables, the study can reveal the key factors influencing the distribution and health of tall trees and offers a beacon of hope in the face of climate change and deforestation, providing a roadmap for more effective conservation efforts. Practical application of the data must be accompanied by a deep understanding of ecological interactions and an adaptive approach to environmental management.

The relationship between tall tree density and local factors, such as soil characteristics, dispersal strategies, and gap formation, is a promising area for future investigation. The hypothesis that local density correlates with reproductive success and recruitment can be tested in different soil types and terrains. Furthermore, exploring density patterns across topographic gradients and relating them to climatic and edaphic variables can provide insights into the interaction between regional and local factors. Regional factors, such as wind and lightning, shape tall tree density patterns in the Amazon. However, they are refined at local scales by processes such as dispersal, topography, and gap formation. These results highlight the importance of integrating multiple scales of analysis to understand the ecology of giant trees in the Amazon and provide support for their conservation under a climate change scenario.

This study contributes to the discussion on the relevance of approaches based on high-resolution data, such as those provided by LiDAR sensors, to better understand tropical forests' structural

dynamics. Furthermore, combining tall tree density data with climate change projections could provide insights into the future of carbon storage in the Amazon. Considering their disproportionate importance in biomass stocks, conserving tall trees should be a priority in global conservation initiatives such as REDD+ (Reducing Emissions from Deforestation and Forest Degradation). Investments in remote monitoring and forest inventories that integrate metrics from emerging trees can improve understanding of forest dynamics and support global mitigation strategies.

Although this study does not aim to model the potential impacts of climate change, it is concerning to note that several climate variables strongly associated with giant tree density may undergo significant changes in future climate crisis scenarios. Environmental changes induced by rising temperatures are already being observed. Changes in environmental variables associated with disturbances can significantly negatively impact the density and survival of large trees. For example, the frequency of anomalous events, such as increased storms and lightning strikes, has already been observed (Dale *et al.*, 2001; Seidl *et al.*, 2017; Bauman *et al.*, 2022; Kamimura *et al.*, 2022). Furthermore, in the current scenarios projected by the sixth climate report of the Intergovernmental Panel on Climate Change (IPCC), the average global temperature is estimated to increase by 1.5°C between 2030 and 2052. If temperature continues to increase and precipitation levels decrease, drought scenarios in the Amazon will become more frequent, directly impacting all biodiversity (Flores *et al.*, 2024). Even in this scenario, only *c.* 15% of the Brazilian Amazon is protected by conservation units, covering only *c.* 58% of the remaining vegetation. Given the ecological importance of giant trees, understanding the effects of climate change on density patterns is critical and it should be explored quickly at finer scales within the Amazon. This knowledge is critical to refining conservation perspectives in a changing world – for example, to what extent are protected areas in the Amazon currently susceptible to impacts induced by climate disruptions, and how can the giant trees withstand or respond to these changes? Efforts to understand how deforestation and climate change interact and mitigate their impacts are urgently needed in light of the high and increasing rates of deforestation in the Brazilian Amazon, which directly threaten sanctuaries of ancient trees throughout the biome.

Acknowledgements

Funding was provided by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES; Finance Code 001); Conselho Nacional de Desenvolvimento Científico e Tecnológico (Processes 403297/2016-8, 301432/2022-8, 301661/2019-7, 306333/2024-4, and 444350/2024-1), Amazon Fund (grant 14.2.0929.1), National Academy of Sciences and US Agency for International Development (grant AID-OAA-A-11-00012), R. de Lima was supported by the Universidade do Estado do Amapá (Processes 0022.0279.1202.0018/2021). M. Nunes was supported by the funding from National Aeronautics and Space Administration (NASA) contract NNL 15AA03C for the development and execution of the GEDI mission. P. Groenendyk was supported by financial support from the São Paulo

Research Foundation – FAPESP, Young Investigator Grant (2018/01847-0). D. Almeida was supported by the São Paulo Research Foundation (#2018/21338-3 and #2019/14697-0). N. Higuchi and P. Aparicio were supported by INCT-Madeiras da Amazônia and Next Generation Ecosystem Experiments-Tropics (NGEE-Tropics), as part of DOE's Terrestrial Ecosystem Science Program Contract No. DE-AC02-05CH11231. T. Jackson was supported by the UK Natural Environment Research Council grant NE/S010750/1. J. Aguirre-Gutiérrez was supported by NERC NE/T011084/1 and NE/Z504191/1, the Leverhulme Trust RPG-2024-342, and the Royal Society RG\R1\251370. EBG was supported by FAPEMIG PQ-00943-21 and APQ-00185-22 by CNPq projects 403297/2016-8, 401053/2019-9, and 306386-2022-4. The Article Processing Charge for the publication of this research was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (ROR identifier: 00x0ma614). The Article Processing Charge for the publication of this research was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (ROR identifier: 00x0ma614).

