A dataset on the structural diversity of European forests

Marco Girardello*[1,2], Gonzalo Oton*[1], Matteo Piccardo[3] Mark Pickering[3], Agata Elia[3], Guido Ceccherini[1], Mariano Garcia[4] Mirco Migliavacca[1] Alessandro Cescatti[1]

[1]European Commission, Joint Research Centre, Italy
[2]Discipline of Geography, School of Natural Sciences, Trinity College Dublin, Dublin, Ireland
[3]Consultant of European Commission, Joint Research Centre, Italy
[4]Universidad de Alcalá, Department of Geology, Geography and the Environment, Environmental Remote Sensing Research Group, C/ Colegios, 2, Alcalá de Henares, 28801, Madrid, Spain

Correspondence to: Marco Girardello (marco.girardello@gmail.com), Gonzalo Oton (gonzalo.oton@ec.europa.eu)

* These authors contributed equally

Abstract. Forest structural diversity, defined as the heterogeneity of canopy structural elements in space, is an important axis of functional diversity and is central to understanding the relationship between canopy structure, biodiversity, and ecosystem functioning. Despite the recognised importance of forest structural diversity, the development of specific data products has been hindered by the challenges associated with collecting information on forest structure over large spatial scales. However, the advent of novel spaceborne LiDAR sensors like the Global Ecosystem Dynamics Investigation (GEDI) is now revolutionising the assessment of forest structural diversity by providing high-quality information on forest structural parameters with a quasi-global coverage. Whilst the availability of GEDI data and the computational capacity to handle large datasets have opened up new opportunities for mapping structural diversity, GEDI only collects sparse measurements of vegetation structure. Continuous information of forest structural diversity over large spatial domains may be needed for a variety of applications. The aim of this study was to create wall-to-wall maps of canopy structural diversity in European forests using a predictive modelling framework based on machine learning. We leverage multispectral and Synthetic Aperture Radar (SAR) data to create a series of input features that were related to eight different structural diversity metrics, calculated using GEDI. The models proved to be robust, indicating that active radar and passive optical data can effectively be used to predict structural diversity. Our dataset finds applications in a range of disciplines, including ecology, hydrology, and climate science. As our models can be regularly rerun as new images become available, it can be used to monitor the impacts of climate change and land use management on forest structural diversity.

1 Introduction

Information on forest canopy structure is important for several disciplines, including Earth System Science, Ecology, Hydrology, and Climate Science. Forest canopy structure plays a fundamental role in ecosystem functioning by affecting carbon storage and cycling, regulating the hydrological cycle, and influencing local and regional climate patterns (Migliavacca et al., 2021; Shugart et al., 2010; Sun et al., 2018). In addition, canopy structure is critical for maintaining high levels of biodiversity by supporting a high diversity of ecological niches (Larue et al., 2019).

The concept of structural diversity or complexity, herein defined as the heterogeneity or variability of canopy structural elements in vertical or horizontal space (Ehbrecht et al., 2021; Hakkenberg et al., 2023; LaRue et al., 2019), is central to understanding the relationship between canopy structure, biodiversity, and ecosystem functioning. Structurally diverse forests can host a wide variety of functionally complementary species, which tend to increase resource-use efficiency and promote feedbacks that enhance resource availability (Gough et al., 2019; Murphy et al., 2022). As a result, these forests can capture light more efficiently, leading to increased ecosystem productivity (Atkins et al., 2018; Toda et al., 2023). Therefore, the availability of data on forest structural diversity over large spatial scales is critical for predicting and managing the response of forest ecosystems to global change.

Mapping forest structural diversity over large spatial scales proved challenging due to the lack of comprehensive datasets and consistent data collection methodologies, hindering our ability to predict ecosystem function at large geographic scales. Whilst forest structural parameters can be measured in various ways, traditional field-based measures of stand structure are generally labour-intensive and have been limited to small areas (Goodbody et al., 2023). Laser scanning, or LiDAR, has been proved a sound alternative for measuring tree height from 3D data measured through echoes (Coops et al., 2021). However, data from airborne LiDAR have been limited in spatial and temporal coverage to specific regions (Hancock et al., 2021). Recent advances in satellite remote sensing technology and computational capabilities have made it possible to measure a range of structural variables at larger scales than ever before. Notably, the Global Ecosystem Dynamics Investigation (GEDI) (Dubayah et al., 2020a) instrument, placed on board the International Space Station (ISS) in December 2018, collecting LiDAR samples until March 2023, has revolutionized the assessment of forest structure. Recent studies have shown how structural data collected by GEDI can be used in several applications ranging from biomass estimation to the monitoring of biodiversity and ecosystem disturbances (Crockett et al., 2023; Hakkenberg et al., 2023; Holcomb et al., 2024). These early examples demonstrate the future potential of the GEDI mission.

Whilst the availability of GEDI data and the computational capacity to handle large datasets have opened up new opportunities to map structural diversity, GEDI only collects sparse measurements of vegetation structure. Although the GEDI mission has recently been extended, it is expected to cover only a minimal fraction of the land surface. Depending on the application of interest, continuous information on structural diversity over forests may be needed. Combining GEDI with other types of satellite remote sensing data within a machine learning framework may thus be necessary for the creation of structural diversity data products that have continuous coverage and that extend beyond the timeframe covered by the GEDI mission. In this context, predictor variables derived from complementary satellite observations are used to bridge the gap between sparse GEDI measurements and the need for wall-to-wall maps of forest structural diversity. These predictors can act as observable proxies for canopy structural complexity, enabling the spatial extrapolation of the GEDI-derived structural diversity metrics across Europe. Several recent studies have successfully combined GEDI data with other remote sensing data sources to predict canopy structure in areas not covered by GEDI, paving the way for mapping specific structural features of vegetation regionally and globally (Aragoneses et al., 2024; Lang et al., 2023; Potapov et al., 2021; Schwartz et al., 2024). Additionally, preliminary efforts to assess the potential of GEDI data to capture canopy diversity over different regions have been carried out (Schneider et al., 2020). However, despite these significant advances, no efforts have been made to map forest structural diversity at a continental scale in Europe.

To address the lack of readily available structural diversity data, we combined a suite of structural diversity indicators calculated using GEDI data with active radar and passive optical data from the Sentinel-1, Sentinel-2, and ALOS-PALSAR missions. This sensor combination was specifically selected to enable spatially and temporally consistent, wall-to-wall estimates suitable for large-scale and monitoring, while capturing complementary structural information across different canopy layers: Sentinel-2 multispectral data are sensitive to canopy biochemical and structural properties at the crown surface, including vegetation density and phenological state. Sentinel-1 C-band SAR interacts primarily with the upper canopy and smaller structural components, capturing variations in canopy roughness ALOS-PALSAR-2 L-band SAR, owing to its longer wavelength, exhibits enhanced sensitivity to larger structural elements and sub-canopy features, providing information on forest vertical complexity. These different sources of data were then integrated using a predictive modelling framework, based on a machine learning method. The resulting models were used to predict structural diversity across Europe. Although Sentinel-1, Sentinel-2 and ALOS-PALSAR-2 data have been previously used for predicting canopy height and other structural components of forests, their joint use for mapping forest structural diversity has not yet been attempted (Liu et al., 2025; Wang et al., 2024). Our analysis includes a total of eight structural diversity metrics, including metrics that quantify the vertical and horizontal heterogeneity of the canopy, as well as metrics that quantify the heterogeneity of forest structure among GEDI observations within a given area. The dataset presented here is readily available for use as input in various environmental models and analyses.

2 Methods

We calculated eight forest structural diversity metrics using NASA GEDI observations (Dubayah et al., 2020b). A list of the metrics is reported in Table 1. A machine learning (ML) framework was used to model the relations between each metric and a series of predictors derived from passive optical and active radar remote sensing data. The model was then used to create a structural diversity dataset that covers the whole forested domain of Europe, extending up to ~52° North, which corresponds to the northern latitudinal limit of the GEDI mission. The creation of the dataset involved five main steps (Fig. 1): (i) satellite remote sensing data pre-processing, (ii) structural diversity metric calculation (iii) model training, (iv) model validation, and (v) prediction (Fig. 1). Data from GEDI, Sentinel 1 and 2, ALOS-Palsar-2 were pre-processed and downloaded from Google Earth Engine (GEE), a cloud-based infrastructure that combines a multi-petabyte catalogue of satellite imagery and geospatial datasets with planetary-scale analysis capabilities (Gorelick et al., 2017).

We used data covering forests that had remained ecologically stable, meaning they experienced no canopy loss, from 2000 to 2021, as identified through the Global Forest Change product by Hansen et al. (2013). Furthermore, our analyses were limited to areas where tree cover exceeded 30% and which bordered at least 6 out of 8 neighbouring pixels, also with tree cover exceeding 30%. While our threshold is more stringent than Food and Agriculture Organization (FAO) definition of forest (FAO, 2000), which specifies an area spanning more than 0.5 hectares with trees taller than 5 metres and a canopy cover of more than 10%, it was chosen to capture areas with substantial arboreal density. Although our selection was guided by the FAO's broader forest criteria, we customized these guidelines to suit our research focus. This threshold was intentionally adopted to focus on landscapes with well-developed and spatially continuous forest canopies. This more conservative threshold reduces the influence of sparsely treed or transitional land-cover types and improves the robustness and interpretability of GEDI-based structural diversity metrics

2.1 Structural diversity metrics

Structural diversity can be characterized in a variety of ways depending on the data from which it is calculated and the intended application. In this work we adopted the common definition where diversity is defined in the vertical dimension as heterogeneity of vegetation height and in the horizontal dimension as canopy heterogeneity (Hakkenberg and Goetz, 2021). We chose a set of metrics that would characterize the heterogeneity within and among structural features for a given area, reflecting both local (alpha) and regional (beta) measures of structural diversity. These complementary metrics have been demonstrated to be particularly crucial for predicting tree diversity and ecosystem functioning (Coverdale and Davies, 2023; Ma et al., 2022; Zhai et al., 2024). A summary of the metrics with the input data used is reported in Table 1.Several GEDI, including Foliage Height Diversity (FHD from the GEDI L2B product) and the Waveform Structural Complexity Index (WSCI, from the GEDI L4C product), already capture important aspects of forest structural heterogeneity and have proven highly valuable for large-scale analyses. However, recent work has shown that these indices exhibit strong scaling relationships with top-of-canopy height (RH98) (de Conto et al., 2024).

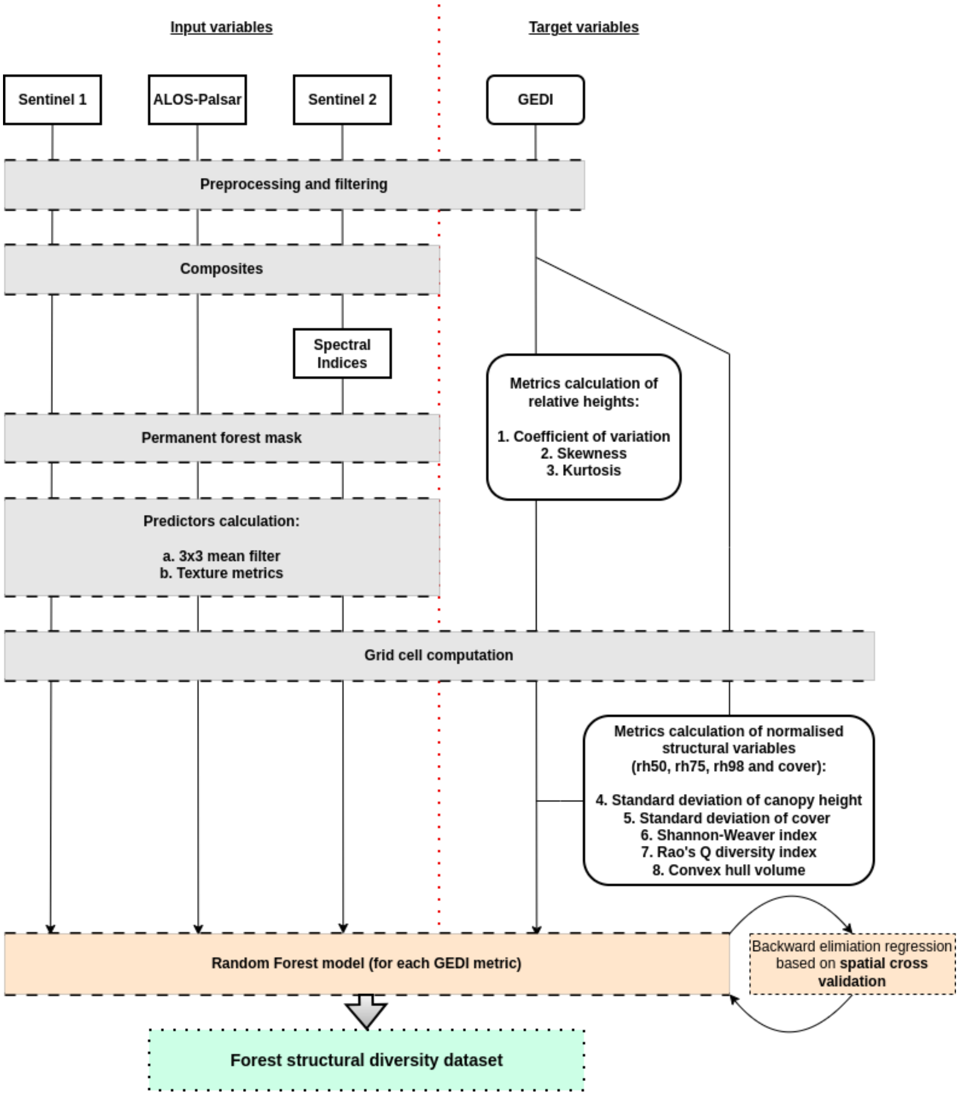


Figure 1. General workflow employed in the creation of the forest structural diversity dataset. The workflow is segmented by a red dashed line, delineating the Remote Sensing predictors from the inputs to the target GEDI data. Boxes with solid edges represent the data that were directly utilised to train the Random Forest models. The grey boxes indicate the preliminary steps undertaken before the model training phase. The process culminates with orange boxes, which signifies the development of the predictive model itself, leading to the green box that represents the final output outcome—the forest structural diversity dataset.

