Giants of the Amazon: How does environmental variation drive the diversity patterns of large trees?

Antônio Carlos da Silva Zanzini63 | Clarissa Rosa5 | Wagner Tadeu Oliveira64 | Lucas Pereira Zanzini65 | Ricardo José Silva66 | Ana Luisa Mangabeira Albernaz67

1Laboratório de Manejo Florestal, Universidade do Estado do Amapá, Macapá, Brazil
2Departamento de Engenharia Florestal, Universidade Federal do Vales do Jequitinhonha e Mucuri, Diamantina, Brazil
3Instituto Federal de Educação Ciência e Tecnologia do Amapá, Laranjal do Jari, Brazil
4Laboratório de Manejo de Florestas Naturais “José Serafim Feitoza Ferraz”, Departamento de Ciência Florestal, Universidade Federal Rural de Pernambuco, Recife, Brazil
5Coordenação de Pesquisas em Biodiversidade, Instituto Nacional de Pesquisas da Amazônia, Manaus, Brazil
6Tropical Botany, Naturalis Biodiversity Center, Leiden, The Netherlands
7Programa de Pós-Graduação em Botânica Aplicada, Departamento de Biologia, Universidade Federal de Lavras, Lavras, Brazil
8Instituto Nacional de Pesquisas da Amazônia, Av. André Araújo, Manaus, Brazil
9Programa de Pós-Graduação em Ecologia, Instituto Nacional de Pesquisas da Amazônia, Manaus, Brazil
10Programa de Pós-Graduação em Botânica, Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus, Brazil
11Instituto Nacional de Pesquisas da Amazônia—Coordenação de Pesquisas em Silvicultura Tropical, Manaus, Brazil
12Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford, UK
13Lancaster Environment Centre, Lancaster University, Lancaster, UK
14Embrapa Amazônia Oriental, Rede Amazônia Sustentável, Belém, Brazil
15Programa de Pós-Graduação em Ecologia, Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, Brazil
16Embrapa, Boa Vista, Brazil
17Embrapa, Brasília, Brazil
18Promotoria de Justiça do Meio Ambiente, Conflitos Agrários, Ministério Público, Macapá, Brazil
19Programa de Pós-Graduação em Biodiversidade e Biotecnologia—Rede Blonorte, Universidade Federal do Amapá, Macapá, Brazil
20School of Geography, Earth and Environmental Science, University of Birmingham, Birmingham, UK
21Empresa Brasileira de Pesquisa Agropecuária, Brasília, Brazil
22Universidade Federal do Amapá, Macapá, Brazil
23Laboratório de Sementes Florestais, Universidade do Estado do Amapá, Rua Presidente Vargas, Macapá, Brazil
24Universidade do Estado de Mato Grosso, Nova Xavantina, Brazil
25Faculdade de Ciências Agrárias, Biológicas e Sociais Aplicadas, Universidade do Estado de Mato Grosso, Nova Xavantina, Brazil
26Universidade do Estado de Mato Grosso, Campus de Nova Xavantina, Nova Xavantina, Brazil
27Programa de Pós-Graduação em Ambiente e Sistemas deProdução Agrícola, Universidade do Estado de Mato Grosso, Tangará da Serra, Brazil
28Museu Paraense Emílio Goeldi, Belém, Brazil
29Departamento de Biologia, Universidade Federal do Amazonas, Manaus, Brazil
30Department of Forestry, University of Brasilia, Brasilia, Brazil
31Department of Forest Sciences, “Luiz de Queiroz” College of Agriculture, University of São Paulo (USP/ESALQ), Piracicaba, Brazil
32Laboratório de Botânica e Ecologia Vegetal, Universidade Federal do Acre, Rio Branco, Brazil
33Universidade Federal de Mato Grosso, Cuiabá, Brazil
34Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
35School of Geography, University of Leeds, Leeds, UK
36University of Exeter, Exeter, UK
37Laboratório de Ecologia Aplicada à Conservação, Universidade Estadual de Santa Cruz, Salobrinho, Brazil
38National, Botanical Garden, Florida, USA
39Program of Environmental and Forest Science, Federal University of Mato Grosso, Cuiabá, Brazil
40Faculdade de Ciências Agrárias e Biológicas, Unemat, Brazil
41Museu Universitário, Universidade Federal do Acre, Rio Branco, Brazil
42Universidade Federal de Roraima (UFRR), Boa Vista, Brazil
43Serviço Florestal Brasileiro, Santarém, Brazil
44CNHS, Universidade Federal de Mato Grosso, Sinop, Brazil
45Universidade Federal de Mato Grosso (UFMT), Núcleo de Estudos da Biodiversidade da Amazônia Mato-grossense, Sinop, Brazil
46Universidade Federal do Acre, Rio Branco, Brazil
47Universidade Federal de Rondônia, Porto Velho, Brazil
48Instituto Federal de Educação Ciência e Tecnologia de Rondônia, Porto Velho, Brazil
49James Cook University, Douglas, Queensland, Australia
Abstract