Author contributions

RBL designed research; RBL, DAS, MHN, PRLB, PG, JA-G, CPO, RLCF, JAAS, TJ, DGS, JPO, and EBG performed research; RBL, MHN, TJ, NH, DRAA, JPO, and EBG provided funding; RBL, CPO, and EBG analyzed data; and RBL, MHN, PRLB, JA-G, CPO, DAS, PG, and EBG wrote the paper, and all authors edited the manuscript.

ORCID

Jesús Aguirre-Gutiérrez  <https://orcid.org/0000-0001-9190-3229>
 José A. Aleixo da Silva  <https://orcid.org/0000-0003-0675-3524>
 Danilo R. Alves de Almeida  <https://orcid.org/0000-0002-8747-0085>
 Perseu da Silva Aparício  <https://orcid.org/0000-0002-6696-529X>
 Paulo R. de Lima Bittencourt  <https://orcid.org/0000-0002-1618-9077>
 Rinaldo L. Caraciolo Ferreira  <https://orcid.org/0000-0001-7349-6041>
 Eric Bastos Görgens  <https://orcid.org/0000-0003-2517-0279>
 Daniela Granato-Souza  <https://orcid.org/0000-0002-7011-0892>
 Peter Groenendyk  <https://orcid.org/0000-0003-2752-6195>
 Marcelino Carneiro Guedes  <https://orcid.org/0000-0003-2702-5614>
 Niro Higuchi  <https://orcid.org/0000-0002-1203-4502>
 Toby Jackson  <https://orcid.org/0000-0001-8143-6161>
 Robson Borges de Lima  <https://orcid.org/0000-0001-5915-4045>
 Matheus Henrique Nunes  <https://orcid.org/0000-0001-9979-6456>

Cinthia Pereira de Oliveira  <https://orcid.org/0000-0001-5158-7866>

Jean Pierre Ometto  <https://orcid.org/0000-0002-4221-1039>

Diego Armando Silva da Silva  <https://orcid.org/0000-0003-1018-3640>

Joselane P. Gomes da Silva  <https://orcid.org/0000-0002-2496-9516>

José Julio de Toledo  <https://orcid.org/0000-0002-1778-0117>

Fabien H. Wagner  <https://orcid.org/0000-0002-9623-1182>

Data availability

The data that support the findings of this study are available in blank; in blank; Zenodo at https://data.niaid.nih.gov/resources?id=zenodo_4091221. These data were derived from the following resources available in the public domain: <https://zenodo.org/records/4091222#.YDc7l2hKjIV>.

References

- Aguirre-Gutiérrez J, Oliveras I, Rifai S, Fauset S, Adu-Bredu S, Affum-Baffoe K, Baker TR, Feldpausch TR, Gvozdevaite A, Hubau W *et al.* 2019. Drier tropical forests are susceptible to functional changes in response to a long-term drought. *Ecology Letters* 22: 855–865.
- Ali A, Sanaei A, Li M, Asadi Nalivan O, Pour MJ, Valipour A, Karami J, Aminpour M, Kaboli H, Askari Y. 2020. Big-trees – energy mechanism underlies forest diversity and aboveground biomass. *Forest Ecology and Management* 461: 117968.
- Ali A, Wang L-Q. 2021. Big-sized trees and forest functioning: current knowledge and future perspectives. *Ecological Indicators* 127: 107760.
- Arasa-Gisbert R, Arroyo-Rodríguez V, Galán-Acedo C, Meave JA, Martínez-Ramos M. 2021. Tree recruitment failure in old-growth forest patches across human-modified rainforests. *Journal of Ecology* 109: 2354–2366.
- Araújo I, Marimon B, Marimon-Junior BH, Oliveira C, da Silva JW, Beú R, Silva I, Simioni P, Tavares J, Phillips O *et al.* 2024. Taller trees exhibit greater hydraulic vulnerability in southern Amazonian forests. *Environmental and Experimental Botany* 226: 105905.
- Arellano G, Medina NG, Tan S, Mohamad M, Davies SJ. 2019. Crown damage and the mortality of tropical trees. *New Phytologist* 221: 169–179.
- Asner GP, Powell GVN, Mascaro J, Knapp DE, Clark JK, Jacobson J, Kennedy-Bowdoin T, Balaji A, Paez-Acosta G, Victoria E *et al.* 2010. High-resolution forest carbon stocks and emissions in the Amazon. *Proceedings of the National Academy of Sciences, USA* 107: 16738–16742.
- Baret F, Weiss M, Lacaze R, Camacho F, Makhmara H, Pacholczyk P, Smets B. 2013. GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part1: principles of development and production. *Remote Sensing of Environment* 137: 299–309.
- Barros F d V, Bittencourt PRL, Brum M, Restrepo-Coupe N, Pereira L, Teodoro GS, Saleska SR, Borma LS, Christoffersen BO, Penha D *et al.* 2019. Hydraulic traits explain differential responses of Amazonian forests to the 2015 El Niño-induced drought. *New Phytologist* 223: 1253–1266.
- Bartholomew DC, Banin LF, Bittencourt PRL, Suis MAF, Mercado LM, Nilus R, Burslem DFRP, Rowland L. 2022. Differential nutrient limitation and tree height control leaf physiology, supporting niche partitioning in tropical dipterocarp forests. *Functional Ecology* 36: 2084–2103.
- Bastin J, Rutishauser E, Kellner JR, Saatchi S, Péliissier R, Hérault B, Slik F, Bogaert J, De Cannière C, Marshall AR *et al.* 2018. Pan-tropical prediction of forest structure from the largest trees. *Global Ecology and Biogeography* 27: 1366–1383.
- Bauman D, Fortunel C, Delhay G, Malhi Y, Cernusak L, Bentley L, Rifai S, Aguirre-Gutiérrez J, Oliveras I, Phillips O *et al.* 2022. Tropical tree mortality has increased with rising atmospheric water stress. *Nature* 608: 528–533.