In the context of this study, our objective was therefore not to directly use existing GEDI products as target variables, but to develop structural diversity metrics that explicitly quantify heterogeneity while minimising direct dependence on canopy height. This motivated the selection of eight complementary metrics based on the distributional properties of GEDI relative height profiles and canopy cover, ecologically interpretable, minimally redundant with each other and with height, and suitable for spatial aggregation and wall-to-wall mapping using multi-sensor satellite data.

The eight structural diversity metrics were designed to span three complementary dimensions of forest structural diversity (Table 1). Vertical heterogeneity within individual canopy profiles is characterised using distributional metrics derived from GEDI relative height profiles, namely the coefficient of variation, skewness, and kurtosis, which capture differences in vertical layering and profile shape. Horizontal heterogeneity is described by the variability among GEDI observations within a spatial unit, quantified through the standard deviation of canopy height (CH) and canopy cover (CC), reflecting spatial variation in canopy structure across the landscape. Finally, combined structural diversity is represented using multivariate metrics (Shannon index (SW), Rao's quadratic entropy (RAO), and convex hull volume (CVH)), which synthesize information from multiple GEDI measurements into a single integrated index per spatial unit. Together, these metrics provide an interpretable yet comprehensive characterisation of forest structural diversity while minimising redundancy among metrics and with top-of-canopy height.

2.1.1 GEDI input data and general framework

GEDI data are collected from a full waveform LiDAR sensor operating onboard the International Space Station (ISS) from April 2019 until January 2023. Due to the orbital path of the International Space Station (ISS), GEDI's coverage is primarily limited to latitudes between ~50° North and south. The instrument provides sparse measurements (hereinafter sample plots or shots) of vegetation structure over an area defined by a sampling footprint of about 25 m diameter.

Input data included the GEDI Level 2A Relative Heights (RH), and the Level 2B total Canopy Cover (CC) values (see Table 1). In the literature, rh^{98} is taken as a reference for the top canopy height (CH) (Lang 2022), CC is the proportion of the shot covered by the vertical projection of the tree crowns. The GEDI data were downloaded from Google Earth Engine after applying a filtering procedure to remove low-quality and unreliable observations and to reduce noise in the input data, based on standard GEDI quality flags and thresholds (Table S1 in the Supplement).

Structural diversity metrics were computed for each spatial analysis unit, defined as regular grid cells (hereafter referred to pixels) at 1 km, 5 km, and 10 km spatial resolution. For each pixel, all number *(M) of* valid GEDI shots overlapping the pixel between April 2019 and January 2023 were collected. The structural diversity metrics of a given pixel are calculated by aggregating all M overlapping that pixel. Each GEDI shot i was characterized by its RH distribution $RH_i = \{rh_i^k, rh_i^{k+1}, \ldots, rh_i^{100}\}$ with k: $rh_i^k \geq 0$ (i.e. only the positive values were considered) and total canopy cover cc_i. To ensure robust estimation of structural diversity, pixels with fewer GEDI observations than a minimum sampling threshold were excluded. This threshold was defined as the median number of valid GEDI shots across all pixels at a given spatial resolution. In addition, extreme metric values were filtered using a z-score criterion, with values exhibiting a deviation greater than 3 across pixels discarded as outliers. The remaining pixels were then used to compute the structural diversity metrics. A post-processing step was applied to remove extreme values: pixels with structural diversity values exhibiting a z-score greater than 3 were discarded as outliers.

We computed the eight structural diversity metrics at three spatial resolutions (1 km, 5 km, and 10 km) using grids in the Lambert Azimuthal Equal Area (LAEA) projection. The choice of these spatial resolutions reflects a trade-off between spatial detail, GEDI sampling density, and metric robustness. Because GEDI provides sparse footprint-level measurements, reliable estimation of structural diversity requires a sufficient number of observations within each spatial unit. Coarser resolutions improve metric stability, while finer resolutions provide greater spatial detail at the cost of higher uncertainty. The 10 km resolution was identified as the most robust scale for continental-scale analyses and is compatible with the spatial aggregation typically used in regional ecosystem and Earth system models. The 5 km and 1 km products support applications requiring finer spatial detail, such as regional ecological analyses and biodiversity assessments. We acknowledge that applications focused on small-scale disturbances, edge effects, or forest fragmentation would benefit from even finer resolution; however, at such scales, the limited density of GEDI observations substantially constrains reliable metric estimation over large areas.

In the following sections, we detail the methodology employed for calculating the diversity metrics and predictor variables, which makes use of the mean $\mu(X)$, standard deviation $\sigma(X)$, skewness $\gamma(X)$, excess kurtosis $\kappa(X)$, coefficient of variation $cv(X)$ of a variable $X = \{x_1, \ldots, x_M\}$, where X represents a vector of observations (see Appendix A for the explicit formulations).

2.1.2 Vertical Diversity Metrics

RH metrics provide information on the vertical distribution of the plant elements, that is, the vertical profile (VP) of the vegetation (see Fig. S1 in the Supplement). The VP in a sample can be reconstructed from the corresponding RH distribution, and the profile's moments (i.e. mean, standard deviation, skewness, kurtosis) are well approximated by the RH distribution's moments (Fig. S1 in the Supplement).

The following calculated indicators characterise the heterogeneity of the vertical profile:

1. the average coefficient of variation of the vertical profiles

$$\tau_{CV} = \mu(CV)$$

with $CV = \{cv(RH_1), \ldots, cv(RH_M)\}$. The coefficient of variation$cv(RH)$quantifies the extent of verticalvariability in relation within a vertical profile as the ratio between the standard deviation and the mean of the relative height distribution. τ_{CV} therefore indicates greater vertical dispersion relative to mean profile height and hence greater vertical heterogeneity;

2. the average skewness of the vertical profiles

$$\tau_{SK} = \mu(\Gamma)$$

where $\Gamma = \{\gamma(RH_1),\ \ldots,\ \gamma(RH_M)\}$. represents the set of skewness values computed from the RH distribution of each of the M GEDI observations within the pixel. Skewness $\gamma(RH)$, or third standardized moment, is a measure of the asymmetry of the VP about its mean, and it can be positive, negative, or zero (Fig. S2, a and d panels in the Supplement). If VP is a unimodal distribution (a distribution with a single peak), positive skewness generally indicates an asymmetric tail extending toward larger height values (overstorey heterogeneity), while negative skewness suggests a tail extending toward smaller height values (understorey heterogeneity). However, note that in the cases where one tail is long, but the other tail is fat, or the distribution is multi-modal, skewness does not always obey this simple rule.

3. the average excess kurtosis of the vertical profiles

$$\tau_{KU} = \mu(\mathrm{K})$$

where $\mathrm{K} = \{\kappa(RH_1),\ \ldots,\ \kappa(RH_M)\}$ represents the set of excess kurtosis values computed from the RH distribution of each of the M GEDI observations within the pixel Excess $\kappa(RH)$ is a measure of the "tailedness" of the VP, and it is equal to 0 for any univariate normal distribution, (Fig. S2 a and d panels in the Supplement). Distributions with negative/positive excess kurtosis are said to be platykurtic/leptokurtic. Platykurtic distributions show fewer and/or less extreme outliers than the normal distribution. In this case, the vegetation mass is more concentrated around the VP mean than near the vertical extremes (i.e. the ground and to top canopy height). However, it is important to note that while kurtosis, the fourth moment, does play a role in characterizing the shape of VP, its influence is comparatively smaller than that of the standard deviation, the second moment, and skewness, the third moment. For instance, two distinct VPs may exhibit identical excess kurtosis while displaying markedly disparate distributions in terms of standard deviations.

2.1.3 Horizontal Diversity Metrics

We calculated 5 vertical horizontal diversity indices.

1. the standard deviation of the canopy heights

$$\tau_{CH} = \sigma(CH)$$

where $CH = \{rh_1^{98},\ \ldots,\ rh_M^{98}\}$ represents the set of rh^{98}values within the τ_{CH} τ_{CH} indicates the spread of the canopy heights in the area.

2. the standard deviation of the total canopy cover

$$\tau_{CC} = \sigma(CC)$$

where $CC = \{cc_1,\ \ldots,\ cc_M\}$ represents the set of canopy cover values within the pixel..τ_{CC} indicates the spread of the total canopy cover in the area.

2.1.4 Combined Vertical and Horizontal and Diversity Metrics

3. the Shannon-Weaver index

$$\tau_{SW} = -\sum_{\log p_{\varepsilon\pi o\omega}} p_{\varepsilon\pi o\omega}\ \log p_{\varepsilon\pi o\omega}$$

in a 4D cartesian space defined on the basis $(rh^{50},\ rh^{75},\ rh^{98},\ cc)$, where $p_{\varepsilon\pi o\omega}$ is the fraction of the GEDI samples within the pixel falling in a specific bin (see Appendix A2). We used a 5-unit bin size on each axis, and the GEDI CC values were amplified by 10. τ_{SW} measures the uncertainty or disorder inherent to the variable's possible outcomes. $\tau_{SW} = 0$ when all observations are confined within a single bin, otherwise $\tau_{SW}$is larger than zero. Higher values indicate heterogeneity, while lower values suggest homogeneity.

4. Rao's quadratic diversity index

$$\tau_{RAO} = \sum_{\varepsilon\pi o\omega}\sum_{\varepsilon'\pi'o'\omega'} p_{\varepsilon\pi o\omega}\ D_{\varepsilon\pi o\omega}^{\varepsilon'\pi'o'\omega'}\ p_{\varepsilon'\pi'o'\omega'}$$

in the 4D cartesian space defined on the basis $(rh^{50},\ rh^{75},\ rh^{98},\ cc)$, where $p_{\varepsilon\pi o\omega}$ is the fraction of the GEDI samples within the pixel falling in a specific bin and $D_{\varepsilon\pi o\omega}^{\varepsilon'\pi'o'\omega'}$ the cartesian distance between two bins (see Appendix A2). We used a 1-unit bin size on each axis, and the GEDI CC values were amplified by 10. τ_{RAO} ranges from zero, indicating no diversity, to positive numbers. Differently from τ_{SW}index, τ_{RAO} considers both abundance ($p_{\varepsilon\pi o\omega}$ terms) and dissimilarity in the sampled data ($D_{\varepsilon\pi o\omega}^{\varepsilon'\pi'o'\omega'}$ term).

5. convex hull volume

$$\tau_{CVH} = CVH(SHT)$$

in the 4D cartesian space defined on the basis $(rh^{50},\ rh^{75},\ rh^{98},\ cc)$. We used a 1-unit bin size on each axis, and the GEDI CC values were amplified by 10. CVH is the function calculating the convex hull volume on the ensemble $SHT = \{sht_1,\ \ldots,\ sht_M\}$, with $sht_i = \left(rh_i^{50},\ rh_i^{75},\ rh_i^{98},\ cc_i\right)$. Larger volumes indicate increased heterogeneity.

Table 1. Structural diversity metrics computed in this study.

