

1 **A dataset on the structural diversity of European forests**

2

3 Marco Girardello*¹, Gonzalo Oton*¹, Matteo Piccardo² Mark Pickering², Agata Elia², Guido Ceccherini¹,
4 Mariano Garcia³ Mirco Migliavacca¹ Alessandro Cescatti¹

5 ¹European Commission, Joint Research Centre, Italy

6 ²Consultant of European Commission, Joint Research Centre, Italy

7 ³Universidad de Alcalá, Department of Geology, Geography and the Environment, Environmental Remote Sensing Research
8 Group, C/ Colegios, 2, Alcalá de Henares, 28801, Madrid, Spain

9

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

11 * These authors contributed equally

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

28

29 **1 Introduction**

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

35

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

44

45 Mapping forest structural diversity over large spatial scales proved challenging due to the lack of comprehensive datasets and
46 consistent data collection methodologies, hindering our ability to predict ecosystem function at large geographic scales. Whilst
47 forest structural parameters can be measured in various ways, traditional field-based measures of stand structure are generally
48 labour-intensive and have been limited to small areas (Goodbody et al., 2023). Laser scanning, or LiDAR, has been proved a
49 sound alternative for measuring tree height from 3D data measured through echoes (Coops et al., 2021). However, data from
50 airborne LiDAR have been limited in spatial and temporal coverage to specific regions (Hancock et al., 2021). Recent advances
51 in satellite remote sensing technology and computational capabilities have made it possible to measure a range of structural
52 variables at larger scales than ever before. Notably, the Global Ecosystem Dynamics Investigation (GEDI) (Dubayah et al.,
53 2020a) instrument, placed on board the International Space Station (ISS) in December 2018, collecting LiDAR samples until
54 March 2023, has revolutionized the assessment of forest structure. Recent studies have shown how structural data collected by
55 GEDI can be used in several applications ranging from biomass estimation to the monitoring of biodiversity and ecosystem
56 disturbances (Crockett et al., 2023; Hakkenberg et al., 2023; Holcomb et al., 2024). These early examples demonstrate the
57 future potential of the GEDI mission.

58

59 Whilst the availability of GEDI data and the computational capacity to handle large datasets have opened up new opportunities
60 to map structural diversity, GEDI only collects sparse measurements of vegetation structure. Although the GEDI mission has
61 recently been extended, it is expected to cover only a minimal fraction of the land surface. Depending on the application of

62 interest, continuous information on structural diversity over forests may be needed. Combining GEDI with other types of
63 satellite remote sensing data within a machine learning framework may thus be necessary for the creation of structural diversity
64 data products that have continuous coverage and that extend beyond the timeframe covered by the GEDI mission. In this
65 context, predictor variables derived from complementary satellite observations are used to bridge the gap between sparse GEDI
66 measurements and the need for wall-to-wall maps of forest structural diversity. These predictors can act as observable proxies
67 for canopy structural complexity, enabling the spatial extrapolation of the GEDI-derived structural diversity metrics across
68 Europe. Several recent studies have successfully combined GEDI data with other remote sensing data sources to predict canopy
69 structure in areas not covered by GEDI, paving the way for mapping specific structural features of vegetation regionally and
70 globally (Aragoneses et al., 2024; Lang et al., 2023; Potapov et al., 2021; Schwartz et al., 2024). Additionally, preliminary
71 efforts to assess the potential of GEDI data to capture canopy diversity over different regions have been carried out (Schneider
72 et al., 2020). However, despite these significant advances, no efforts have been made to map forest structural diversity at a
73 continental scale in Europe.

74

75 To address the lack of readily available structural diversity data, we combined a suite of structural diversity indicators
76 calculated using GEDI data with active radar and passive optical data from the Sentinel-1, Sentinel-2, and ALOS-PALSAR
77 missions. **This sensor combination was specifically selected to enable spatially and temporally consistent, wall-to-wall
78 estimates suitable for large-scale and monitoring, while capturing complementary structural information across different
79 canopy layers: Sentinel-2 multispectral data are sensitive to canopy biochemical and structural properties at the crown surface,
80 including vegetation density and phenological state. Sentinel-1 C-band SAR interacts primarily with the upper canopy and
81 smaller structural components, capturing variations in canopy roughness ALOS-PALSAR-2 L-band SAR, owing to its longer
82 wavelength, exhibits enhanced sensitivity to larger structural elements and sub-canopy features, providing information on
83 forest vertical complexity.** These different sources of data were then integrated using a predictive modelling framework, based
84 on a machine learning method. The resulting models were used to predict structural diversity across Europe. Although Sentinel-
85 1, Sentinel-2 and ALOS-PALSAR-2 data have been previously used for predicting canopy height and other structural
86 components of forests, their joint use for mapping forest structural diversity has not yet been attempted (Liu et al., 2025; Wang
87 et al., 2024). Our analysis includes a total of eight structural diversity metrics, including metrics that quantify the vertical and
88 horizontal heterogeneity of the canopy, as well as metrics that quantify the heterogeneity of forest structure among GEDI
89 observations within a given area. The dataset presented here is readily available for use as input in various environmental
90 models and analyses.