- Bennett AC, McDowell NG, Allen CD, Anderson-Teixeira KJ. 2015. Larger trees suffer most during drought in forests worldwide. *Nature Plants* 1: 1–5.
- Birdsey RA, DellaSala DA, Walker WS, Gorelik SR, Rose G, Ramirez CE. 2023. Assessing carbon stocks and accumulation potential of mature forests and larger trees in U.S. federal lands. *Frontiers in Forests and Global Change* 5: 1074508.
- Bittencourt PRL, Oliveira RS, da Costa ACL, Giles AL, Coughlin I, Costa PB, Bartholomew DC, Ferreira LV, Vasconcelos SS, Barros FV *et al.* 2020. Amazonia trees have limited capacity to acclimate plant hydraulic properties in response to long-term drought. *Global Change Biology* 26: 3569–3584.
- Brando P. 2018. Tree height matters. *Nature Geoscience* 11: 390–391.
- Bruijnzeel LA, Kappelle M, Mulligan M, Scatena FN. 2011. Tropical montane cloud forests: state of knowledge and sustainability perspectives in a changing world. In: Bruijnzeel LA, Scatena FN, Hamilton LS, eds. *Tropical montane cloud forests: science for conservation and management. International hydrology series*. Cambridge University Press, 691–740.
- Calhoun P, Hallett M, Su X, Cafri G, Levine R, Fan J. 2020. Random forest with acceptance–rejection trees. *Computational Statistics* 35: 983–999.
- Caron TMF, Chuma VJUR, Sandi AA, Norris D. 2021. Big trees drive forest structure patterns across a lowland Amazon regrowth gradient. *Scientific Reports* 11: 3380.
- Carvalho RL, Resende AF, Barlow J, França FM, Moura MR, Maciel R, Alves-Martins F, Shutt J, Nunes CA, Elias F *et al.* 2023. Pervasive gaps in Amazonian ecological research. *Current Biology* 33: 3495–3504.e4.
- Clark CJ, Poulsen JR, Bolker BM, Connor EF, Parker VT. 2005. Comparative seed shadows of bird-, monkey-, and wind-dispersed trees. *Ecology* 86: 2684–2694.
- Coomes DA, Dalponte M, Jucker T, Asner GP, Banin LF, Burslem DFRP, Lewis SL, Nilus R, Phillips OL, Phua M-H *et al.* 2017. Area-based vs tree-centric approaches to mapping forest carbon in Southeast Asian forests from airborne laser scanning data. *Remote Sensing of Environment* 194: 77–88.
- Couvreur V, Ledder G, Manzoni S, Way DA, Muller EB, Russo SE. 2018. Water transport through tall trees: a vertically explicit, analytical model of xylem hydraulic conductance in stems. *Plant, Cell & Environment* 41: 1821–1839.
- Cramer J, Mesquita R, Williamson G. 2007. Forest fragmentation differentially affects seed dispersal of large and small-seeded tropical trees. *Biological Conservation* 137: 415–423.
- Crowther TW, Glick HB, Covey KR, Bettigole C, Maynard DS, Thomas SM, Smith JR, Hintler G, Duguid MC, Amatulli G *et al.* 2015. Mapping tree density at a global scale. *Nature* 525: 201–205.
- Cushman K, Burley J, Imbach B, Saatchi S, Silva C, Vargas O, Zraggen C, Kellner J. 2021. Impact of a tropical forest blowdown on aboveground carbon balance. *Scientific Reports* 11: 11279.
- Cutler F, Wiener R. 2022. *RANDOMFOREST: Breiman and cutler's random forests for classification and regression*.
- Dale VH, Joyce LA, McNulty S, Neilson RP, Ayres MP, Flannigan MD, Hanson PJ, Irland LC, Lugo AE, Peterson CJ *et al.* 2001. Climate change and forest disturbances: climate change can affect forests by altering the frequency, intensity, duration, and timing of fire, drought, introduced species, insect and pathogen outbreaks, hurricanes, windstorms, ice storms, or landslides. *Bioscience* 51: 723–734.
- De Lima RB, Gorgens EB, Da Silva DAS, De Oliveira CP, Batista APB, Caraciolo Ferreira RL, Costa FRC, Ferreira de Lima RA, Da Silva Aparício P, De Abreu JC *et al.* 2023. Giants of the Amazon: how does environmental variation drive the diversity patterns of large trees? *Global Change Biology* 29: 4861–4879.
- Domec J-C, Lachenbruch B, Meinzer FC, Woodruff DR, Warren JM, McCulloh KA. 2008. Maximum height in a conifer is associated with conflicting requirements for xylem design. *Proceedings of the National Academy of Sciences, USA* 105: 12069–12074.
- Draper FC, Costa FRC, Arellano G, Phillips OL, Duque A, Macía MJ, ter Steege H, Asner GP, Berenguer E, Schiatti J *et al.* 2021. Amazon tree dominance across forest strata. *Nature Ecology & Evolution* 5: 757–767.