Metric	Description	Units	GEDI data	Diversity
τ_{CV}	VP coefficient of variation	-	RH	vertical
τ_{SK}	VP skewness	-	RH	vertical
τ_{KU}	VP excess kurtosis	-	RH	vertical
τ_{CH}	CH standard deviation	m	rh^{98}	horizontal
τ_{CC}	CC standard deviation	-	cc	horizontal
τ_{SW}	Shannon-Weaver index	-	$(rh^{50}, rh^{75}, rh^{98}, cc)$	combined
τ_{RAO}	Rao's quadratic entropy index	-	$(rh^{50}, rh^{75}, rh^{98}, cc)$	combined
τ_{CVH}	Convex Hull volume	-	$(rh^{50}, rh^{75}, rh^{98}, cc)$	combined

2.2 Predictor variables

The variables used as ML predictors were calculated from Sentinel-1, Sentinel-2, and ALOS-PALSAR-2 observed data, which provide complementary information on forest canopy structure derived from optical reflectance, C- and L-band SAR backscatter, and associated textural properties. The predictor calculation involved the following steps:

1. appropriate bands/indices ϕ_α were calculated from the remote sensing raster images;
2. the $\phi_{\alpha,i}^{\beta}$ values, with β equal to SM, ASM, ENT, or DISS, are calculated from the pixels within the 7x7 window aligned with the footprint of the GEDI shot i . $\phi_{\alpha,i}^{SM}$ is the spatial mean. $\phi_{\alpha,i}^{ASM}$, $\phi_{\alpha,i}^{ENT}$, and $\phi_{\alpha,i}^{DISS}$ are the texture metrics Angular Second Moment (ASM), entropy, and dissimilarity index, respectively (see Appendix A3).
3. the raster images of the predictors were computed as $\phi_{\alpha}^{\beta} = \mu(\Phi_{\alpha}^{\beta})$, where the mean is calculated on the M values $\Phi_{\alpha}^{\beta} = \{\phi_{\alpha,1}^{\beta}, \dots, \phi_{\alpha,M}^{\beta}\}$ corresponding to the geographical positions that overlap the image pixels.

In the following, we present what satellite remote sensing data were used and how they were combined for the calculation of the indices. A total of 47 predictors were derived. A summary of the predictors is reported in Table S2 in the Supplement.

2.2.1 Sentinel-1 radar data

The European Space Agency's (ESA) Sentinel-1 (S1) comprises a constellation of two polar-orbiting satellites, sun-synchronous orbit with a 12-day repeat cycle, which operate day and night a C-band (λ = 5.5 cm) Synthetic Aperture Radar (SAR) to capture data at a spatial resolution of approximately 10 meters. The radar enables the acquirement of imagery regardless of the weather, and the C-band frequency is particularly effective in interacting with fine vegetative elements such as leaves and branches (Naidoo et al., 2015). In our study, from Sentinel-1 we utilized both backscatter and coherence data. Backscatter is the portion of the outgoing radar signal that the target redirects directly back towards the radar antenna. The backscatter characteristics provide crucial insights into the physical properties of forest canopies. For the year 2020, we focused on the signal dual-polarization VV and VH Sentinel-1A (S1A) and Sentinel-1B (S1B) Ground Range Detected (GRD) data, acquired in the Interferometric Wide (IW) swath mode, as it predominantly covers land masses (Kellndorfer et al., 2022). VV(H) is a mode that transmits vertical waves and receives vertical (horizontal) waves to create the SAR image. We selected data from the descending orbit, which has been shown to exhibit fewer correlations with evapotranspiration (ET) (Mueller et al., 2022). Sentinel-1 data used in this study were obtained from Google Earth Engine, where they had already undergone some pre-processing. Preprocessing steps carried out by the Google Earth Engine team include applying the orbit file for geocoding, removing GRD border noise and thermal noise, and performing radiometric calibration. We performed a radiometric terrain correction following (Vollrath et al., 2020), as well as the removal of stripes and edges. Following radiometric terrain correction and the removal of stripes and edge artefacts, we selected all valid Sentinel-1 observations captured over Europe within a six-month window centred on the date of maximum NDVI identified independently for each pixel from the Sentinel-2 dataset (see Section 2.2.3). We then derived:

1. the S1 backscatter six-month mean $\phi_{S1VVgs\mu}$ and $\phi_{S1VHgs\mu}$, where the mean is intended to mitigate speckle noise while emphasizing the vegetation growing season;
2. the S1 backscatter standard deviation growing season $\phi_{S1VVgs\sigma}$ and $\phi_{S1VHgs\sigma}$;
3. the S1 backscatter bi-monthly mean $\phi_{S1VVpre\mu}$ and $\phi_{S1VHpre\mu}$ for a window extending two months before the month before the peak, $\phi_{S1VVact\mu}$ and $\phi_{S1VHact\mu}$ for the period spanning one month before to one month after the peak, and $\phi_{S1VVpost\mu}$ and $\phi_{S1VHpost\mu}$ for the two months after the month after the peak.

Coherence is the relationship between waves in a beam of electromagnetic (EM) radiation. Two wave trains of EM radiation are coherent when they are in phase. In radar, the term coherence is also used to describe systems that preserve the phase of the received signal. Coherence measurements serve as a valuable tool for monitoring temporal changes in forested environments (Bruggisser et al., 2021; Cartus et al., 2022). The coherence data utilized in this study were extracted from the dataset developed by Kellndorfer et al. (2022). This dataset is the product of multi-temporal, repeat-pass interferometric processing of S1 SAR images. It incorporates signal dual-polarization VV and VH data from S1A and S1B in Single Look Complex (SLC) format, utilizing the IW swath mode from the year 2020. The product is divided into seasonal sets, and we selected summer (June-August) coherence metrics ϕ_{CO}, aligned with the growing season, employing a 12-day repeat-pass interval to optimize the balance between image continuity and temporal resolution. This interval was chosen to minimize gaps in the image series, compared to shorter intervals (such as 6 days), while longer intervals (e.g., 18, 24, 36, or 48 days) could result in excessive decorrelation. With a relatively unchanged scene between acquisitions, higher coherence values are achieved, which correlate strongly with the radar signal and hence, reduce noise levels. Furthermore, we prioritized signal VV polarization to enhance our understanding of the data, as it minimizes vegetation decorrelation effects (Pan et al., 2022)

2.2.2 ALOS-PALSAR-2 radar data

The ALOS-PALSAR-2 (Advanced Land Observing Satellite - Phased Array type L-band Synthetic Aperture Radar) system, developed by the Japan Aerospace Exploration Agency (JAXA), operates in the L-band frequency (λ = 23.62 cm) at a spatial resolution of 25 meters. The L-band is particularly effective at penetrating canopy layers to provide backscatter signals from larger vegetative features such as branches and trunks, and even from the ground. For our analysis, we made use of the global mosaic of backscatter annual composites, which incorporate signal dual-polarization HH and HV data (Shimada et al., 2014) from the years 2019 and 2020, accessed via GEE. In instances where the data availability was constrained for an annual composite, the dataset was supplemented with observations from adjacent years. To ensure the reliability of our dataset and account for possible gaps in observations, we averaged data across two years to generate $\phi_{AP2HH\mu}$ and $\phi_{AP2HV\mu}$ data. This approach helps mitigate noise and stabilize the composite images.

2.2.3 Sentinel-2 optical data

The ESA Sentinel-2 (S2) mission comprises a constellation of two polar-orbiting satellites placed in the same sun-synchronous orbit, phased at 180° to each other. Its high revisit time (10 days at the equator with one satellite, and 5 days with 2 satellites at best) allows monitoring of the Earth's surface changes. The Multi-Spectral Instrument (MSI) on board the 2 platforms collects the sunlight reflected from the Earth and supplies high-resolution multispectral imagery with resolutions of 10 and 20 meters. Data are acquired at 10 m spatial resolution for Visible (Blue, Green, Red) and Near-Infra-Red (NIR) bands, and at 20 m spatial resolution for VNIR-Red Edge (RE1, RE2, RE3, RE4) and Short Wave Infra-Red (SWIR) bands (SWIR1, SWIR2). The Level-2A product provides atmospherically corrected Surface Reflectance (SR) images. In this study we used all the Level-2A images from 2000 to 2021 identified by a scene-level cloud and snow cover smaller than 70% and 5%, respectively, as provided by Google Earth Engine. We then calculated:

1. the Normalized Difference Vegetation Index

$$\phi_{NDVI} = \frac{\rho NIR - \rho Red}{\rho NIR + \rho Red}$$

as proposed by (Rouse et al., 1974) it is a widely recognized index strongly correlated with vegetation health and primary productivity;

2. the Normalized Difference Water Index

$$\phi_{NDWI} = \frac{\rho NIR - \rho SWIR1}{\rho NIR + \rho SWIR1}$$

as proposed by (Gao, 1996), it is correlated with leaf water content.

3. the Normalized Difference Red Edge Index

$$\phi_{NDRE} = \frac{\rho NIR - \rho RE1}{\rho NIR + \rho RE1}$$

as proposed by Gitelson and Merzlyak, (1994) it offers sensitivity to chlorophyll content and is useful in assessing forest composition and canopy cover;

4. the Modified Soil Adjusted Vegetation Index

$$\phi MSAVI = \frac{2 \cdot \rho NIR + 1 - \sqrt{(2 \cdot \rho NIR + 1)^2 - 8 \cdot (\rho NIR - \rho Red)}}{2}$$

as proposed by (Qi et al., 1994), it is suited to monitoring vegetation density and dynamics, particularly during early growth stages when bare soil is prevalent, thereby minimizing soil background effects;

5. the Green Normalized Difference Vegetation Index

$$\phi_{GNDVI} = \frac{\rho NIR - \rho Green}{\rho NIR + \rho Green}$$

as proposed (Gitelson and Merzlyak, 1998), it responds to chlorophyll concentration and is indicative of vegetation composition, structure, habitat conditions, and species diversity;

6. the standard deviation of NDVI

$$\phi_{NDVI\sigma} = \sigma(\phi NDVI)$$

as noted by (Perrone et al., 2024), it accounts for a significant portion of the variability observed in-situ plant diversity.

2.3 Model training and validation

We used a machine learning method - Random Forest (Breiman, 2001) - to quantify the relations between the remote sensing predictors and the eight metrics. Random Forest is an ensemble learning method based on decision trees that is widely employed for regression tasks. A key advantage of Random Forests is that model fitting is relatively fast and hyperparameter optimization requires only a moderate amount of tuning, compared to other machine learning methods. Optimization of the Random Forest model typically involves tuning a number of hyperparameters. These include the size of the forest (i.e. the number of decision trees), the method of bootstrapping samples, and the setting of the maximum depth for the trees. We specified a fixed number of trees, 600; bootstrapping, a technique that involves random sampling with replacement, which contributes to the diversity of the decision trees in the model and helps prevent overfitting; and we did not impose any limitations on the depth of the individual decision trees, allowing them to expand fully. To evaluate the performance of the Random Forest model, we used mean squared error (MSE) as the metric.

To mitigate the potential for overfitting, we used a backward stepwise selection process that begins with a full model including all available predictors. The algorithm then iteratively removes the least important feature, as determined by its contribution to model performance. The relative importance of predictors was assessed using a permutation procedure (Altmann et al., 2010). At each iteration, the model complexity is reduced by one predictor, and the resulting model is evaluated. We compared each newly simplified model to the immediate predecessor to determine whether there was an improvement in performance or a decrease that was less than 1% worse. The elimination process is halted if the removal of additional predictors causes the model's performance to decrease by more than 1% compared to the previous iteration. At each step, a spatial cross-validation procedure is used to assess the performance of the model. The metric we utilized to assess model performance throughout this process was the coefficient of determination (R^2).

To validate the reduced models, we used two types of validation techniques to assess their predictive accuracy and robustness:

- Random train-validation split: in this approach, the dataset was randomly split, allocating 33% for model validation. Random validation is a common method that provides a quick and often effective means of evaluating model performance on unseen data. However, it has a notable drawback when dealing with spatial data: it disregards the spatial structure inherent in the dataset (i.e. points close to each other are, generally, more similar than points further away). By ignoring this spatial autocorrelation, random validation may inadvertently conceal overfitting issues, leading to an overly optimistic perception of the model's predictive capabilities.
- 10-Fold Spatial Cross-Validation (Roberts et al., 2016): we implemented a 10-fold spatial cross-validation procedure to address the shortcomings of random validation, thus reducing the overfitting. This more sophisticated method partitions the data into ten spatially distinct subsets, or folds, ensuring that each fold comprises disjointed sets that are geographically separated. The partitioning is achieved by clustering data points according to their spatial coordinates, which preserves the spatial structure and autocorrelation present in the dataset. During the validation process, each fold is used once as a validation set while the remaining folds serve as the training set. This technique provides a more realistic evaluation of the model's performance and its ability to generalize across different spatial regions, thereby offering a safeguard against overfitting and ensuring a more reliable assessment of the model's true predictive power.