91

92 **2 Methods**

93 We calculated eight forest structural diversity metrics using NASA GEDI observations (Dubayah et al., 2020b). A list of the
94 metrics is reported in Table 1. A machine learning (ML) framework was used to model the relations between each metric and

95 a series of predictors derived from passive optical and active radar remote sensing data. The model was then used to create a
96 structural diversity dataset that covers the whole forested domain of Europe, extending up to $\sim 52^\circ$ North, which corresponds
97 to the northern latitudinal limit of the GEDI mission. The creation of the dataset involved five main steps (Fig. 1): (i) satellite
98 remote sensing data pre-processing, (ii) structural diversity metric calculation (iii) model training, (iv) model validation, and
99 (v) prediction (Fig. 1). Data from GEDI, Sentinel 1 and 2, ALOS-Palsar-2 were pre-processed and downloaded from Google
100 Earth Engine (GEE), a cloud-based infrastructure that combines a multi-petabyte catalogue of satellite imagery and geospatial
101 datasets with planetary-scale analysis capabilities (Gorelick et al., 2017).

102

103 We used data covering forests that had remained ecologically stable, meaning they experienced no canopy loss, from 2000 to
104 2021, as identified through the Global Forest Change product by Hansen et al. (2013). Furthermore, our analyses were limited
105 to areas where tree cover exceeded 30% and which bordered at least 6 out of 8 neighbouring pixels, also with tree cover
106 exceeding 30%. While our threshold is more stringent than Food and Agriculture Organization (FAO) definition of forest
107 (FAO, 2000), which specifies an area spanning more than 0.5 hectares with trees taller than 5 metres and a canopy cover of
108 more than 10%, it was chosen to capture areas with substantial arboreal density. Although our selection was guided by the
109 FAO's broader forest criteria, we customized these guidelines to suit our research focus. **This threshold was intentionally
110 adopted to focus on landscapes with well-developed and spatially continuous forest canopies. This more conservative threshold
111 reduces the influence of sparsely treed or transitional land-cover types and improves the robustness and interpretability of
112 GEDI-based structural diversity metrics**