- Dubayah RO, Sheldon SL, Clark DB, Hofton MA, Blair JB, Hurtt GC, Chazdon RL. 2010. Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *Journal of Geophysical Research: Biogeosciences* 115: 1–17.
- Durgante FM, Higuchi N, Ohashi S, Householder JE, Lima AJN, Ishizuka M, Wittmann F, dos Santos J, Carneiro VMC, Xu X *et al.* 2023. Soil fertility and drought interact to determine large variations in wood production for a hyperdominant Amazonian tree species. *Frontiers in Forests and Global Change* 5: 1065645.
- Ehbrecht M, Seidel D, Annighöfer P, Kreft H, Köhler M, Zemp DC, Puettmann K, Nilus R, Babweteera F, Willim K *et al.* 2021. Global patterns and climatic controls of forest structural complexity. *Nature Communications* 12: 519.
- Enquist BJ, Abraham AJ, Harfoot MBJ, Malhi Y, Doughty CE. 2020. The megabiota are disproportionately important for biosphere functioning. *Nature Communications* 11: 699.
- Fauset S, Baker TR, Lewis SL, Feldpausch TR, Affum-Baffoe K, Foli EG, Hamer KC, Swaine MD. 2012. Drought-induced shifts in the floristic and functional composition of tropical forests in Ghana. *Ecology Letters* 15: 1120–1129.
- Feng Y, Negrón-Juárez RI, Romps DM, Chambers JQ. 2023. Amazon windthrow disturbances are likely to increase with storm frequency under global warming. *Nature Communications* 14: 101.
- Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37: 4302–4315.
- Flores BM, Montoya E, Sakschewski B, Nascimento N, Staal A, Betts RA, Levis C, Lapola DM, Esquivel-Muelbert A, Jakovac C *et al.* 2024. Critical transitions in the Amazon forest system. *Nature* 626: 555–564.
- ForestPlots.net, Blundo C, Carilla J, Grau R, Malizia A, Malizia L, Osinaga-Acosta O, Bird M, Bradford M, Catchpole D *et al.* 2021. Taking the pulse of Earth's tropical forests using networks of highly distributed plots. *Biological Conservation* 260: 108849.
- Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, Husak G, Rowland J, Harrison L, Hoell A *et al.* 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data* 2: 150066.
- García MN, Domingues TF, Oliveira RS, Costa FRC. 2023. The biogeography of embolism resistance across resource gradients in the Amazon. *Global Ecology and Biogeography* 32: 2199–2211.
- Genier R, Poggi J-M. 2020. *Random Forests with R*. Cham, Switzerland: Springer International Publishing.
- Gora EM, Bitzer PM, Burchfield JC, Gutierrez C, Yanoviak SP. 2021. The contributions of lightning to biomass turnover, gap formation and plant mortality in a tropical forest. *Ecology* 102: e03541.
- Gora EM, Burchfield JC, Muller-Landau HC, Bitzer PM, Yanoviak SP. 2020. Pan-tropical geography of lightning-caused disturbance and its implications for tropical forests. *Global Change Biology* 26: 5017–5026.
- Gora EM, Esquivel-Muelbert A. 2021. Implications of size-dependent tree mortality for tropical forest carbon dynamics. *Nature Plants* 7: 384–391.
- Gora EM, McGregor IR, Muller-Landau HC, Burchfield JC, Cushman KC, Rubio VE, Mori GB, Sullivan MJP, Chmielewski MW, Esquivel-Muelbert A. 2025. Storms are an important driver of change in tropical forests. *Ecology Letters* 28: e70157.
- Gorgens EB, Motta AZ, Assis M, Nunes MH, Jackson T, Coomes D, Rosette J, Aragão LEO e C, Ometto JP. 2019. The giant trees of the Amazon basin. *Frontiers in Ecology and the Environment* 17: 373–374.
- Gorgens EB, Nunes MH, Jackson T, Coomes D, Keller M, Reis CR, Valbuena R, Rosette J, Almeida DRA, Gimenez B *et al.* 2021. Resource availability and disturbance shape maximum tree height across the Amazon. *Global Change Biology* 27: 177–189.
- Green J, Gentine P, Berry J, Zhang Y, Ciais P. 2020. Amazon rainforest photosynthesis increases in response to atmospheric dryness. *Science Advances* 6: eabb7232.
- Harris DJ, Ndolo Ebika ST, Sanz CM, Madingou MPN, Morgan DB. 2021. Large trees in tropical rain forests require big plots. *Plants, People, Planet* 3: 282–294.
- Hengl T, Mendes de Jesus J, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, Blagotić A, Shangquan W, Wright MN, Geng X, Bauer-Marschallinger B *et al.* 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* 12: e0169748.