For more than three decades, major efforts in sampling and analyzing tree diversity in South America have focused almost exclusively on trees with stems of at least 10 and 2.5 cm diameter, showing highest species diversity in the wetter western and northern Amazon forests. By contrast, little attention has been paid to patterns and drivers of diversity in the largest canopy and emergent trees, which is surprising given these have dominant ecological functions. Here, we use a machine learning approach to quantify the importance of environmental factors and apply it to generate spatial predictions of the species diversity of all trees (dbh ≥ 10 cm) and for very large trees (dbh ≥ 70 cm) using data from 243 forest plots (108,450 trees and 2832 species) distributed across different forest types and biogeographic regions of the Brazilian Amazon. The diversity of large trees and of all trees was significantly associated with three environmental factors, but in contrasting ways across regions and forest types. Environmental variables associated with disturbances, for example, the lightning flash rate and wind speed, as well as the fraction of photosynthetically active radiation, tend to govern the diversity of large trees. Upland rainforests in the Guiana Shield and Roraima regions had a high diversity of large trees. By contrast, variables associated with resources tend to govern tree diversity in general. Places such as the province of Imeri and the northern portion of the province of Madeira stand out for their high diversity of species in general. Climatic and topographic stability and functional adaptation mechanisms promote ideal conditions for species diversity. Finally, we mapped general patterns of tree species diversity in the Brazilian Amazon, which differ substantially depending on size class.

KEYWORDS

big trees, diversity map, forest ecology, forest inventory, remote sensing, species richness
INTRODUCTION

Understanding how the diversity of tree species is distributed is a great challenge, especially in tropical forests, a hyperdiverse biome. Once multiple plot inventories are combined, then they can reveal key large-scale patterning in community diversity, for instance, that some Amazon forests are extremely diverse (Gentry, 1988), or that wet Neotropical climates support high arboreal diversity (Clinebell et al., 1995). Multiple, recent initiatives that combine forest inventories have allowed further analysis of ecological and spatial variation over large scale (e.g., Bastin et al., 2018; Crowther et al., 2015; ForestPlots.net et al., 2021; Silk et al., 2015). These studies have confirmed that parts of South America and particularly western Amazonia, as well as the Andean foothills, are global epicenters for plant diversity and trees in particular (Cazzolla Gatti et al., 2022; Liang et al., 2022; Sabatini et al., 2022; Ter Steege et al., 2013).

More generally, it has long been known that the diversity of tree species follows latitudinal variations, consistent with the metabolic theory of biodiversity (Liang et al., 2022), or is affected by grain-dependent effects of both environment and biogeographic regions (Keil & Chase, 2019). In the Amazon basin tree diversity at the one-hectare scale are driven by gradients of precipitation and soil (Esquivel Muelbert et al., 2016; Ter Steege et al., 2003). However, most work focuses on trees with stems ≥10cm in diameter, and sometimes as small as 2.5cm diameter or less (Draper et al., 2021; Sabatini et al., 2022), meaning that diversity patterns among relatively small trees tend to dominate our understanding of tropical forest diversity. By contrast, few studies have attempted to analyze the largest trees at larger scales, let alone explore how environmental factors influence their diversity.

Large trees have a dominant influence on the aboveground biomass, productivity, and sustainable use of tropical forests (Ali et al., 2022; Fauset et al., 2015; Kearsley et al., 2019; Lindenmayer, 2017) and they are considered indicators of stable ecosystems (de Lima et al., 2022; Lindenmayer & Laurance, 2016; Roswell et al., 2021; Silk et al., 2013). However, even for the community of all tree species of all sizes, spatial diversity models are still scarce in Amazonia (Ter Steege et al., 2003, 2013), and primarily regional spatial diversity models for large trees are poorly studied. Thus, while we may expect divergent diversity patterns between small and large tree species, but have lacked the data or analyses to explore this in detail, and thus, key gaps remain concerning the main environmental factors driving diversity in the Amazon basin.

There are a few reasons to expect the diversity patterns of large trees to differ from those of smaller trees: (1) Areas of occupancy: Large trees can occupy larger areas than small trees, which means they can support more species in their habitat (Liu et al., 2019); (2) Habitat heterogeneity: Large trees can create different microhabitats in their structure, such as ecological niches at different tree heights, allowing different species to occupy specific areas (Mendieta-Leiva et al., 2022); (3) Habitat connectivity: Large trees can also serve as habitat connectors, linking different areas and allowing species to move between them. This can increase species diversity in an ecosystem (Craven et al., 2016; Keeley et al., 2021). Positive relationships between plant diversity and environmental factors are common in various vegetation types (Balvanera & Aguirre, 2006; Beck et al., 2011; Silk et al., 2009; Zheng et al., 2019) and ecological gradients (Maharjan et al., 2021; Song et al., 2021; Tolmos et al., 2022; Venter et al., 2017; Wittmann et al., 2006). The available remote sensing environmental data provide spatially refined information on landscape and vegetation heterogeneity over the Amazon basin that can be readily incorporated into spatial models to study diversity distribution patterns. For example, the maximum large tree height in different biogeographic regions in the Amazon is influenced by edaphic, climatic, and topographic variables (Gorgens et al., 2021).