Both validation approaches were retained because they address complementary evaluation objectives. Spatial cross-validation provides a conservative assessment of model generalisation in the presence of spatial autocorrelation and is therefore more appropriate for evaluating transferability across regions. Random validation, while potentially optimistic for spatial data, was included to facilitate comparison with previous studies that rely on this approach and to characterise overall model behaviour under standard machine-learning evaluation settings. Throughout the manuscript, spatial cross-validation results are emphasised when discussing model robustness and applicability.

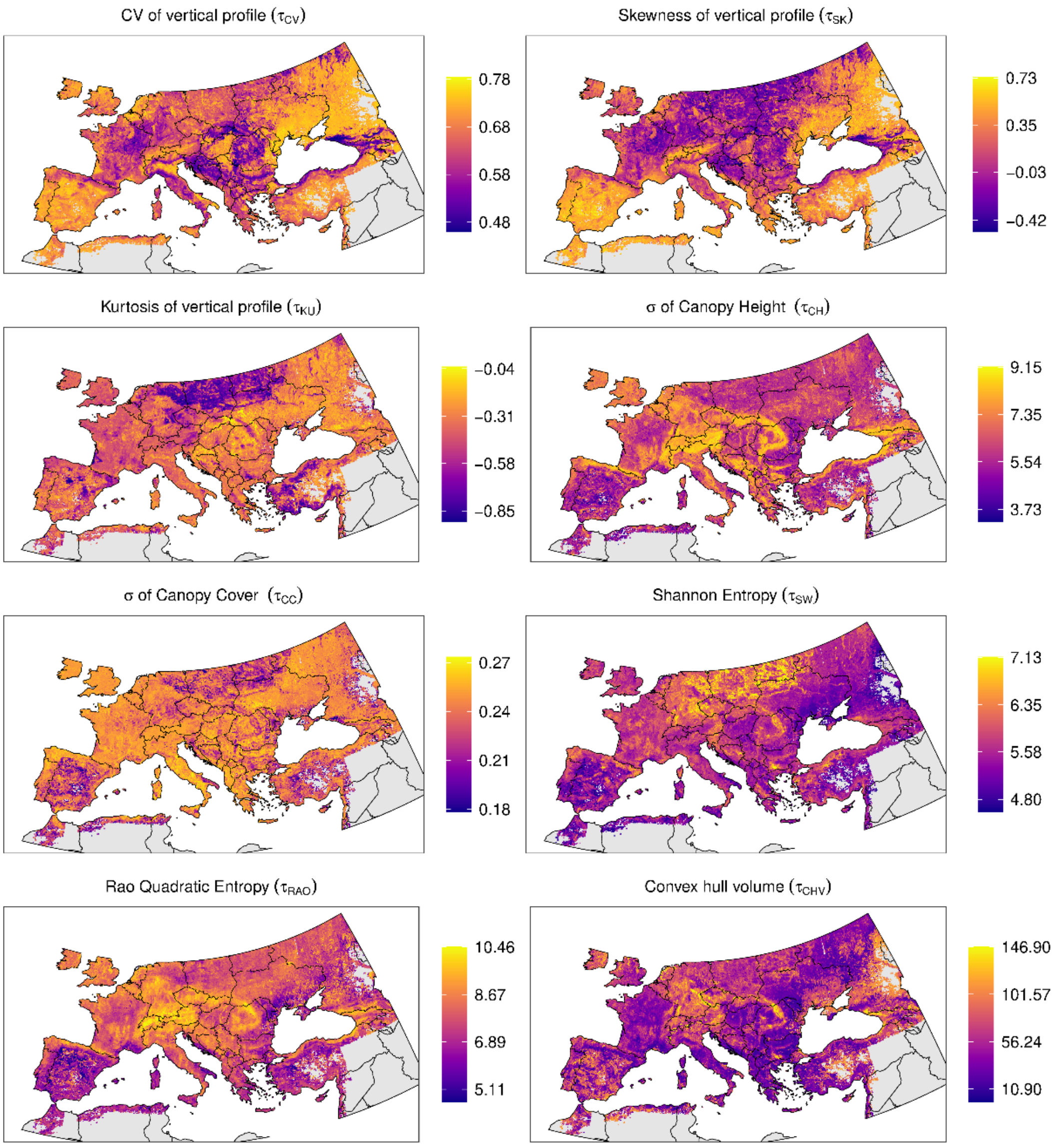
Models were fitted to datasets created at different resolutions including data calculated at 10 km, 5 km and 1 km. Prediction uncertainty was quantified by calculating the standard deviation of predictions across the ensembles of decision trees in the Random Forest models. This metric captures the variability in predictions among individual trees within each model, providing a measure of uncertainty associated with predictions for the different response variables.

3 Results

3.1 Spatial patterns of forest structural diversity

The dataset includes spatial grids for eight structural diversity metrics at three different resolutions (10 km, 5km and 1 km). These metrics show a significant variation in structural diversity across the European forests as shown in Fig. 2 (see also Fig. S3 and Fig. S4 in the Supplement for 5km and 1km resolution datasets).

An examination of the variability in the 10 km resolution metrics in climate space revealed distinct patterns along temperature and precipitation gradients (Fig. 3). Patterns of variability in metrics describing vertical heterogeneity showed significant differences when comparing the coefficient of variation (τ_{CV}) and skewness (τ_{SK}) against kurtosis (τ_{KU}). The coefficient of variation and skewness primarily exhibited high values at the extremes of the climatic gradient. This is observed in warm and arid climates where total annual precipitation is below ~500 mm and Annual Mean Temperature is above ~10 °C, as well as in colder climates where Annual Mean Temperature is below ~5 °C. Patterns of variability in the kurtosis were more nuanced, consistently showing negative values across the European domain, which suggests a tendency for a platykurtic distribution in the vertical profile of canopies under diverse environmental conditions. The most pronounced negative kurtosis values were observed for the northern part of the temperate climate zone (Fig. 3). By contrast, more heterogeneous patterns occurred in other areas such as those with a Mediterranean climate, showing high variability (Fig. 3). Diversity metrics describing structural heterogeneity in horizontal space, as well as combined metrics, (τ_{CH} ,τ_{CC} , τ_{SW} , τ_{RAO} , τ_{CVH}) also showed considerable variability along precipitation and temperature gradients. With the exception of the convex hull (τ_{CVH}), all metrics displayed low diversity values in hot and dry climates.



CV of vertical profile (τ_{CV})

Skewness of vertical profile (τ_{SK})

Kurtosis of vertical profile (τ_{KU})

σ of Canopy Height (τ_{CH})

σ of Canopy Cover (τ_{CC})

Shannon Entropy (τ_{SW})

Rao Quadratic Entropy (τ_{RAO})

Convex hull volume (τ_{CHV})

Figure 2. Mapped structural diversity at a 10 km resolution, derived from the Random Forest modelling. Each panel illustrates the geographic distribution of a specific metric (see methods for metric details). The colour palette transitions from purple to yellow, denote an increasing gradient of structural diversity, with warmer colours signifying higher values.

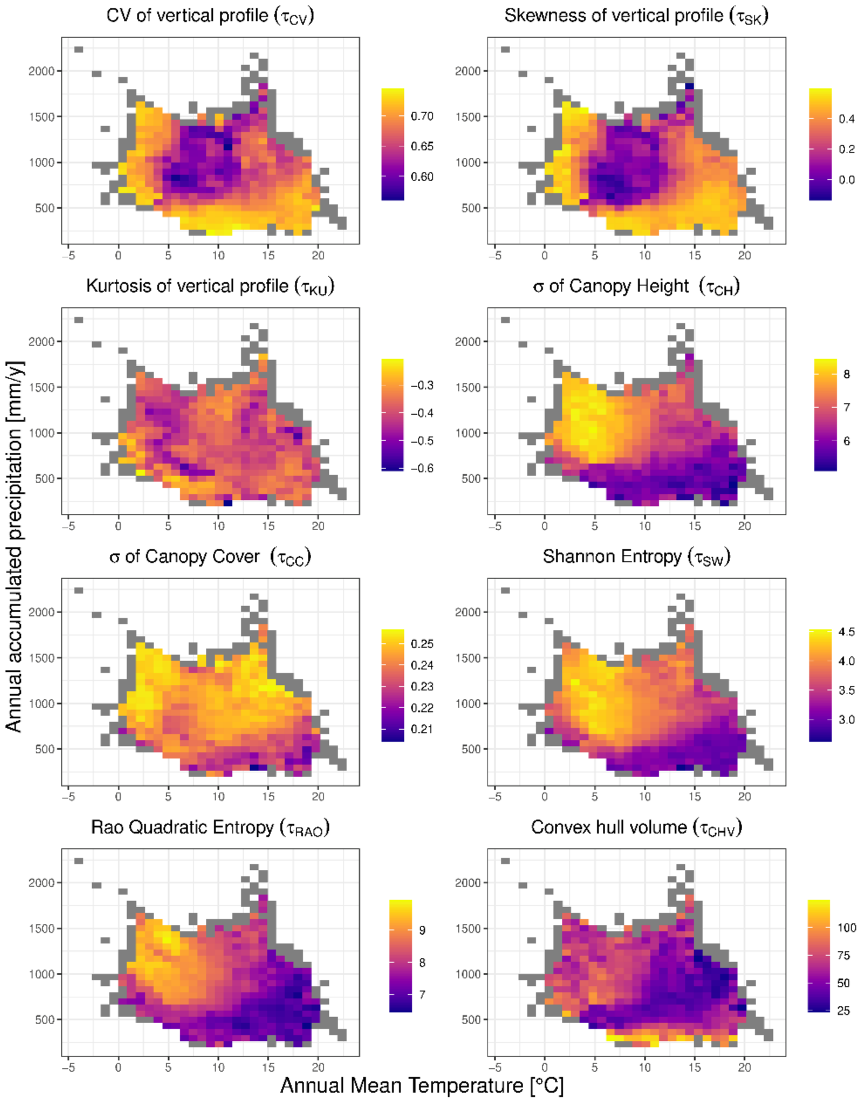


Figure 3. Structural diversity variables mapped against mean climate (temperature and precipitation). The results refer to the dataset at 10 km resolution. Coloured bins depict variation in structural diversity, calculated as the average of the structural diversity values falling within each bin. Grey bins indicate those containing fewer than 5 observations, for which the average was not calculated.

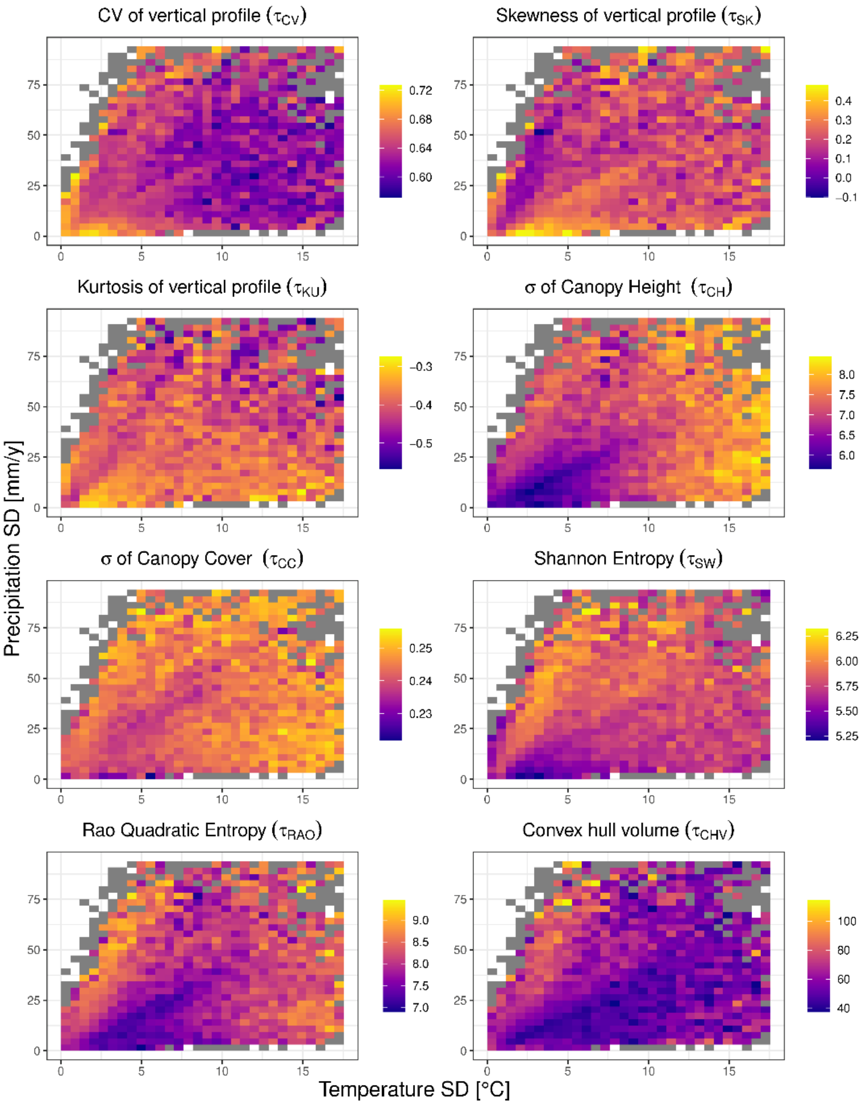


Figure 4. Structural diversity variables mapped against climate variability (temperature and precipitation SD in space). The results refer to the dataset at 10 km resolution. Coloured bins depict variation in structural diversity, calculated as the average of the structural diversity values falling within each bin. Grey bins indicate those containing fewer than 5 observations, for which the average was not calculated.