- Hijmans R. 2025. *RASTER: geographic data analysis and modeling*. R package v.3.6-32. [WWW document] URL <https://github.com/rspatial/raster>.
- Hubbell SP, Foster RB, O'Brien ST, Harms KE, Condit R, Wechsler B, Wright SJ, de Lao SL. 1999. Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science* 283: 554–557.
- Ibanez T, Bauman D, Aiba S, Arsouze T, Bellingham PJ, Birkinshaw C, Birnbaum P, Curran TJ, DeWalt SJ, Dwyer J *et al.* 2024. Damage to tropical forests caused by cyclones is driven by wind speed but mediated by topographical exposure and tree characteristics. *Global Change Biology* 30: e17317.
- Jackson TD, Shenkin AF, Majalap N, Bin Jami J, Bin Sailim A, Reynolds G, Coomes DA, Chandler CJ, Boyd DS, Burt A *et al.* 2021. The mechanical stability of the world's tallest broadleaf trees. *Biotropica* 53: 110–120.
- Jiang M, Felzer BS, Nielsen UN, Medlyn BE, Fortin M. 2017. Biome-specific climatic space defined by temperature and precipitation predictability. *Global Ecology and Biogeography* 26: 1270–1282.
- Joswig JS, Wirth C, Schuman MC, Kattge J, Reu B, Wright IJ, Sippel SD, Rüger N, Richter R, Schaepman ME *et al.* 2022. Climatic and soil factors explain the two-dimensional spectrum of global plant trait variation. *Nature Ecology & Evolution* 6: 36–50.
- Jucker T, Ali A. 2023. Top canopy height and stem size variation enhance aboveground biomass across spatial scales in seasonal tropical forests. *Plants* 12: 1343.
- Jucker T, Hardwick SR, Both S, Elias DMO, Ewers RM, Milodowski DT, Swinfield T, Coomes DA. 2018. Canopy structure and topography jointly constrain the microclimate of human-modified tropical landscapes. *Global Change Biology* 24: 5243–5258.
- Kamimura K, Nanko K, Matsumoto A, Ueno S, Gardiner J, Gardiner B. 2022. Tree dynamic response and survival in a category-5 tropical cyclone: the case of super typhoon Trami. *Science Advances* 8: eabm7891.
- Kassambara A, Mundt F. 2020. *FACTOEXTRA: Extract and visualize the results of multivariate data analyses*. R package v.1.0.7. [WWW document] URL <https://github.com/kassambara/factoextra>.
- Koch GW, Sillett SC, Jennings GM, Davis SD. 2004. The limits to tree height. *Nature* 428: 851–854.
- Lang N, Jetz W, Schindler K, Wegner JD. 2023. A high-resolution canopy height model of the Earth. *Nature Ecology & Evolution* 7: 1778–1789.
- Larjavaara M. 2014. The world's tallest trees grow in thermally similar climates. *New Phytologist* 202: 344–349.
- Lewis GP, Siqueira GS, Banks H, Bruneau A. 2017. The majestic canopy-emergent genus *Dinizia* (Leguminosae: Caesalpinioideae), including a new species endemic to the Brazilian state of Espírito Santo. *Kew Bulletin* 72: 48.
- Liang J, Gamarra JGP, Picard N, Zhou M, Pijanowski B, Jacobs DF, Reich PB, Crowther TW, Nabuurs GJ, de-Miguel S *et al.* 2022. Co-limitation towards lower latitudes shapes global forest diversity gradients. *Nature Ecology & Evolution* 6: 1423–1437.
- de Lima RB, Gorgens EB, Batista APB, da Silva DAS, de Oliveira CP, Sousa CSC. 2022. Diversity and big tree patterns in the Brazilian Amazon. *Diversity* 14: 503.
- Lindenmayer DB. 2017. Conserving large old trees as small natural features. *Biological Conservation* 211: 51–59.
- Lindenmayer DB, Laurance WF. 2017. The ecology, distribution, conservation and management of large old trees: ecology and management of large old trees. *Biological Reviews* 92: 1434–1458.
- Loubota Panzou GJ, Fayolle A, Feldpausch TR, Ligt G, Doucet J-L, Forni E, Zombo I, Mazengue M, Loumeto J-J, Gourlet-Fleury S. 2018. What controls local-scale aboveground biomass variation in central Africa? Testing structural, composition and architectural attributes. *Forest Ecology and Management* 429: 570–578.
- Lutz JA, Furniss TJ, Johnson DJ, Davies SJ, Allen D, Alonso A, Anderson-Teixeira KJ, Andrade A, Baltzer J, Becker KML *et al.* 2018. Global importance of large-diameter trees. *Global Ecology and Biogeography* 27: 849–864.
- Madrigal-González J, Calatayud J, Ballesteros-Cánovas JA, Escudero A, Cayuela L, Marqués L, Rueda M, Ruiz-Benito P, Herrero A, Aponte C *et al.* 2023. Global patterns of tree density are contingent upon local determinants in the world's natural forests. *Communications Biology* 6: 47.