These patterns might be expected to impact large tree diversity over space and/or time, depending on which factors are key—for example, the role of increased lightning occurrences and wind speed in limiting tree height suggests potential for concomitant changes in the diversity patterns of large trees (Bennett et al., 2015; Clark et al., 2019; Gora & Esquivel-Muelbert, 2021; Yanoviak et al., 2020).

The significant regional variation within the large Brazilian Amazon, and growing efforts to inventory tree plot biomass and species, favors the development of large area ecological and distribution models to understand how and which environmental factors influence the diversity of large and small trees. Here, using a multisite dataset from across the Amazon, we aim to explore how environmental factors influence patterns of total tree species diversity, and patterns of large trees, contributing to understanding how the forest is spatially structured and according to its biodiversity and environmental components.

MATERIALS AND METHODS

2.1 Study area

Individual tree data were obtained from 243 forest plots distributed across 65 sites in the Brazilian Amazon (Figure 1). Plot sizes and shape varied according to the objectives of each acquisition source. While the majority were standardized at 100×100m (10,000m²), others are 40×250m (10,000m²), 10×250m (2500m²), 20×200m (4000m²), and 20×50m (1000m²). Only plots with at least 90% of the trees identified genus level and at least 60% to at species level, were included in the analysis. Thus, we included 108,450 trees with a stem diameter of 1.3m from the ground (dbh—diameter at breast height) ≥10 cm in the analyses.

A subset of the data was created by filtering only trees with dbh ≥70 cm, highlighting the largest trees, as suggested by Harris et al. (2021). The choice of large trees with a diameter ≥70 cm can be consulted for more details in Appendix S3. This subset had 240 plots and 3863 trees. All plots contribute to the wider ForestPlots.net meta–network and database (https://www.foresplots.net/), which curates inventories conducted over the last 25 years (1997–2021) (ForestPlots.net et al., 2021). Our sites span different forest types and cover several biogeographic regions. To ensure the comparison of plots with different sample sizes, plot data were...
standardized to the 1 ha sampling effort (Bastin et al., 2018; Clark & Clark, 2000).

Taxonomic nomenclature was checked using the Taxonomic Name Resolution Service (https://tnrs.iplantcollaborative.org) and Missouri Botanical Garden (https://www.missouribotanicalgarden.org/). To verify any uncertainty in the correct name of the species, we used the taxonomic conference implemented in the flora package (Carvalho, 2020) for R software version 4.2.1 (R Core Team, 2022).

2.2 | Species diversity

Diversity was analyzed for all trees (dbh ≥ 10 cm) and for large trees only (dbh ≥ 70 cm), using the Fisher’s alpha-diversity index (α) and Shannon entropy diversity index calculated for each plot:

\[ S = \alpha \ln\left(1 + \frac{n}{\alpha}\right) \]  
\[ e^{H'} = \exp\left(-\sum_{i=1}^{S} \frac{P_i \times \ln(P_i)}{S}\right) \]

Tree alpha diversity, expressed as Fisher’s alpha (\(\alpha\)), is a measure which corrects for the sample size (Fisher et al., 1943). The exponential Shannon entropy index is a transformation of Shannon’s index that weights species in proportion to their frequency (Jost, 2006); \(n\) is the number of individuals. \(H'\) is the Shannon diversity index; \(S\) is the total number of species for a given plot; \(P_i\) is the proportion of the total number of individuals in the sample belonging to species \(i\). The Shannon index provides information on abundance and equability without favoring any species, as it counts all species according to their frequencies (Borges Silva et al., 2022). To complement the information on the diversity index, the richness resulting from the total number of species was also calculated for each plot. Richness and diversity are considered indicators of general biodiversity (Gastauer et al., 2021). All data were analyzed using the vegan (Oksanen et al., 2022) and BiodiversityR packages (Kindt, 2022) in R.

Rarefaction curves of cumulative individuals were created, one for each dataset, using the iNEXT package (Hsieh et al., 2022).

2.3 | Environmental variables

Sixteen candidate spatially explicit environmental variables were selected and grouped into three distinct categories: topographic, climatic, and edaphic (Table 1). To allow comparison of our results with those from other studies, we used the same variables as considered in the work developed by Gorgens et al. (2021). Data were cropped considering the limits of the Brazilian Amazon and re-sampled to a spatial resolution of 30 arc seconds (~1 km) when necessary.
Temperature, precipitation, and the 19 bioclimatic variables were downloaded from WorldClim version 2 (Fick & Hijmans, 2017). The average number of cloudless days throughout the year was obtained from the surface reflectance products of the MODIS (Moderate Resolution Imaging Spectroradiometer) sensor. The Terra product MOD09GA Version 6 was used, which provides an estimate of the spectral reflectance of the MODIS surface, corrected for atmospheric conditions.

The average number of days per year with precipitation above 20 mm has been calculated from the precipitation time series of the Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) dataset (Funk et al., 2015). Potential evapotranspiration was obtained from TerraClimate data, which combine climatological normals from WorldClim, the Climate Research Unit (CRU) Ts4.0, 55-year Japanese Reanalysis (JRA-55) data, and the Penman–Monteith methodology. Fraction of absorbed photosynthetically active radiation (fapar) was obtained from the calibrated and corrected Earth surface reflectance product of the Advanced Very High-Resolution Radiometer (AVHRR) of the National Oceanic and Atmospheric Administration (NOAA). The fapar layer provides information on the photosynthetic activity of plants (Baret et al., 2013).