Specifically, a combination of precipitation levels below ~500 mm and annual mean temperatures above ~10°C (Fig. 3) was associated with the lowest levels of diversity. By contrast, the highest levels of diversity generally occurred in areas with higher precipitation levels (> 500 mm). Patterns of variability in the metrics in climate space for 5km and 1 km resolution (see Fig. S5 and Fig. S6 in the Supplement) dataset were broadly concordant with the 10km dataset, indicating that the results are insensitive to the grain size at which they were calculated.

An examination of the 10 km resolution structural diversity metrics along gradients of temperature and precipitation variability (Fig. 4) revealed distinct but less pronounced patterns compared to those observed in mean climate space (Fig. 3).

Several metrics showed contrasting responses to climate variability. The coefficient of variation of the vertical profile (τ_{CV}) exhibited an inverse pattern, with highest values occurring at low climate variability (low SD in both temperature and precipitation), suggesting that stable climates may promote more heterogeneous vertical canopy structures. In contrast, the standard deviation of canopy height (τ_{CH}) showed the most striking pattern, exhibiting high values when temperature variability exceeded 10°C and precipitation variability fell below 25 mm. Kurtosis (τ_{KU}) showed relatively modest variation across climate variability space, with a weak tendency toward more negative values (more platykurtic distributions) at higher precipitation variability. The standard deviation of canopy cover (τ_{CC}) showed a clear vertical gradient, with values increasing primarily as a function of precipitation variability, relatively independent of temperature variability. Similarly, Shannon entropy (τ_{SW}) was highest at high levels of precipitation variability combined with low to moderate temperature variability. In contrast, Rao's quadratic entropy (τ_{RAO}) exhibited a bimodal pattern, reaching its highest values both at very low precipitation and low temperature and, to a lesser extent, at high precipitation variability. The convex hull volume (τ_{CHV}) showed the most uniform distribution across climate variability space, with generally low values and limited systematic variation. Overall, metrics displayed considerably more scatter along gradients of climatic variability than in mean climate space, suggesting that climate variability plays a more complex role in shaping forest structural diversity than mean climate conditions.

An examination of structural diversity patterns by biogeographic region largely reinforced the patterns revealed in the climate space plots (Fig. 5, see Fig. S7 and Fig. S8 in the Supplement for the 5 km and 1 km datasets), while revealing additional topographic effects. Combined horizontal-vertical diversity metrics (Shannon, Rao) were lowest in the Mediterranean region, consistent with the low diversity observed in hot, arid climates, and highest in Continental, Atlantic, and Alpine regions. Conversely, vertical canopy heterogeneity (Skewness τ_{SK}) was highest in the Mediterranean region, reinforcing the pattern of elevated skewness at climatic extremes. Canopy height variability (τ_{CH}) was markedly elevated in the Alpine region ($M_{ed} = 8.11$) compared to Mediterranean ($M_{ed} = 6.00$), Continental ($M_{ed} = 6.26$), and Atlantic ($M_{ed} = 7.09$), regions likely reflecting greater topographic heterogeneity. CV of vertical profile remained consistent across all regions (~0.65-0.70), while (τ_{KU}) and σ of canopy cover τ_{CC}, showed minimal biogeographic variation.

A Principal Component Analysis (PCA) biplot (Fig. 6D) was used to explore the degree of intercorrelation among the eight structural diversity metrics. The PCA shows that the metrics are distributed across the principal component space with generally wide angular separation among vectors, indicating low to modest correlations. These patterns are consistent across spatial resolutions (Fig. S9, Fig. S11D and Fig. S12D in the Supplement).

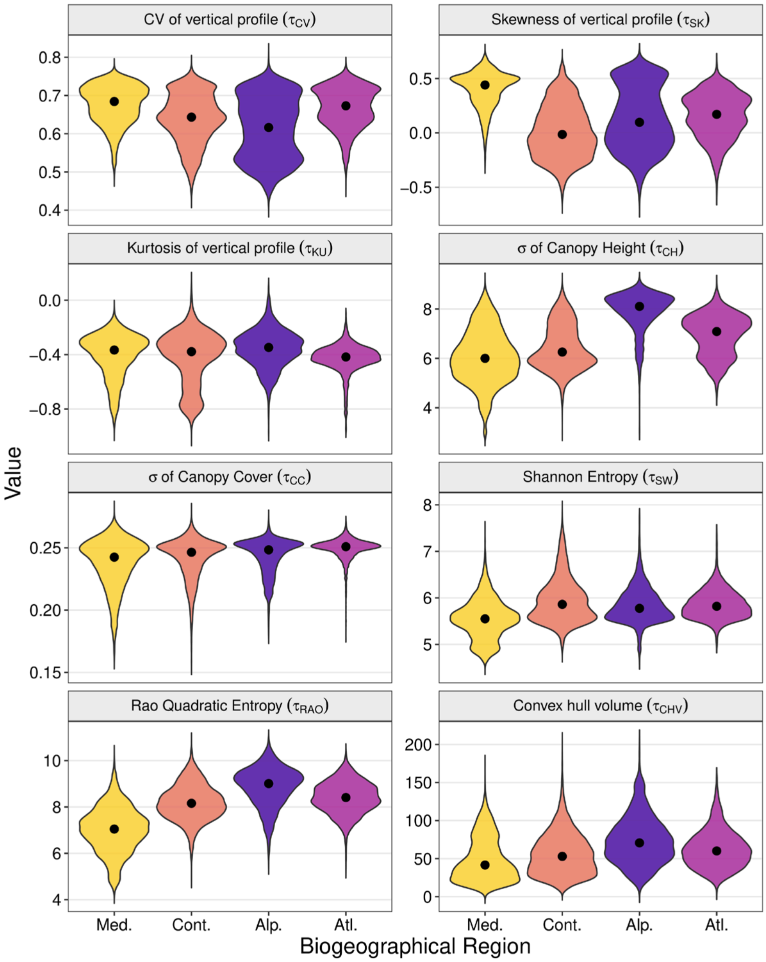


Figure 5. Patterns of variability in forest structural diversity metrics across European biogeographic regions. Violin plots depict the probability density distribution of eight metrics at 10 km resolution. Width represents the kernel density estimate, with black points indicating median values. Colors denote the five biogeographic regions analysed. Med. = Mediterranean region, Cont. = Continental region, Alp. = Alpine region, Atl. = Atlantic region.

3.2 Variable importance and model performance

The final models, derived from the stepwise backward elimination procedure, retained between 7 and 23 predictors, representing the extremes observed across various resolutions of input data and output variable types. The number of selected predictors generally increased with the resolution of the input data (Fig S10 in the Supplement). Models trained for standard deviation of canopy cover (τ_{CC}) and convex hull (τ_{CVH}) retained the highest number of predictors. In contrast, models for skewness (τ_{SK}) and Rao quadratic entropy (τ_{RAO}) retained the lowest number of predictors (Fig S10 in the Supplement).

An examination of the type of predictors selected in the final models highlighted the importance of radar-related predictors, over optical ones as shown in Fig. 6A (see Fig. S11A and Fig. S12A in the Supplement for the 5 km and 1 km datasets). The average proportion of radar-related variables selected across all diversity metrics and resolutions was 0.64, although there was considerable variability. In general, as the resolution of the input dataset increased, the proportion of radar-related variables selected through the feature elimination procedure also increased (Fig. 6A; for the 5 km and 1km datasets see Fig. S11A and Fig. S12A in the Supplement). The diversity variables for which the highest number of radar-related predictors were selected was the convex hull (τ_{CVH}). On the other hand, the one for which the highest number of optical-related predictors were selected, was canopy cover (τ_{CC}).

Among the predictors retained in the final models, texture-related types were the most commonly selected, followed by backscatter, spectral indices, and coherence (Fig. 6B; for the 5 km and 1km datasets see Fig. S11B and Fig. S12B in the Supplement). Notably, texture metrics constituted, on average, the largest proportion of selected variables at a 10 km resolution (Fig 6B). Conversely, the proportion of backscatter-related variables and spectral indices increased in models using the finer resolution input data (Fig. S11B and Fig. S12B in the Supplement).

Model validation revealed that random cross-validation consistently outperformed spatial cross-validation across all resolutions. At 10 km resolution, the model for the Shannon the model for Shannon index τ_{SW} achieved the highest scores, with 0.73 in random validation and 0.64 in spatial validation (Fig 6C and Tables B1 and B2). Conversely, the model with convex hull (τ_{CVH}) as a variable showed the lowest performance, scoring 0.29 in random cross-validation and 0.20 in spatial cross-validation. This difference likely reflects the contrasting statistical properties of the metrics: Shannon entropy integrates information across the full distribution of GEDI observations within a spatial unit and is therefore more robust to sampling variability, whereas convex hull–based metrics are more sensitive to outliers and local sampling density.The best-performing models at 5 km and 1 km differed from those at 10 km, with skewness models (τ_{SK}) yielding the best results at both 5 km and 1 km, while canopy cover variability (τ_{CC}) was the lowest-performing at 1 km and convex hull (τ_{CVH}) at 5 km (Tables B1 and B2; Fig. S11C and Fig. S12C in the Supplement).

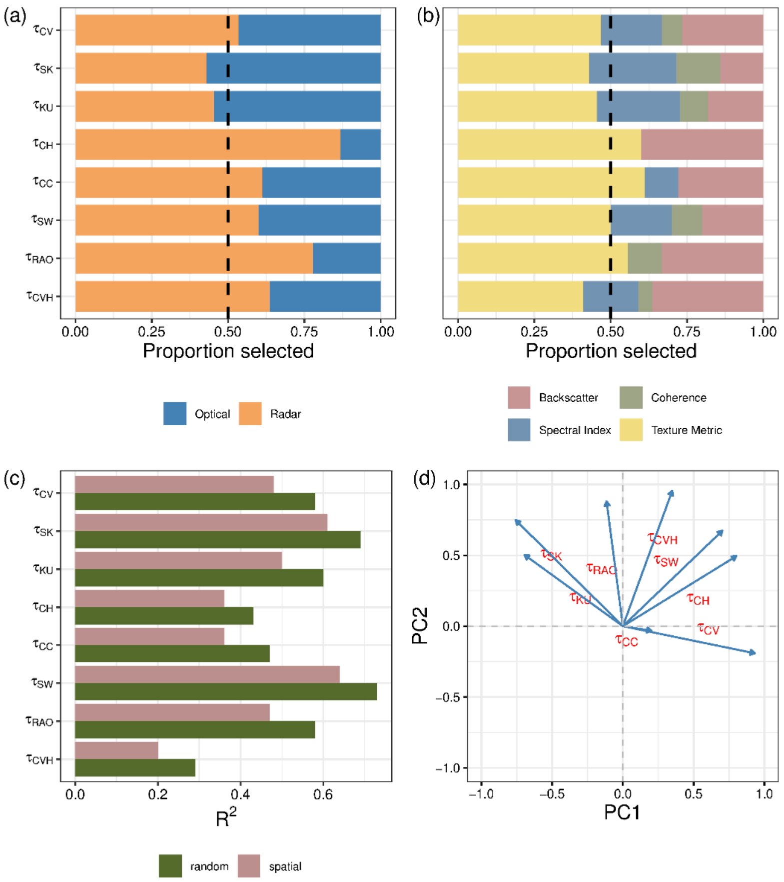


Figure 6. Results of the random forest modelling exercise at 10 km resolution. Panels display the variable selection frequencies (A and B) and model performance, as indicated by the R^2 values derived from two types of validation methods (C). Panel D shows the results of the Principal Component Analysis (PCA) conducted on the estimated structural diversity metrics at this resolution.