- Marra DM, Chambers JQ, Higuchi N, Trumbore SE, Ribeiro GHM, dos Santos J, Negrón-Juárez RI, Reu B, Wirth C. 2014. Large-scale wind disturbances promote tree diversity in a central Amazon Forest. *PLoS ONE* 9: e103711.
- Mattos CRC, Hirota M, Oliveira RS, Flores BM, Miguez-Macho G, Pokhrel Y, Fan Y. 2023. Double stress of waterlogging and drought drives forest–savanna coexistence. *Proceedings of the National Academy of Sciences, USA* 120: e2301255120.
- McDowell N, Allen CD, Anderson-Teixeira K, Brando P, Brienen R, Chambers J, Christoffersen B, Davies S, Doughty C, Duque A *et al.* 2018. Drivers and mechanisms of tree mortality in moist tropical forests. *New Phytologist* 219: 851–869.
- Meinzer FC, Clearwater MJ, Goldstein G. 2001. Water transport in trees: current perspectives, new insights and some controversies. *Environmental and Experimental Botany* 45: 239–262.
- Mills MB, Malhi Y, Ewers RM, Kho LK, Teh YA, Both S, Burslem DFRP, Majalap N, Nilus R, Huaraca Huasco W *et al.* 2023. Tropical forests post-logging are a persistent net carbon source to the atmosphere. *Proceedings of the National Academy of Sciences, USA* 120: e2214462120.
- Morrone J. 2014. Biogeographical regionalisation of the neotropical region. *Zootaxa* 3782: 1–110.
- Muller-Landau HC, Cushman KC, Arroyo EE, Martinez Cano I, Anderson-Teixeira KJ, Backiel B. 2021. Patterns and mechanisms of spatial variation in tropical forest productivity, woody residence time, and biomass. *New Phytologist* 229: 3065–3087.
- Oliveira RS, Costa FRC, van Baalen E, de Jonge A, Bittencourt PR, Almanza Y, Barros F d V, Cordoba EC, Fagundes MV, Garcia S *et al.* 2019. Embolism resistance drives the distribution of Amazonian rainforest tree species along hydro-topographic gradients. *New Phytologist* 221: 1457–1465.
- Ometto JP, Gorgens EB, De Souza Pereira FR, Sato L, De Assis MLR, Cantinho R, Longo M, Jacon AD, Keller M. 2023. A biomass map of the Brazilian Amazon from multisource remote sensing. *Scientific Data* 10: 668.
- Pebesma E, Bivand R, Racine E, Sumner M, Cook I, Keitt T, Lovelace R, Wickham H, Ooms J, Müller K. 2024. sf: simple features for R. Standardized support for spatial vector data. *The R Journal* 10: 439–446.
- Pinho BX, Peres CA, Leal IR, Tabarelli M. 2020. Critical role and collapse of tropical mega-trees: a key global resource. In: *Advances in ecological research*. Academic Press, Elsevier, 253–294.
- Piovesan G, Biondi F. 2021. On tree longevity. *New Phytologist* 231: 1318–1337.
- Quesada CA, Lloyd J, Anderson LO, Fyllas NM, Schwarz M, Czimczik CI. 2011. Soils of Amazonia with particular reference to the RAINFOR sites. *Biogeosciences* 8: 1415–1440.
- R Core Team. 2024. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. [WWW document] URL <https://www.R-project.org/>.
- Reis CR, Jackson TD, Gorgens EB, Dalagnol R, Jucker T, Nunes MH, Ometto JP, Aragão LEOC, Rodriguez LCE, Coomes DA. 2022. Forest disturbance and growth processes are reflected in the geographical distribution of large canopy gaps across the Brazilian Amazon. *Journal of Ecology* 110: 2971–2983.
- Rifai SW, Urquiza Muñoz JD, Negrón-Juárez RI, Ramírez Arévalo FR, Tello-Espinoza R, Vanderwel MC, Lichstein JW, Chambers JQ, Bohlman SA. 2016. Landscape-scale consequences of differential tree mortality from catastrophic wind disturbance in the Amazon. *Ecological Applications* 26: 2225–2237.
- Ripley B, Venables B, Bates DM, Firth D. 2022. *Modern applied statistics with S, Fourth edition*. New York: Springer. ISBN: 0-387-95457-0. [WWW document] URL <https://www.stats.ox.ac.uk/pub/MASS4/>.
- Sanchez-Martinez P, Dexter KG, Draper FC, Baraloto C, Leão do Amaral I, de Souza Coelho L, de Almeida Matos FD, de Andrade Lima Filho D, Salomão RP, Wittmann F *et al.* 2025. Phylogenetic conservatism in the relationship between functional and demographic characteristics in Amazon tree taxa. *Functional Ecology* 39: 181–198.
- Schöngart J, Piedade MTF, Ludwigshausen S, Horna V, Worbes M. 2002. Phenology and stem-growth periodicity of tree species in Amazonian floodplain forests. *Journal of Tropical Ecology* 18: 581–597.
- Seidl R, Thom D, Kautz M, Martin-Benito D, Peltoniemi M, Vacchiano G, Wild J, Ascoli D, Petr M, Honkaniemi J *et al.* 2017. Forest disturbances under climate change. *Nature Climate Change* 7: 395–402.