Lightning frequency is associated with climatic events and tree mortality (Gora et al., 2020). This variable was obtained from the Lightning Imaging Sensor (LIS) instrument aboard the Tropical Rainfall Measuring Mission provided by the Global Hydrology Resource Center of NASA’s Earth Observation System (EOSDIS).

Wind speed data were made available from the maximum daily average wind speeds over the last 5 years produced by the fifth reanalysis of the global model (ERA5) of the European Center for Medium-Term Weather Forecasts (ECMWF). Two metrics of wind speeds were used: u-speed, which is the zonal speed (a component of the wind toward the east), and v-speed, which is the meridional speed (a component of the wind toward the north). Studies indicate that winds are related to disturbances that result in tree mortality in the Amazon (Marra et al., 2014; Rifai et al., 2016).

Edaphic variables were obtained from the SoilGrids based on the World Reference Base (WRB) and USDA classification systems (ca. 280 raster layers in total; Hengl et al., 2017). The layers of clay content (% fine particles <2 μm) and water content (% volumetric at field capacity at 30 cm), both with a spatial resolution of 250 m, were estimated from machine learning applied over a global profile.
compilation and soil layers (Hengl et al., 2017). All variables were processed using ArcMap 10.1 (ESRI, 2019), and analysis was performed using R software version 4.2.1 (R Core Team, 2022).

2.4 | Spatial modelling

Choosing the suitable algorithm for modelling depends on the problem and available data. Here we used the Random Forest algorithm to model the relationship between environmental variables and diversity using 500 random, regression–type decision trees. According to the approach carried out by Meyer and Pebesma (2022), the Random Forest (RF) algorithm is widely used for data analysis of plant diversity, mainly to understand patterns, due to several reasons: (1) it effectively handles high-dimensional data: rainforests contain a wide variety of species, which results in highly complex, high-dimensional data. RF can handle this complex data and produce accurate results; (2) it is a nonparametric algorithm: as the relationships between species in a rainforest can be highly nonlinear, a nonparametric algorithm such as RF is best suited to handle these data types; (3) Can handle missing data: according to Cai et al. (2023), data collected in tropical forests may need to be completed or have missing values. RF can handle this missing data without dropping or replacing it; (4) can handle overfitting: the presence of rare species in a tropical forest can lead to an overfitting problem, where the model is overfitted to the available data and RF is less prone to this problem due to its ensemble model nature, where many smaller models are combined to produce an overall result; (5) it is highly scalable: The RF algorithm is highly scalable and can be easily applied to larger datasets, which is an advantage in a rainforest environment where the number of species can be large.

For all these reasons, RF is widely used for data analysis in tropical forests to understand diversity patterns. However, this does not mean that other machine-learning algorithms cannot be used to analyze rainforest data or other data types. Choosing the suitable algorithm depends on the problem and available data.

The number of variables randomly sampled as candidates in each division was defined as the number of explanatory variables divided by 3 (default algorithm parameterization). Evaluation of the model’s accuracy was performed using the standard error of the estimate (RMSE) and coefficient of determination ($R^2$), considering the $k$-fold cross-validation ($k=15$). The importance of environmental variables was analyzed using marginal graphs, keeping the other variables constant at an average value. The importance of variables is measured by permuting variables in the model and measuring the increase in estimation error (MSE).

The RF model was applied to the environmental layers at the pixel level, for the entire Amazon region, through map algebra. The model parameters were applied independently to each pixel to produce the diversity maps for all trees and for the giant trees of the Amazon forest (Figure 2). Statistical analysis was applied in an R environment (R Core Team, 2022), using the MASS (Ripley et al., 2022) and RandomForest (Cutler & Wiener, 2022) packages.

3 | RESULTS

3.1 | Species diversity

Considering only the large trees (dbh ≥ 70 cm), 240 plots had 3863 individuals (~3.6% of the measured trees) grouped into 449 species (for every 8.6 trees, one large species was found). The lack of a plateau indicates that many additional species are likely to be discovered if sampling increases (Figure 3a). Considering all trees measured (dbh ≥ 10 cm), the rarefaction curve reached a plateau with approximately 25,000 individuals distributed in approximately 240 species in a total of 2832 species sampled (Figure 3a). Fisher’s alpha diversity of all trees showed a poor, but significant correlation ($r = .2$, $p < .001$) with the Fisher’s alpha diversity of large trees (Figure 3b). The median Fisher’s alpha-diversity index for all trees (101) was approximately 94% higher that for large trees (4).

3.2 | Spatial modelling and environmental factors

The fitted RF model showed a standard error of 0.69 (MAE = 3.22; RMSE = 4.81) for large trees and 0.65 (MAE = 29.14; RMSE = 41.7) for all trees. The variation explained by the RF models was markedly high for Fisher’s alpha diversity all trees ($R^2 = .71$) and for Fisher’s alpha diversity large trees ($R^2 = .85$). The graphical analysis of observed versus predicted diversity data shows positive associations, with a higher degree of adherence for all trees compared to large trees (Figure 4).