Models estimating metrics describing horizontal variability, particularly the standard deviation of canopy cover (τ_{CC}) and convex hull volume (τ_{CVH}), showed lower predictive performance, especially at finer spatial resolutions ($R^2 < 0.30$ at 1 km). These metrics rely more directly on variability among individual GEDI observations within a spatial unit, which is less consistently preserved when locally derived optical and SAR predictors are spatially aggregated to the grid-cell level. In contrast, models estimating metrics describing vertical heterogeneity and combined structural diversity (e.g. τ_{SK}, τ_{CV}, τ_{SW}), particularly at 10 km resolution, exhibited higher validation scores and greater stability across validation schemes.

An examination of the standard deviation of model outputs revealed generally increasing trend of prediction uncertainty across resolutions (Fig. S13, Fig. S14 and Fig. S15 in the Supplement), except in Rao (τ_{RAO}) and convex hull (τ_{CVH}). Overall, uncertainty was generally low across the spatial domain of interest, reflecting limited variability within the ensemble. Notable exceptions occur in the Mediterranean region for the convex hull, kurtosis, and standard deviation of canopy height metrics. Further variability is observed in Eastern Europe, particularly for the convex hull, skewness, kurtosis, Shannon index, and standard deviation of canopy height.

4. Discussion

4.1. Model-based predictions of structural diversity

Our dataset provides eight metrics describing the structural heterogeneity of European forests. To our knowledge, this is the first attempt to comprehensively map forest structural diversity at a quasi-continental scale (because GEDI is unable to observe anything above 50° North). Datasets such as the one presented here contribute to an emerging landscape of data products based on spaceborne LiDAR data, ranging from regional to global scales (e.g. Lang et al., 2023; Shendryk, 2022; Sothe et al., 2022). However, while most efforts have primarily focused on mapping top canopy height, we aimed to create a set of complementary metrics describing the diversity of canopy structure, an ecologically important yet neglected aspect in research.

Some of the ecological indices employed in this study are routinely applied to optical data to quantify landscape-level heterogeneity using multispectral data (Tuanmu and Jetz, 2015). For instance, the Rao and Shannon diversity indices, which can be calculated from spectral indices, have been widely used to quantify the heterogeneity of vegetation and are often proposed as indicators of ecosystem heterogeneity (Rocchini et al., 2021). These heterogeneity indicators have proved to be useful in a variety of contexts, including biodiversity modelling and quantifying the vulnerability of forest ecosystems to disturbances (Forzieri et al., 2021; Taddeo et al., 2021). However, indices based solely on optical data fail to capture crucial aspects of structural heterogeneity, which are related to the three-dimensional arrangement of vegetative elements in the canopy (Fassnacht et al., 2022). Our study addresses a critical gap by introducing the first consistent dataset that maps structural diversity across the forested domain in Europe. This development will contribute to a more detailed and robust regional analysis on ecosystem dynamics, which critically depend on vegetation structure (Migliavacca et al., 2021) and structural diversity (LaRue et al., 2023), and other facets of biodiversity, which requires information on the vertical profile of plants (Fassnacht et al., 2022).

Our findings revealed that model performance differed according to the spatial resolutions and diversity metrics, with several models achieving R^2 values indicative of moderate to strong predictive accuracy, particularly at coarser spatial resolutions (Appendix B, Tables B1 and B2). This variation highlights the critical role of resolution in model performance, indicating that, depending on the application of interest, coarser resolutions may optimize the utility of the models. As expected, spatial cross-validation consistently yielded lower R^2 values than random train-validation random validation across most metrics and resolutions. This outcome reflects the challenges inherent on machine learning methods (Meyer and Pebesma, 2021) of predicting outcomes in areas geographically distinct from the training data. Neverthless, the decrease was generally modest, affirming the broad applicability of our models beyond the training domain.

The recursive feature elimination procedure highlighted the importance of textural variables (Fig. S10 in the Supplement) across diversity metrics and spatial resolutions. Entropy, derived from ALOS-PALSAR-2 data, stood out as the most influential variable, corroborating research that demonstrates textural metrics' effectiveness in capturing spatial heterogeneity in structural diversity (Bae et al., 2019). Additionally, the significant role of coherence, which aligns with evidence of its predictive power for forest structural features (Bruggisser et al., 2021; Cartus et al., 2022), suggests its potential in reflecting changes in forest structural density and composition. Collectively, our findings underscore the benefits of integrating various sensor data to enhance the prediction of structural diversity, as evidenced by the diverse contributions of optical and radar-based predictors.

4.2. Potential applications

We envisage that our structural diversity dataset will significantly advance future research and practical applications across several disciplines. We identify three key areas where the dataset could be utilised.

Firstly, the dataset could aid in the development of different biodiversity indicators. Ecosystem structure has been identified as an Essential Biodiversity Variable (EBV) (Valbuena et al., 2020), and a wide range of studies have shown a strong correlation between LiDAR-based metrics and ground-based biodiversity measurements (Marselis et al., 2020). The metrics developed here could be used to identify areas with unique structural features that harbour high levels of biodiversity. Furthermore, integrating them with data from other sensors, such as Sentinel 1 and Sentinel 2, offers a promising avenue for generating accurate spatial explicit estimates of different indicators, thus paving the way for the development of frameworks for monitoring long-term biodiversity changes.

Secondly, the dataset offers a valuable resource for quantifying the observed impacts of global change drivers on the functioning of European forest ecosystems. The increasing recognition of the role of structural diversity in driving ecosystem processes (Ali et al., 2016; Aponte et al., 2020; Listopad et al., 2015) underscores the importance of our metrics. Consequently, our dataset provides a crucial tool for enabling comprehensive, data-driven assessments of the impact of climate and land cover changes on the functioning of forest ecosystems across large scales, addressing the previous limitations posed by the unavailability of structural diversity data over extensive spatial scales.

Thirdly, the dataset could be used for improving Earth system models. Historically, plant canopy structure has not been adequately represented in these models (Atkins et al., 2018; Schneider et al., 2020). This lack of detailed representation can lead to significant errors in predicting energy balance, carbon cycling, and ecosystem responses to environmental changes (Duveiller et al., 2023). Integrating structural diversity into these models has the potential to enhance the accuracy of simulations by incorporating more realistic representations of light interception, photosynthetic rates, and energy fluxes. In particular, these applications are directly relevant to contemporary Earth system modelling frameworks such as CMIP6 and forthcoming CMIP7 simulations, which underpin IPCC climate assessments and projections.

5. Data availability

The structural diversity metrics generated in this study can be accessed at Figshare: https://doi.org/10.6084/m9.figshare.26058868 (Girardello et al., 2024). All maps are available at three spatial resolutions (1 km, 5 km, and 10 km) in the EPSG:3035 (LAEA) spatial reference system. All eight at all three spatial resolutions are provided; users are encouraged to consult the validation results (Tables B1 and B2) to assess the suitability of individual metrics for specific applications.

6. Code availability

Google Earth Engine code for data preparation and data for reproducing the figures are available at https://github.com/drmarcogir/structuraldiversity

7. Conclusions

We generated a spatially-explicit dataset on eight forest structural diversity metrics at multiple resolutions (10km, 5km, 1km) encompassing temperate, Mediterranean, and continental regions of Europe. Models developed to create the dataset were robust. The dataset generated in our study represents a novel contribution to the Essential Biodiversity Variables (EBV) framework, and the metrics can be used in various applications, ranging from the study of biodiversity to ecosystem functioning. We conclude that combining GEDI data with those from other satellite sensors paves the way for developing a consistent and scalable framework to monitor structural diversity across Europe.

Appendices

Appendix A: Supplementary Methods

A1 Statistical indicators

The statistical indicators used in this study are detailed below. The mean μ, standard deviation σ, skewness γ, excess kurtosis κ, coefficient of variation cv of a variable $X = \{x_1, \ \ldots, \ x_N\}$ are defined as:

$$\mu(X) = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sigma(X) = \left\{\frac{1}{N}\sum_{i=1}^{N}[x_i - \mu(X)]^2\right\}$$

$$\gamma(X) = \frac{\sum_{i=1}^{N}[x_i - \mu(X)]^3}{[\sigma(X)]^3}$$

$$\kappa(X) = \frac{\sum_{i=1}^{N}[x_i - \mu(X)]^4}{[\sigma(X)]^4}$$

$$cv(X) = \frac{\sigma(X)}{\mu(X)} \qquad \text{(A1)}$$

A2 Binning in cartesian 4d space

$p_{\varepsilon\pi o\omega}$indicates the fraction of the GEDI shots falling in the bin identified by the indices $(\varepsilon, \pi, o, \omega)$ in the 4D cartesian space defined on the basis$(e^{\varepsilon}, e^{\pi}, e^{o}, e^{\omega})$, see Figure S2, with

$$\textstyle\sum_{\varepsilon\pi o\omega} p_{\varepsilon\pi o\omega} = 1 \qquad \text{(A2)}$$

where $\sum_{\varepsilon\pi o\omega}$ stands for $\sum_{\varepsilon=1}^{N_{bins}^{e^{\varepsilon}}}\sum_{\pi=1}^{N_{bins}^{e^{\pi}}}\sum_{o=1}^{N_{bins}^{e^{o}}}\sum_{\omega=1}^{N_{bin}^{e^{\omega}}}$, with $N_{bins}^{e^{\varepsilon}}$ number of bins in the e^{ε} dimension, and $D_{\varepsilon\pi o\omega}^{\varepsilon'\pi' o'\omega'}$ indicates the cartesian distance between $(\varepsilon, \pi, o, \omega)$ and $(\varepsilon', \pi', o', \omega')$ bin.

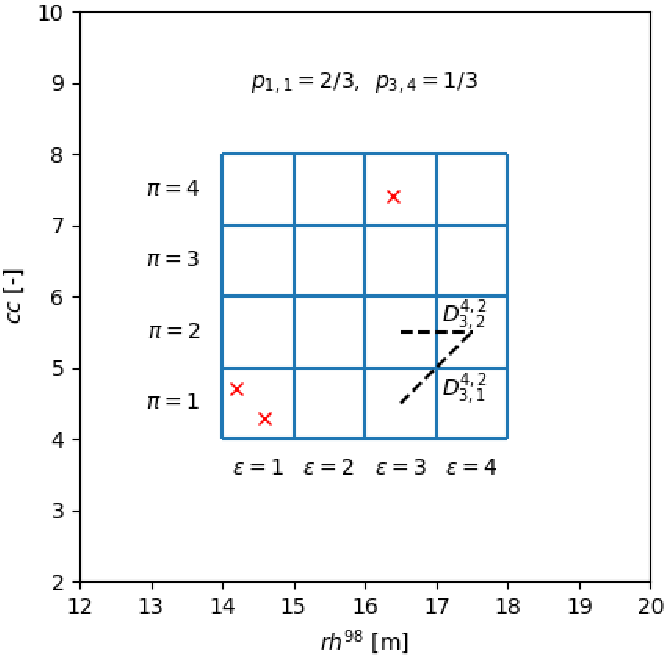


Figure A1. example of $p_{\varepsilon\pi}$ and $D_{\varepsilon\pi}^{\varepsilon'\pi'}$ estimation in the 2D cartesian space defined on the basis $(rh^{98},\ cc)$. The GEDI shots are reported with the red X, GEDI cc values have been amplified by 10.

A3 Predictor calculation

Starting with appropriate bands/indices (step 2 of the workflow in the main text), the four scalars $\phi_{\alpha,i}^{\beta}$, where $\beta \in \{SM,\ ASM,\ ENT,\ DISS\}$, are calculated from the cluster of 7x7 pixels $\phi_{\alpha,i}(p,q)$ overlapping the footprint of the GEDI shot i, where p and q represent the pixel indices within the window. In details, we calculated:

1. the spatial mean (SM)

$$\phi_{\alpha,i}^{SM} = \mu\left(\phi_{\alpha,i}(p,q)\right) \qquad \text{(A3)}$$

which is performed to compensate for potential footprint geolocation inaccuracies, and reduce the presence of noise, and three texture metrics. Texture metrics provide spatial content information (Nichol and Sarker 2011), and are highly effective in capturing the pixels heterogeneity. Defining $\overline{\phi}_{\alpha,i}(p,q)$ as the grey-levels matrix, which is calculated from $\phi_{\alpha,i}(p,q)$ by normalizing the values* within the range of [0, 1] based on the 1st and 99th percentiles, $C_{\alpha,i}(m,n)$ as the corresponding grey-levels co-occurrence matrix (GLCM), with dimension 256x256 (Haralick et al. 1973):

$$C_{\alpha,i}(m,n) = \sum_{p=1}^{7}\sum_{q=1}^{6} 1\,,\ if\ \overline{\phi}_{\alpha,i}(p,q) = m\ and\ \overline{\phi}_{\alpha,i}(p,q+1) = q;\ 0,\ otherwise$$

and $p_{\alpha,i}(m,n)$ as the probability that grey-level m occurs close to the grey-level n :

$$p_{\alpha,i}(m,n) = \frac{C_{\alpha,i}(m,n)}{\sum_{p=0}^{255}\sum_{q=0}^{255} C_{\alpha,i}(p,q)} \qquad \text{(A4)}$$

we calculated:

2. the angular second moment (ASM)

$$\phi_{\alpha,i}^{ASM} = -\sum_{m=0}^{255}\sum_{n=0}^{255}\left[p_{\alpha,i}(m,n)\right]^2 \qquad \text{(A5)}$$

ASM is a measure of the homogeneity or uniformity of pixel values within a neighbourhood. It reflects the degree to which pixel values deviate from the mean, providing insights into the texture's smoothness or roughness;

3. the entropy

$$\phi_{\alpha,i}^{ENT} = -\sum_{m=0}^{255}\sum_{n=0}^{255} p_{\alpha,i}(m,n)\log p_{\alpha,i}\,(m,n) \qquad \text{(A6)}$$

Entropy is a measure of the randomness or disorder in the distribution of grey levels. It quantifies image non-uniformity, with higher entropy values indicating a more random distribution of pixel values within a neighbourhood;

4. the dissimilarity index

$$\phi_{\alpha,i}^{DISS} = \sum_{m=0}^{255}\sum_{n=0}^{255} p_{\alpha,i}(m,n)|m-n|ogp_{\alpha,i}(m,n) \qquad \text{(A7)}$$

Dissimilarity measures the complexity and the nature of grey-level transitions among neighbouring pixels (Conners et al. 1984). It quantifies image contrast, with higher dissimilarity values reflecting pronounced differences among neighbouring pixel values.