- Skiadaresis G, Schwarz J, Stahl K, Bauhus J. 2021. Groundwater extraction reduces tree vitality, growth and xylem hydraulic capacity in *Quercus robur* during and after drought events. *Scientific Reports* 11: 5149.
- Silva CA, Hudak AT, Vierling LA, Valbuena R, Cardil A, Mohan M, de Almeida DR, Broadbent EN, Almeyda Zambrano AM, Wilkinson B *et al.* 2022. TREETOP: a shiny-based application and R package for extracting forest information from LiDAR data for ecologists and conservationists. *Methods in Ecology and Evolution* 13: 1164–1176.
- Slik F, Pinho B, Griffith D, Webb E, Raghubanshi AS, Quaresma AC, Cuni Sanchez A, Sultana A, Souza AF, Ensslin A *et al.* 2024. Wind dispersed tree species have greater maximum height. *Global Ecology and Biogeography* 33: e13878.
- Slik JWF, Paoli G, McGuire K, Amaral I, Barroso J, Bastian M, Blanc L, Bongers F, Boundja P, Clark C *et al.* 2013. Large trees drive forest aboveground biomass variation in moist lowland forests across the tropics: large trees and tropical forest biomass. *Global Ecology and Biogeography* 22: 1261–1271.
- Slik JWF, Raes N, Aiba S-I, Brearley FQ, Cannon CH, Meijaard E, Nagamasu H, Nilus R, Paoli G, Poulsen AD *et al.* 2009. Environmental correlates for tropical tree diversity and distribution patterns in Borneo. *Diversity and Distributions* 15: 523–532.
- Spanner GC, Gimenez BO, Wright CL, Menezes VS, Newman BD, Collins AD, Jardine KJ, Negrón-Juárez RI, Lima AJN, Rodrigues JR *et al.* 2022. Dry season transpiration and soil water dynamics in the central Amazon. *Frontiers in Plant Science* 13: 825097.
- ter Steege H, Pitman N, Sabatier D, Baraloto C, Salomão R, Guevara Andino J, Phillips O, Castilho C, Magnusson W, Molino J-F *et al.* 2013. Hyperdominance in the Amazonian Tree Flora. *Science* 342: 1243092.
- Stephenson NL. 1990. Climatic control of vegetation distribution: the role of the water balance. *The American Naturalist* 135: 649–670.
- Sullivan M, Lewis S, Affum-Baffoe K, Castilho C, Costa F, cuni sanchez A, Ewango C, Hubau W, Marimon B, Monteagudo A *et al.* 2020. Long-term thermal sensitivity of earth's tropical forests. *Science* 368: 869–874.
- Ter Steege H, Pitman NCA, Do Amaral IL, De Souza Coelho L, De Almeida Matos FD, De Andrade Lima Filho D, Salomão RP, Wittmann F, Castilho CV, Guevara JE *et al.* 2023. Mapping density, diversity and species-richness of the Amazon tree flora. *Communications Biology* 6: 1130.
- Trenberth KE, Dai A, van der Schrier G, Jones PD, Barichivich J, Briffa KR, Sheffield J. 2014. Global warming and changes in drought. *Nature Climate Change* 4: 17–22.
- Tuanmu M-N, Jetz W. 2015. A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modelling. *Global Ecology and Biogeography* 24: 1329–1339.
- Venter M, Dwyer J, Dieleman W, Ramachandra A, Gillieson D, Laurance S, Cernusak LA, Beehler B, Jensen R, Bird MI. 2017. Optimal climate for large trees at high elevations drives patterns of biomass in remote forests of Papua New Guinea. *Global Change Biology* 23: 4873–4883.