The 16 environmental variables considered in this study showed different degrees of importance when analyzing the association between large trees and trees in general (Figure S1). Four different degrees of association were observed: positive (diversity increases as the factor increases), negative (diversity decreases as the factor increases), paraboloid behavior with downward concavity (diversity initially increases but then decreases as the factor increases), and upward concave paraboloid behavior (diversity initially decreases but increases as the factor increases). Furthermore, these environmental variables can be divided into two groups: environmental variables associated with tree growth resources and those with disturbances. Although some variables can reflect both resources and disturbances, to facilitate the discussion, we here associate variables of this type with the most relevant of the two categories. The environmental variables associated with disturbances (uspeed, vspeed, lighting, days 20) were the most relevant for the diversity of large trees. The variables associated with tree growth (clearDays, pseasoon, pweattest, elevation, pannual, pet, tannual, and fapar) were the most relevant for the diversity of trees in general (Figures 5 and 6).

Among the variables associated with disturbance, wind speed, lightning occurrence rate, the number of days with precipitation above 20 mm, and the maximum temperature stand out. These variables are generally directly related to tree mortality and were more relevant to the diversity of large trees than to the diversity of all trees (Figure 5b–e). While four of these variables above are among the six most relevant for the diversity of giant trees, only three are
FIGURE 2  Data processing and analysis flowchart: (a) plot network organization; (b) stacking of environmental spatial layers; (c) crossing the field plots and the stack of environmental variables; (d) for each forest typology and biogeographic region, we submitted plot records to identify the least collinear collection of covariates; (e) fitted of the Random Forest model for diversity; (f) a top-tier predictive model was selected or created by averaging the model using cross-validation; (g) diversity estimation for the entire Amazon at the pixel level; and (h) distribution of diversity for all trees and for large trees. [Colour figure can be viewed at wileyonlinelibrary.com]
among the six most relevant considering all trees. Climatic metrics derived from precipitation (pannual and pseason) are of greater predictive importance for diversity indices analyzed and RF models. Among the environmental variables associated with resources are photosynthetically active radiation, precipitation metrics (annual average, wettest month, and coefficient of variation), temperature metrics (annual average and coefficient of variation), number of days without clouds, evapotranspiration, elevation above sea level. These variables were the most strongly associated with all tree diversity (Figure 6a, c, e-g, l-n).

Average annual precipitation, the number of annual days with precipitation greater than 20 mm, evapotranspiration, precipitation of the wettest month, elevation, and soil clay content were positively associated with the Fisher’s alpha diversity of big trees. However, the factors lightning occurrence rate, component meridional wind speed, zonal wind speed, precipitation variation coefficient (pseason), mean annual temperature, number of annual days without cloud, and maximum annual temperature showed a negative association. The factors photosynthetically active radiation, coefficient of temperature variation, and soil water content showed a paraboloid association with Fisher’s alpha diversity, with upward concavity (Figure 5).

The Fisher’s alpha diversity of large trees increased substantially with average annual precipitation, elevation, and a fraction of photosynthetically active radiation with a threshold above 80% and number of days with precipitation greater than 40 mm. In contrast, increased wind speed and lightning occurrence rate are associated with lower Fisher’s alpha diversity of large trees (Figure 5).

Considering the Fisher’s alpha diversity of all trees, the number of annual days with precipitation greater than 20 mm, lightning occurrence rate, average annual precipitation, elevation, and photosynthetically active radiation, and average annual temperature showed a positive association with diversity. In contrast, the factors coefficient of variation of precipitation, number of annual days without cloud, evapotranspiration, and soil water content showed a negative association. Precipitation of the wettest month, zonal wind speed, meridional wind speed, maximum temperature annual, coefficient of temperature variation, and soil clay content showed a paraboloid association with upward concavity with the total tree diversity (Figure 6).

The highest Fisher’s alpha-diversity values for large trees were higher in terra-firme forest in the northern portion of the Guiana Shield biogeographic province and the eastern portion of the Roraima biogeographic province (about −54 to −51° longitude, Figure 7). Fisher’s alpha...
of large trees in this region exceeded 40. The predictive maps also reveal many areas with intermediate alpha-diversity values for large trees (10 < 20) in areas west of Imerí, Madeira, and north of Xingu-Tapajós and Pantepui. Large areas in central Amazonia between Xingu-Tapajós and Madeira and its tributaries are dominated by forests with low alpha diversity of large trees (<5). The central and eastern provinces (Madeira, Xingu-Tapajós, and Pará) are dominated by open ombrophilous forests (also known as rainforest or tropical rainforest, is a type of ecosystem that develops in tropical and subtropical regions with high rainfall throughout the year). The areas of ecological tension or transition belt showed the lowest diversity values for all trees and large trees (Figure 7). These regions coincide with the arc of deforestation.

The relationships between environmental factors and the Shannon diversity index produced trends similar to those based on the Fisher's alpha for both large (Figure S4) and all trees (Figure S5). However, a spatial appreciation for this index suggests a significant divergence in the potential distribution of the diversity of large species and the set of all analyzed species (Figure S6). The greatest Shannon diversity for all trees was observed in the Imerí province, northwest of the Madeira and Pantepui regions, closely associated with forests with high rainfall. The supplementary information results can reveal more details of the correlation analysis for the estimated diversity metrics versus the inventoried plot size (Figure S2) and the number of large trees (Figure S3).