* For Sentinel-2 data, we retained only pixels with NDVI values greater than 0, as values below 0 are more likely to represent non-vegetative features.

Appendix B: Model validation results

Table B1. Results of the random validation procedure conducted for the forest structural metrics at three spatial resolutions: 1x1km, 5x5km, and 10x10km. The validation outcomes are presented in terms of the coefficient of determination (R^2), which quantifies the proportion of the variance in the dependent variable that is predictable from the independent variables.

Metric	1 km	5 km	10 km
CV of vertical profile (τ_{CV})	0.36	0.51	0.58
Skewness of vertical profile (τ_{SK})	0.47	0.64	0.69
Kurtosis of vertical profile (τ_{KU})	0.28	0.48	0.6
σ of Canopy Height (τ_{CH})	0.26	0.39	0.43
σ of Canopy Cover (τ_{CC})	0.16	0.37	0.47
Shannon Entropy (τ_{SW})	0.39	0.63	0.73
Rao Quadratic Entropy (τ_{RAO})	0.32	0.52	0.58
Convex Hull Volume (τ_{CHV})	0.26	0.34	0.29

Table B2. Results of the spatial cross-validation procedures conducted for the forest structural metrics at three spatial resolutions: 1x1km, 5x5km, and 10x10km. The validation outcomes are presented in terms of the coefficient of determination (R^2).

Metric	1 km	5 km	10 km
CV of vertical profile (τ_{CV})	0.33	0.43	0.48
Skewness of vertical profile (τ_{SK})	0.43	0.57	0.61
Kurtosis of vertical profile (τ_{KU})	0.24	0.41	0.5
σ of Canopy Height (τ_{CH})	0.25	0.34	0.36
σ of Canopy Cover (τ_{CC})	0.14	0.29	0.36
Shannon Entropy (τ_{SW})	0.36	0.55	0.64
Rao Quadratic Entropy (τ_{RAO})	0.29	0.44	0.47
Convex Hull Volume (τ_{CHV})	0.2	0.25	0.2

Author contribution MGir, GO and AC conceived the ideas with contributions from MM, GC, and MPicc. MGar, MPick, and AE contributed to the discussion on metric development and interpretation. MGir, GO, and MPicc collated and analysed the data. MGir led the writing with inputs from MPicc and GO. All authors contributed to the revision of the manuscript and approved the final version.

Competing interests The authors declare no competing financial interests

Acknowledgements The study was funded by the Exploratory Project ForBioRes of the European Commission, Joint Research Centre

References

Ali, A., Yan, E.-R., Chen, H. Y. H., Chang, S. X., Zhao, Y.-T., Yang, X.-D., and Xu, M.-S.: Stand structural diversity rather than species diversity enhances aboveground carbon storage in secondary subtropical forests in Eastern China, Biogeosciences, 13, 4627–4635, https://doi.org/10.5194/bg-13-4627-2016, 2016.

Altmann, A., Toloşi, L., Sander, O., and Lengauer, T.: Permutation importance: a corrected feature importance measure, Bioinformatics, 26, 1340–1347, https://doi.org/10.1093/bioinformatics/btq134, 2010.

Aponte, C., Kasel, S., Nitschke, C. R., Tanase, M. A., Vickers, H., Parker, L., Fedrigo, M., Kohout, M., Ruiz-Benito, P., Zavala, M. A., and Bennett, L. T.: Structural diversity underpins carbon storage in Australian temperate forests, Global Ecology and Biogeography, 29, 789–802, https://doi.org/10.1111/geb.13038, 2020.

Aragoneses, E., García, M., Ruiz-Benito, P., and Chuvieco, E.: Mapping forest canopy fuel parameters at European scale using spaceborne LiDAR and satellite data, Remote Sens Environ, 303, 114005, https://doi.org/10.1016/j.rse.2024.114005, 2024.

Atkins, J. W., Fahey, R. T., Hardiman, B. S., and Gough, C. M.: Forest Canopy Structural Complexity and Light Absorption Relationships at the Subcontinental Scale, J Geophys Res Biogeosci, 123, 1387–1405, https://doi.org/10.1002/2017JG004256, 2018.

Bae, S., Levick, S. R., Heidrich, L., Magdon, P., Leutner, B. F., Wöllauer, S., Serebryanyk, A., Nauss, T., Krzystek, P., Gossner, M. M., Schall, P., Heibl, C., Bässler, C., Doerfler, I., Schulze, E.-D., Krah, F.-S., Culmsee, H., Jung, K., Heurich, M., Fischer, M., Seibold, S., Thorn, S., Gerlach, T., Hothorn, T., Weisser, W. W., and Müller, J.: Radar vision in the mapping of forest biodiversity from space, Nat Commun, 10, 4757, https://doi.org/10.1038/s41467-019-12737-x, 2019.

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

Bruggisser, M., Dorigo, W., Dostálová, A., Hollaus, M., Navacchi, C., Schlaffer, S., and Pfeifer, N.: Potential of Sentinel-1 C-Band Time Series to Derive Structural Parameters of Temperate Deciduous Forests, Remote Sens (Basel), 13, 798, https://doi.org/10.3390/rs13040798, 2021.

Cartus, O., Santoro, M., Wegmuller, U., Labriere, N., and Chave, J.: Sentinel-1 Coherence for Mapping Above-Ground Biomass in Semiarid Forest Areas, IEEE Geoscience and Remote Sensing Letters, 19, 1–5, https://doi.org/10.1109/LGRS.2021.3071949, 2022.

Coops, N. C., Tompalski, P., Goodbody, T. R. H., Queinnec, M., Luther, J. E., Bolton, D. K., White, J. C., Wulder, M. A., van Lier, O. R., and Hermosilla, T.: Modelling lidar-derived estimates of forest attributes over space and time: A review of approaches and future trends, Remote Sens Environ, 260, 112477, https://doi.org/10.1016/j.rse.2021.112477, 2021.

Coverdale, T. C. and Davies, A. B.: Unravelling the relationship between plant diversity and vegetation structural complexity: A review and theoretical framework, Journal of Ecology, 111, 1378–1395, https://doi.org/10.1111/1365-2745.14068, 2023.

Crockett, E. T. H., Atkins, J. W., Guo, Q., Sun, G., Potter, K. M., Ollinger, S., Silva, C. A., Tang, H., Woodall, C. W., Holgerson, J., and Xiao, J.: Structural and species diversity explain aboveground carbon storage in forests across the United States: Evidence from GEDI and forest inventory data, Remote Sens Environ, 295, 113703, https://doi.org/10.1016/j.rse.2023.113703, 2023.

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P. L., Qi, W., and Silva, C.: The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography, Science of Remote Sensing, 1, 100002, https://doi.org/10.1016/j.srs.2020.100002, 2020a.

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., and others: The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography, Science of Remote Sensing, 1, 100002, 2020b.

Duveiller, G., Pickering, M., Muñoz-Sabater, J., Caporaso, L., Boussetta, S., Balsamo, G., and Cescatti, A.: Getting the leaves right matters for estimating temperature extremes, Geosci Model Dev, 16, 7357–7373, https://doi.org/10.5194/gmd-16-7357-2023, 2023.

Ehbrecht, M., Seidel, D., Annighöfer, P., Kreft, H., Köhler, M., Zemp, D. C., Puettmann, K., Nilus, R., Babweteera, F., Willim, K., Stiers, M., Soto, D., Boehmer, H. J., Fisichelli, N., Burnett, M., Juday, G., Stephens, S. L., and Ammer, C.: Global patterns and climatic controls of forest structural complexity, Nat Commun, 12, 519, https://doi.org/10.1038/s41467-020-20767-z, 2021.

FAO: On definitions of forest and forest change, 2000.

Fassnacht, F. E., Müllerová, J., Conti, L., Malavasi, M., and Schmidtlein, S.: About the link between biodiversity and spectral variation, Appl Veg Sci, 25, https://doi.org/10.1111/avsc.12643, 2022.

Forzieri, G., Girardello, M., Ceccherini, G., Spinoni, J., Feyen, L., Hartmann, H., Beck, P. S. A., Camps-Valls, G., Chirici, G., Mauri, A., and Cescatti, A.: Emergent vulnerability to climate-driven disturbances in European forests, Nat Commun, 12, 1081, https://doi.org/10.1038/s41467-021-21399-7, 2021.

Gao, B.: NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space, Remote Sens Environ, 58, 257–266, https://doi.org/10.1016/S0034-4257(96)00067-3, 1996.

Girardello, M., Oton, G., Piccardo, M., and Ceccherini, G.: A dataset on the structural diversity of European forests, https://doi.org/10.6084/m9.figshare.26058868.v1, 2024.

Gitelson, A. and Merzlyak, M. N.: Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves, J Photochem Photobiol B, 22, 247–252, https://doi.org/10.1016/1011-1344(93)06963-4, 1994.

Gitelson, A. A. and Merzlyak, M. N.: Remote sensing of chlorophyll concentration in higher plant leaves, Advances in Space Research, 22, 689–692, https://doi.org/10.1016/S0273-1177(97)01133-2, 1998.

Goodbody, T. R. H., Coops, N. C., Queinnec, M., White, J. C., Tompalski, P., Hudak, A. T., Auty, D., Valbuena, R., LeBoeuf, A., Sinclair, I., McCartney, G., Prieur, J.-F., and Woods, M. E.: sgsR : a structurally guided sampling toolbox for LiDAR-based forest inventories, Forestry, 96, 411–424, https://doi.org/10.1093/forestry/cpac055, 2023.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R.: Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sens Environ, 202, https://doi.org/10.1016/j.rse.2017.06.031, 2017.

Gough, C. M., Atkins, J. W., Fahey, R. T., and Hardiman, B. S.: High rates of primary production in structurally complex forests, Ecology, 100, https://doi.org/10.1002/ecy.2864, 2019.

Hakkenberg, C. R. and Goetz, S. J.: Climate mediates the relationship between plant biodiversity and forest structure across the United States, Global Ecology and Biogeography, 30, 2245–2258, https://doi.org/10.1111/geb.13380, 2021.

Hakkenberg, C. R., Atkins, J. W., Brodie, J. F., Burns, P., Cushman, S., Jantz, P., Kaszta, Z., Quinn, C. A., Rose, M. D., and Goetz, S. J.: Inferring alpha, beta, and gamma plant diversity across biomes with GEDI spaceborne lidar, Environmental Research: Ecology, 2, 035005, https://doi.org/10.1088/2752-664X/acffcd, 2023.

Hancock, S., McGrath, C., Lowe, C., Davenport, I., and Woodhouse, I.: Requirements for a global lidar system: spaceborne lidar with wall-to-wall coverage, R Soc Open Sci, 8, https://doi.org/10.1098/rsos.211166, 2021.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., and Townshend, J. R. G.: High-Resolution Global Maps of 21st-Century Forest Cover Change, Science (1979), 342, 850–853, https://doi.org/10.1126/science.1244693, 2013.

Holcomb, A., Burns, P., Keshav, S., and Coomes, D. A.: Repeat GEDI footprints measure the effects of tropical forest disturbances, Remote Sens Environ, 308, 114174, https://doi.org/10.1016/j.rse.2024.114174, 2024.