- Wagner FH, Hérault B, Bonal D, Stahl C, Anderson LO, Baker TR, Becker GS, Bееckman H, Boanerges Souza D, Botosso PC *et al.* 2016. Climate seasonality limits leaf carbon assimilation and wood productivity in tropical forests. *Biogeosciences* 13: 2537–2562.
- Wright MN, Ziegler A. 2017. RANGER: a fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software* 77: 1–17.
- Wright SJ, Muller-Landau H, Condit R, Hubbell S. 2003. Gap-dependent recruitment, realized vital rates, and size distributions of tropical trees. *Ecology* 84: 3174–3185.
- Wunderle JM. 1997. The role of animal seed dispersal in accelerating native forest regeneration on degraded tropical lands. *Forest Ecology and Management* 99: 223–235.
- Yanoviak SP, Gora EM, Bitzer PM, Burchfield JC, Muller-Landau HC, Detto M, Paton S, Hubbell SP. 2020. Lightning is a major cause of large tree mortality in a lowland neotropical forest. *New Phytologist* 225: 1936–1944.
- Zoletto B, Van Der Sande MT, Sleen P, Sheil D. 2023. Lightning scars on tropical trees—evidence and opportunities. *Ecology and Evolution* 13: e10210.
- Zuleta D, Duque A, Cardenas D, Muller-Landau HC, Davies SJ. 2017. Drought-induced mortality patterns and rapid biomass recovery in a terra firme forest in the Colombian Amazon. *Ecology* 98: 2538–2546.
- Zuleta D, Russo S, Barona-Colmenares A, Barreto-Silva J, Cárdenas D, Castaño N, Davies S, Detto M, Sua S, Turner B *et al.* 2020. Importance of topography for tree species habitat distributions in a terra firme forest in the Colombian Amazon. *Plant and Soil* 450: 133–149.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Dataset S1 Complete dataset with the name and metrics of environmental and geographic variables, tree density, and topographic metrics generated by airborne LiDAR in 900 randomly distributed transects in the Brazilian Amazon biome.

Fig S1 Methodological workflow for mapping giant trees in the Amazon, from LiDAR data collection to Random Forest modeling and spatial density maps.

Fig. S2 Spatial distribution of 16 environmental predictors across the Brazilian Amazon used in the Random Forest model.

Fig. S3 Pairwise scatterplots, correlations, and distributions of tall-tree density and environmental predictors.

Fig. S4 Model performance and residual diagnostics for tall-tree density predictions across 900 Amazon transects.

Fig. S5 Importance of environmental predictors for tall-tree density estimation based on IncNodePurity and %IncMSE metrics.

Fig. S6 Partial dependence plots showing the influence of environmental variables on modeled tall-tree density.

Fig. S7 Principal component analysis of environmental variables across Amazonian biogeographic provinces.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

Disclaimer: The New Phytologist Foundation remains neutral with regard to jurisdictional claims in maps and in any institutional affiliations.