4 | DISCUSSION

4.1 | Species diversity and environmental factors

In this study, we asked what factors explain the diversity of large trees (dbh ≥ 70 cm) in the Amazon and how these patterns differ from those affecting smaller trees (dbh ≥ 10 cm). We show that the diversity of large trees is mainly determined by environmental factors linked to the intensity of disturbances. In contrast, the diversity of smaller trees is more affected by factors that control productivity. Higher diversity values for large trees were found in the extreme Northeast of Amazonia (Amapá), while the highest values for smaller trees were found in the Northwest.

We found that the diversity of large trees decreased with increasing meridional wind speed and lightning occurrence rate. Large trees that emerge above the forest canopy are most susceptible to lightning strikes and toppling by wind due to the greater exposure to these factors (Bennett et al., 2015; Clark et al., 2019; Gora & Esquivel-Muelbert, 2021; Yanoviak et al., 2020). Recent research shows that lightning has been a prominent cause of crown damage in the tropics, often leading to tree mortality (Gora et al., 2017, 2021; Marra et al., 2014).

The highest diversity of large trees was found in terra-firme areas (well-drained rainforest including on hills) in the Guiana Shield and Roraima regions. These regions also have the greatest potential for the growth of very tall trees (Gorgens et al., 2021). This strong relationship between well-drained upland areas and large tree diversity has also been observed in other forests in the tropics and has been associated with ecologically stable environmental conditions for tree establishment and growth (Duivenvoorden, 1995; Fagua et al., 2021; Slik et al., 2009; Ter Steege et al., 2003; Venter et al., 2017).

Recurrent increases in wind speed can significantly modify the structure and composition of forests. For example, the predicted change in tropical cyclone regimes due to climate change can strongly impact forests in the tropics (Hall et al., 2020; Rau et al., 2022; Rifai et al., 2016). Our largely equatorial forests are not affected by cyclones but strong winds associated with convectional storm downbursts, and also cold fronts in southern and western Amazonia, may drive significant mortality for large trees. Since most trees in the forest are not
large (dbh ≥ 70 cm), lightning and wind are much less important for the set of all trees (dbh ≥ 10 cm). Furthermore, wind and lightning are essential factors in the diversity and dynamics of species since they are considered to be agents of natural disturbance in forests and can act by selecting individuals (Hall et al., 2020; Kling & Ackerly, 2021; Laurance & Curran, 2008; Mitchell, 2013; White & Pickett, 1985).

Our results show higher diversity for trees in general (dbh ≥ 10 cm) in low-altitude forest regions with permanent, seasonally, or transitional flooding along the Imeri province (Figure 7). This region is dominated by extensive river systems, floodplains, and lowlands. Depending on their proximity to rivers and their sediment load and topography, forests contain great diversity. These results are possibly expected because these areas are among the lowest densities of human population in the Amazon region, and roads are absent, giving them a particular value in conservation. Areas that are low or humid have trees with smaller crowns (Barbier et al., 2010) but with high species heterogeneity, mainly because they are ecotones or forest transition areas (Amaral et al., 2021; Marques et al., 2011; Silveira et al., 2016). These wetlands are also hotspots for endemic flora and fauna (Berde et al., 2020; Morrone, 2014) and contribute significantly to regional and global biodiversity (Junk et al., 2006; Liang et al., 2022). Similar results have been found in floodplain areas at the biome scale (Duivenvoorden, 1995; Moraes et al., 2021; Silman et al., 2007; Ter Steege et al., 2013; Wittmann et al., 2006). These diversity patterns are also observed in the context of source–sink effects along topographic gradients (Colville et al., 2020; Maharjan et al., 2021; Nanda et al., 2021; Syfert et al., 2018; Zeleny et al., 2010) and along latitudinal gradients (Kraft et al., 2011; Roy et al., 2007; Slik et al., 2009). This would mean that on evolutionary timescales, these ecological mechanisms of lowland forests are essential for maintaining tree diversity in tropical mountain forest (Coelho de Souza et al., 2019; Slik et al., 2009).

The three variables related to precipitation were among the most important associated with the diversity of all trees (dbh ≥ 10 cm).
FIGURE 6  Graphs of the marginal effect of each environmental variable on the diversity of all trees (dbh ≥ 10 cm) in the Random Forest model. The environmental factors are shown in order of decreasing importance.

FIGURE 7  Fisher's alpha-diversity distribution for large trees (dbh ≥ 70 cm) and diversity distribution considering all trees (dbh ≥ 10 cm) estimated by the RF model for the Brazilian Amazon. [Colour figure can be viewed at wileyonlinelibrary.com]
Rainfall in Amazonia averages approximately 2300 mm year\(^{-1}\), and is driven in part by the austral monsoons and generates a well-defined environmental gradient. In areas near the Brazilian borders with Colombia and Venezuela, the annual total rainfall can reach 3500 mm (Figueroa & Nobre, 1990). Here, there is no defined dry period, with potential to flood some forests most of the year (Householder et al., 2021; Junk et al., 2018). These high precipitation values are associated with less seasonal equatorial climates, and near the Andes mountains, rainfall is further enhanced by the orographic rise of moisture transported by the east trade winds of the Intertropical Convergence Zone (ITCZ).