Kellndorfer, J., Cartus, O., Lavalle, M., Magnard, C., Milillo, P., Oveisgharan, S., Osmanoglu, B., Rosen, P. A., and Wegmüller, U.: Global seasonal Sentinel-1 interferometric coherence and backscatter data set, Sci Data, 9, 73, https://doi.org/10.1038/s41597-022-01189-6, 2022.

Lang, N., Jetz, W., Schindler, K., and Wegner, J. D.: A high-resolution canopy height model of the Earth, Nat Ecol Evol, 7, 1778–1789, https://doi.org/10.1038/s41559-023-02206-6, 2023.

Larue, E. A., Hardiman, B. S., Elliott, J. M., and Fei, S.: Structural diversity as a predictor of ecosystem function, Environmental Research Letters, 14, https://doi.org/10.1088/1748-9326/ab49bb, 2019.

LaRue, E. A., Hardiman, B. S., Elliott, J. M., and Fei, S.: Structural diversity as a predictor of ecosystem function, Environmental Research Letters, 14, 114011, https://doi.org/10.1088/1748-9326/ab49bb, 2019.

LaRue, E. A., Knott, J. A., Domke, G. M., Chen, H. Y., Guo, Q., Hisano, M., Oswalt, C., Oswalt, S., Kong, N., Potter, K. M., and Fei, S.: Structural diversity as a reliable and novel predictor for ecosystem productivity, Front Ecol Environ, 21, 33–39, https://doi.org/10.1002/fee.2586, 2023.

Listopad, C. M. C. S., Masters, R. E., Drake, J., Weishampel, J., and Branquinho, C.: Structural diversity indices based on airborne LiDAR as ecological indicators for managing highly dynamic landscapes, Ecol Indic, 57, 268–279, https://doi.org/10.1016/j.ecolind.2015.04.017, 2015.

Liu, C., Gong, W., Shi, S., Wang, T., Xu, T., Shi, Z., and Niu, J.: Deep learning-driven forest canopy height mapping in boreal regions through multi-source remote sensing fusion: Integrating Sentinel-1/2, PALSAR, and ICESat-2/LVIS data, International Journal of Applied Earth Observation and Geoinformation, 143, 104766, https://doi.org/10.1016/j.jag.2025.104766, 2025.

Ma, Q., Su, Y., Hu, T., Jiang, L., Mi, X., Lin, L., Cao, M., Wang, X., Lin, F., Wang, B., Sun, Z., Wu, J., Ma, K., and Guo, Q.: The coordinated impact of forest internal structural complexity and tree species diversity on forest productivity across forest biomes, Fundamental Research, https://doi.org/10.1016/j.fmre.2022.10.005, 2022.

Marselis, S. M., Abernethy, K., Alonso, A., Armston, J., Baker, T. R., Bastin, J., Bogaert, J., Boyd, D. S., Boeckx, P., Burslem, D. F. R. P., Chazdon, R., Clark, D. B., Coomes, D., Duncanson, L., Hancock, S., Hill, R., Hopkinson, C., Kearsley, E., Kellner, J. R., Kenfack, D., Labrière, N., Lewis, S. L., Minor, D., Memiaghe, H., Monteagudo, A., Nilus, R., O'Brien, M., Phillips, O. L., Poulsen, J., Tang, H., Verbeeck, H., and Dubayah, R.: Evaluating the potential of full-waveform lidar for mapping pan-tropical tree species richness, Global Ecology and Biogeography, 29, 1799–1816, https://doi.org/10.1111/geb.13158, 2020.

Meyer, H. and Pebesma, E.: Predicting into unknown space? Estimating the area of applicability of spatial prediction models, Methods Ecol Evol, 12, 1620–1633, https://doi.org/10.1111/2041-210X.13650, 2021.

Migliavacca, M., Musavi, T., Mahecha, M. D., Nelson, J. A., Knauer, J., Baldocchi, D. D., Perez-Priego, O., Christiansen, R., Peters, J., Anderson, K., Bahn, M., Black, T. A., Blanken, P. D., Bonal, D., Buchmann, N., Caldararu, S., Carrara, A., Carvalhais, N., Cescatti, A., Chen, J., Cleverly, J., Cremonese, E., Desai, A. R., El-Madany, T. S., Farella, M. M., Fernández-Martínez, M., Filippa, G., Forkel, M., Galvagno, M., Gomarasca, U., Gough, C. M., Göckede, M., Ibrom, A., Ikawa, H., Janssens, I. A., Jung, M., Kattge, J., Keenan, T. F., Knohl, A., Kobayashi, H., Kraemer, G., Law, B. E., Liddell, M. J., Ma, X., Mammarella, I., Martini, D., Macfarlane, C., Matteucci, G., Montagnani, L., Pabon-Moreno, D. E., Panigada, C., Papale, D., Pendall, E., Penuelas, J., Phillips, R. P., Reich, P. B., Rossini, M., Rotenberg, E., Scott, R. L., Stahl, C., Weber, U., Wohlfahrt, G., Wolf, S., Wright, I. J., Yakir, D., Zaehle, S., and Reichstein, M.: The three major axes of terrestrial ecosystem function, Nature, 598, 468–472, https://doi.org/10.1038/s41586-021-03939-9, 2021.

Mueller, M. M., Dubois, C., Jagdhuber, T., Hellwig, F. M., Pathe, C., Schmullius, C., and Steele-Dunne, S.: Sentinel-1 Backscatter Time Series for Characterization of Evapotranspiration Dynamics over Temperate Coniferous Forests, Remote Sens (Basel), 14, 6384, https://doi.org/10.3390/rs14246384, 2022.

Murphy, B. A., May, J. A., Butterworth, B. J., Andresen, C. G., and Desai, A. R.: Unraveling Forest Complexity: Resource Use Efficiency, Disturbance, and the Structure-Function Relationship, J Geophys Res Biogeosci, 127, https://doi.org/10.1029/2021JG006748, 2022.

Naidoo, L., Mathieu, R., Main, R., Kleynhans, W., Wessels, K., Asner, G., and Leblon, B.: Savannah woody structure modelling and mapping using multi-frequency (X-, C- and L-band) Synthetic Aperture Radar data, ISPRS Journal of Photogrammetry and Remote Sensing, 105, 234–250, https://doi.org/10.1016/j.isprsjprs.2015.04.007, 2015.

Pan, J., Zhao, R., Xu, Z., Cai, Z., and Yuan, Y.: Quantitative estimation of sentinel-1A interferometric decorrelation using vegetation index, Front Earth Sci (Lausanne), 10, https://doi.org/10.3389/feart.2022.1016491, 2022.

Perrone, M., Conti, L., Galland, T., Komárek, J., Lagner, O., Torresani, M., Rossi, C., Carmona, C. P., de Bello, F., Rocchini, D., Moudrý, V., Šímová, P., Bagella, S., and Malavasi, M.: "Flower power": How flowering affects spectral diversity metrics and their relationship with plant diversity, Ecol Inform, 81, 102589, https://doi.org/10.1016/j.ecoinf.2024.102589, 2024.

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C. E., Armston, J., Dubayah, R., Blair, J. B., and Hofton, M.: Mapping global forest canopy height through integration of GEDI and Landsat data, Remote Sens Environ, 253, 112165, https://doi.org/10.1016/j.rse.2020.112165, 2021.

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., and Sorooshian, S.: A modified soil adjusted vegetation index, Remote Sens Environ, 48, 119–126, https://doi.org/10.1016/0034-4257(94)90134-1, 1994.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, Ecography, n/a--n/a, https://doi.org/10.1111/ecog.02881, 2016.

Rocchini, D., Thouverai, E., Marcantonio, M., Iannacito, M., Da Re, D., Torresani, M., Bacaro, G., Bazzichetto, M., Bernardi, A., Foody, G. M., Furrer, R., Kleijn, D., Larsen, S., Lenoir, J., Malavasi, M., Marchetto, E., Messori, F., Montaghi, A., Moudrý, V., Naimi, B., Ricotta, C., Rossini, M., Santi, F., Santos, M. J., Schaepman, M. E., Schneider, F. D., Schuh, L., Silvestri, S., Ŝímová, P., Skidmore, A. K., Tattoni, C., Tordoni, E., Vicario, S., Zannini, P., and Wegmann, M.: rasterdiv—An Information Theory tailored R package for measuring ecosystem heterogeneity from space: To the origin and back, Methods Ecol Evol, 12, 1093–1102, https://doi.org/10.1111/2041-210X.13583, 2021.

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., and others: Monitoring vegetation systems in the Great Plains with ERTS, NASA Spec. Publ, 351, 309, 1974.

Schneider, F. D., Ferraz, A., Hancock, S., Duncanson, L. I., Dubayah, R. O., Pavlick, R. P., and Schimel, D. S.: Towards mapping the diversity of canopy structure from space with GEDI, Environmental Research Letters, 15, https://doi.org/10.1088/1748-9326/ab9e99, 2020.

Schwartz, M., Ciais, P., Ottlé, C., De Truchis, A., Vega, C., Fayad, I., Brandt, M., Fensholt, R., Baghdadi, N., Morneau, F., Morin, D., Guyon, D., Dayau, S., and Wigneron, J.-P.: High-resolution canopy height map in the Landes forest (France) based on GEDI, Sentinel-1, and Sentinel-2 data with a deep learning approach, International Journal of Applied Earth Observation and Geoinformation, 128, 103711, https://doi.org/10.1016/j.jag.2024.103711, 2024.

Shendryk, Y.: Fusing GEDI with earth observation data for large area aboveground biomass mapping, International Journal of Applied Earth Observation and Geoinformation, 115, 103108, https://doi.org/10.1016/j.jag.2022.103108, 2022.

Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R., and Lucas, R.: New global forest/non-forest maps from ALOS PALSAR data (2007–2010), Remote Sens Environ, 155, 13–31, https://doi.org/10.1016/j.rse.2014.04.014, 2014.

Shugart, H. H., Saatchi, S., and Hall, F. G.: Importance of structure and its measurement in quantifying function of forest ecosystems, J Geophys Res Biogeosci, 115, https://doi.org/10.1029/2009JG000993, 2010.

Sothe, C., Gonsamo, A., Lourenço, R. B., Kurz, W. A., and Snider, J.: Spatially Continuous Mapping of Forest Canopy Height in Canada by Combining GEDI and ICESat-2 with PALSAR and Sentinel, Remote Sens (Basel), 14, 5158, https://doi.org/10.3390/rs14205158, 2022.

Sun, J., Yu, X., Wang, H., Jia, G., Zhao, Y., Tu, Z., Deng, W., Jia, J., and Chen, J.: Effects of forest structure on hydrological processes in China, J Hydrol (Amst), 561, 187–199, https://doi.org/10.1016/j.jhydrol.2018.04.003, 2018.

Taddeo, S., Dronova, I., and Harris, K.: Greenness, texture, and spatial relationships predict floristic diversity across wetlands of the conterminous United States, ISPRS Journal of Photogrammetry and Remote Sensing, 175, 236–246, https://doi.org/10.1016/j.isprsjprs.2021.03.012, 2021.

Toda, M., Knohl, A., Luyssaert, S., and Hara, T.: Simulated effects of canopy structural complexity on forest productivity, For Ecol Manage, 538, 120978, https://doi.org/10.1016/j.foreco.2023.120978, 2023.

Tuanmu, M. and Jetz, W.: A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modelling, Global Ecology and Biogeography, 24, 1329–1339, https://doi.org/10.1111/geb.12365, 2015.

Valbuena, R., O'Connor, B., Zellweger, F., Simonson, W., Vihervaara, P., Maltamo, M., Silva, C. A., Almeida, D. R. A., Danks, F., Morsdorf, F., Chirici, G., Lucas, R., Coomes, D. A., and Coops, N. C.: Standardizing Ecosystem Morphological Traits from 3D Information Sources, Trends Ecol Evol, 35, 656–667, https://doi.org/10.1016/j.tree.2020.03.006, 2020.

Vollrath, A., Mullissa, A., and Reiche, J.: Angular-Based Radiometric Slope Correction for Sentinel-1 on Google Earth Engine, Remote Sens (Basel), 12, 1867, https://doi.org/10.3390/rs12111867, 2020.

Wang, C., Zhang, W., Ji, Y., Marino, A., Li, C., Wang, L., Zhao, H., and Wang, M.: Estimation of Aboveground Biomass for Different Forest Types Using Data from Sentinel-1, Sentinel-2, ALOS PALSAR-2, and GEDI, Forests, 15, 215, https://doi.org/10.3390/f15010215, 2024.Zhai, L., Will, R. E., and Zhang, B.: Structural diversity is better associated with forest productivity than species or functional diversity, Ecology, 105, https://doi.org/10.1002/ecy.4269, 2024.