Various studies report changes in tree species composition along tropical altitudinal gradients (Guimaraes et al., 2021; Tolmos et al., 2022; Venter et al., 2017). In our analyses, lower rainfall and lower elevation (<50 m a.s.l.) in certain areas are associated with decreases in the diversity indices both for large trees and for the set with all trees (Figures 5–7). In these environments, the lower areas are characterized by communities of taxa that have difficulty growing to reach large dimensions (Da Cunha et al., 2021).

The annual number of days without clouds (clearDays year\(^{-1}\)) and the seasonality of precipitation (pseason %) were essential predictors to explain the patterns of diversity of all trees (Figure 6). However, they were of only intermediate importance for the diversity of large trees (Figure 5). An increase in cloudless days is associated with direct solar radiation (Barkhordarian et al., 2019; Gorgens et al., 2021), favoring photosynthetic activity for the species pool up to a thermal tolerance threshold (Green et al., 2020; Sullivan et al., 2020). Although atmospheric dryness, caused by the combination of these factors, can increase photosynthetic activity in the Amazon forest when water supply is abundant (Green et al., 2020), reductions in diversity in southern Amazonia may result from the long exposure of plants to high solar radiation and associated extended periods of seasonal water deficit. Bennett et al. (2015) found that droughts consistently negatively impacted larger trees’ growth and mortality rates on a global scale. Other studies report that the number of clear days and the seasonality of precipitation are determinant factors for floristic diversity patterns at regional and continental scales (Brujinzeel et al., 2011; Ehbrecht et al., 2021; Jiang et al., 2017; Wagner et al., 2016; Wilson & Jetz, 2016).

Elevation, here treated as a proxy for distinguishing terra-firme from floodplains, had a significant positive association with the diversity of both trees in general and large trees. Studies in tropical forests of Borneo (Slik et al., 2009), tropical lowlands of South America (Neves et al., 2020), Tanzania (Tallents et al., 2005), Colombia (Fagua et al., 2021; Tolmos et al., 2022), and Papua New Guinea (Venter et al., 2017) have also reported the importance of altitude in species diversity and composition.

### 4.2 Study limitations

In the meta-analysis performed, we explored a variety of data sources and analytical tools that can be applied to develop predictive maps that incorporate the observed spatial pattern of diversity in biome-scale landscapes, forest types, and biogeographic regions. However, available data still lacks good coverage or resolution for some areas, leading to potential limitations.

First, it is common to face problems using inventory data collected differently in each location, mainly associated with taxonomically reliable data, and correctly defined forest typologies in a given locality. As a result, some species common in the biome may go unnoticed, and not all forest types may be equally represented, leading to high standard deviations or biased diversity estimates for each forest type or biogeographic region (Slik et al., 2009). Although the locations were spatially well distributed for the biome, some locations were close to one another. An effort was also made to correct taxonomic names in specialized websites and packages and in using maps of forest typologies in finer detail. This procedure was necessary to correct errors about the representativeness of forest typologies and species richness at the plot level and the biome scale (Ter Steege et al., 2016). The analysis of the diversity patterns discussed is also based on accurate classifications of land cover and markedly anthropic gradients, which may not be available or may occasionally be misaligned with the inventoried plots (Schmit et al., 2006).

Second, different scales of environmental data (10×10 km tables) and inventory form another complication, as plots covering a few hectares are unlikely to represent a 10×10 km area (Slik et al., 2009). In addition, some regions or areas contained more inventoried trees, plots, or surveyed surface area than others, resulting in different levels of diversity captured by the site. Although attempts have been made to correct the diversity values for these sampling biases, likely, they have at least partially influenced the results. Ideally, meta-analyses such as this should be performed on standardized plot layouts in combination with high-resolution environmental data (e.g., Gerstner et al., 2017). Unfortunately, such an approach is still difficult to carry out in more remote tropical areas in the Amazon, where the availability of reliable inventory data is already a problem (ForestPlots.net et al., 2021), not to mention reliable high-resolution soil and climate data (Fick & Hijmans, 2017).

Third, and related to environmental layers and their potential to explain the diversity of large trees, it was noted that the functional importance of edaphic variables (clayContent and waterContent) presented the poorest results among all the variables tested based on related characteristics to predictive importance. These low-importance values obtained can be explained perhaps by the available GIS layers in which the spatially explicit grid (interpolated values) does not correctly explain the inherent values at the plot level, thus requiring an effort to collect data at local scales. A detailed investigation of these factors locally would be fundamental, as it could identify other ecological functions of the edaphic information set for the diversity of tree species in the Amazon (Marca-Zevallos et al., 2022; Terra et al., 2018). Edaphic factors are fundamental in explaining many attributes and ecological patterns of the forest, mainly related to productivity and diversity in tropical forests (Hoffhansl et al., 2020; Homeier & Leuschner, 2021). Joswig et al. (2022) find a dominant joint effect
of climate and soil on trait variation of plant functional characteristics on a global scale. However, our results are consistent with studies of the same tropical geographic regions (e.g., Ter Steege et al., 2013) with highly varied species composition (Slik et al., 2015), which leads us to believe that the main conclusion is satisfactory. Finally, more direct studies on the set of ecological forces acting to define patterns of tree species diversity are needed to clarify these responses.

4.3 | Final considerations

We conclude that the patterns and processes that determine the diversity of large trees clearly differ from those for smaller trees, which has clear implications for ecological theories, conservation, and management.

Although the RF models suggest that the diversity of large trees (dbh ≥ 70 cm) and of all trees (dbh ≥ 10 cm) is associated with different environmental variables, there is a modest and significant association (r = .2, p < .001) between the two Fisher’s alpha-diversity indices (Figure 3b).

Richness is strongly associated with the number of individuals in the sample, especially in the early stages of sampling (Figure 3a). Furthermore, as the sampling proceeds, the rarefaction curves representing communities with trees of different sizes intersect twice, first in a number slightly more significant than 5000 individuals and then in a number slightly <1500 individuals. This means that the relative richness of two datasets (albeit in the same samples), measured on smaller numbers of individuals, does not predict well their relative richness on a larger sample size. This is generally true even when using Chao1 estimators (Chao et al., 2004; Roswell et al., 2021) to predict true diversity. Because richness is so sensitive to sampling effort and relative abundance, its estimate may depend on how samples are patterned. Even the best estimators of asymptotic richness, such as Chao1 (Colwell et al., 2012), cannot reliably predict true community diversity (Jost, 2007). The problem is that both the sample richness and the individual-based richness estimators are strongly influenced by the rarest species, which are precisely the species about which we know less and which are more challenging to identify in the field. This is another way of saying wealth has high uncertainty (Roswell et al., 2021). Indeed, this uncertainty is often insurmountable in estimating and comparing community diversity from samples.

The resulting maps are a potential distribution of diversity bands for both large trees and the set of all trees along the basin. Given the performance of the RF models and the distribution of sampling sites in the basin, no inferences outside the main range of the sampled data and on other ecosystems are allowed. Therefore, our analysis of the distribution of tree diversity is restricted to the Brazilian Amazon basin.

While our analysis does not attempt explicitly to measure the possible impacts of climate change on large tree diversity, it is concerning to note that several climatic variables strongly associated with tree diversity may undergo intense changes under credible climate change scenarios. Climate-induced changes in the frequency of anomalous events, such as the increased occurrence of storms, the frequency of lightning and severe droughts, are already observed (Dale et al., 2001; Kamimura et al., 2022; Seidl et al., 2017). Changes in environmental variables associated with disturbances could have large negative impacts on the diversity and survival of large trees.

Given the enormous tree diversity of tropical forests, understanding the effects of climate change on diversity patterns is critical and should be analyzed as soon as possible at finer scales within Amazonia. Such knowledge is critical to refining conservation perspectives in a changing world—for example, to what extent are currently protected areas in Amazonia susceptible to climate disturbance-induced impacts on their biologically critical large trees? Efforts to understand how deforestation and climate change interact, and to mitigate their impacts, are urgently needed in light of the high and increasing rates of deforestation that have taken place in the Brazilian Amazon in recent years and which threaten the most extraordinary biodiversity on the planet.

AUTHOR CONTRIBUTIONS
Robson Borges de Lima conceived the main idea of the manuscript and accessed relevant Brazilian Amazon Plots in the ForestPlots.net database with the approval of data owners. Robson Borges de Lima and Eric Bastos Görgens led the compilation of field inventory and remote sensing data with the assistance of Cinthia Pereira de Oliveira, Diego Armando S. da Silva, Anderson Pedro B. Batista, Jadsom Coelho de Abreu, and Rinaldo L. Caraciolo Ferreira. All co-authors contributed data. Robson Borges de Lima performed the analyses with the assistance of Eric Bastos Görgens. Robson Borges de Lima wrote the first draft of the manuscript, with all authors providing editorial input.

ACKNOWLEDGMENTS
We thank the numerous researchers and field assistants who collected and provided the field data compiled in the database for the manuscript analyses, without whom this work would not have been possible. This project was funded 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 and 301661/2019-7) (Processes 550467/2010-6 and 301432/2022-8); Amazon Fund (grant 14.2.0929.1); Fundo Iratapuru; Universidade do Estado do Amapá (Processes 0022.0279.1202.0018/2021); and remote sensing data with the assistance of Cinthia Pereira de Oliveira, Diego Armando S. da Silva, Anderson Pedro B. Batista, Jadsom Coelho de Abreu, and Rinaldo L. Caraciolo Ferreira. All co-authors contributed data. Robson Borges de Lima performed the analyses with the assistance of Eric Bastos Görgens. Robson Borges de Lima wrote the first draft of the manuscript, with all authors providing editorial input.

CONFICT OF INTEREST STATEMENT
The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT
Data are available at https://doi.org/10.5281/zenodo.7688243 (accessed on 05 July 2022).


Carvalho, G. (2020). flora: Tools for Interacting with the Brazilian Flora 2020 (0.3.4). https://CRAN.R-project.org/package=flora


**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.