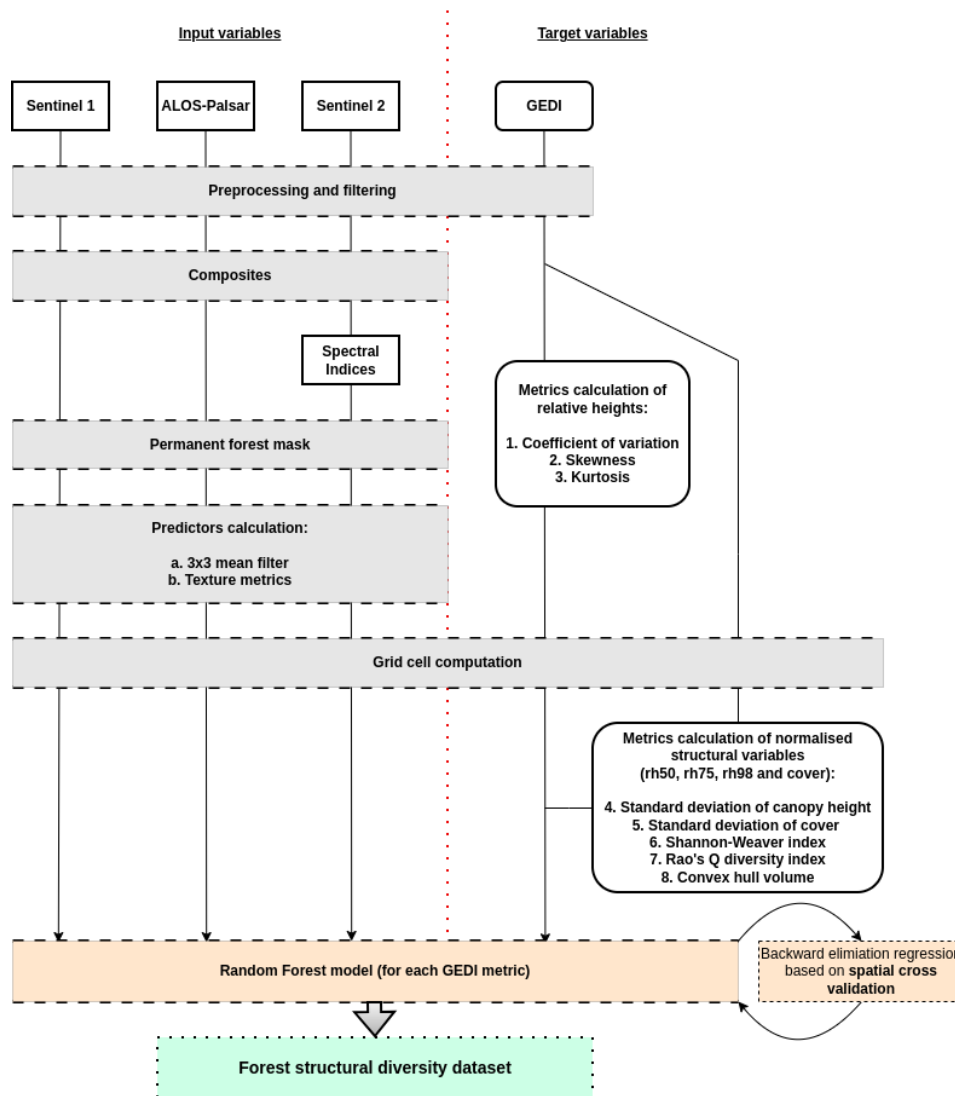
113

114 **2.1 Structural diversity metrics**

115 Structural diversity can be characterized in a variety of ways depending on the data from which it is calculated and the intended
116 application. In this work we adopted the common definition where diversity is defined in the vertical dimension as
117 heterogeneity of vegetation height and in the horizontal dimension as canopy heterogeneity (Hakkenberg and Goetz, 2021).
118 We chose a set of metrics that would characterize the heterogeneity within and among structural features for a given area,
119 reflecting both local (alpha) and regional (beta) measures of structural diversity. These complementary metrics have been
120 demonstrated to be particularly crucial for predicting tree diversity and ecosystem functioning (Coverdale and Davies, 2023;
121 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,
122 including Foliage Height Diversity (FHD from the GEDI L2B product) and the Waveform Structural Complexity Index
123 (WSCSI, from the GEDI L4C product), already capture important aspects of forest structural heterogeneity and have proven
124 highly valuable for large-scale analyses. However, recent work has shown that these indices exhibit strong scaling relationships
125 with top-of-canopy height (RH98) (de Conto et al., 2024).**

126

127



129

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

135

136 In the context of this study, our objective was therefore not to directly use existing GEDI products as target variables, but to
 137 develop structural diversity metrics that explicitly quantify heterogeneity while minimising direct dependence on canopy
 138 height. This motivated the selection of eight complementary metrics based on the distributional properties of GEDI relative

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

151

152 2.1.1 GEDI input data and general framework

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

157

158 Input data included the GEDI Level 2A Relative Heights (RH), and the Level 2B total Canopy Cover (CC) values (see Table
159 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
160 covered by the vertical projection of the tree crowns. The GEDI data were downloaded from Google Earth Engine after
161 applying a filtering procedure to remove low-quality and unreliable observations and to reduce noise in the input data, based
162 on standard GEDI quality flags and thresholds (Table S1 in the Supplement).

163 Structural diversity metrics were computed for each spatial analysis unit, defined as regular grid cells (hereafter referred to
164 pixels) at 1 km, 5 km, and 10 km spatial resolution. For each pixel, all number (M) of valid GEDI shots overlapping the pixel
165 between April 2019 and January 2023 were collected. The structural diversity metrics of a given pixel are calculated by
166 aggregating all M overlapping that pixel. Each GEDI shot i was characterized by its RH distribution $RH_i =$
167 $\{rh_i^k, rh_i^{k+1}, \dots, rh_i^{100}\}$ with $k: rh_i^k \geq 0$ (i.e. only the positive values were considered) and total canopy cover cc_i . To ensure
168 robust estimation of structural diversity, pixels with fewer GEDI observations than a minimum sampling threshold were
169 excluded. This threshold was defined as the median number of valid GEDI shots across all pixels at a given spatial resolution.
170 In addition, extreme metric values were filtered using a z-score criterion, with values exhibiting a deviation greater than 3

171 across pixels discarded as outliers. The remaining pixels were then used to compute the structural diversity metrics. A post-
172 processing step was applied to remove extreme values: pixels with structural diversity values exhibiting a z-score greater than
173 3 were discarded as outliers.

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

184

185 In the following sections, we detail the methodology employed for calculating the diversity metrics and predictor variables,
186 which makes use of the mean $\mu(X)$, standard deviation $\sigma(X)$, skewness $\gamma(X)$, excess kurtosis $\kappa(X)$, coefficient of variation
187 $cv(X)$ of a variable $X = \{x_1, \dots, x_M\}$, where X represents a vector of observations (see Appendix A for the explicit
188 formulations).

189

190 2.1.2 Vertical Diversity Metrics

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

195

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

- 197 1. the average coefficient of variation of the vertical profiles

198

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

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

202

203 2. the average skewness of the vertical profiles

204
$$\tau_{SK} = \mu(\Gamma)$$

205 where $\Gamma = \{\gamma(RH_1), \dots, \gamma(RH_M)\}$ represents the set of skewness values computed from the RH distribution of each of the
206 M GEDI observations within the pixel. Skewness $\gamma(RH)$, or third standardized moment, is a measure of the asymmetry of the
207 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
208 distribution (a distribution with a single peak), positive skewness generally indicates an asymmetric tail extending toward
209 larger height values (overstorey heterogeneity), while negative skewness suggests a tail extending toward smaller height values
210 (understorey heterogeneity). However, note that in the cases where one tail is long, but the other tail is fat, or the distribution
211 is multi-modal, skewness does not always obey this simple rule.

212

213 3. the average excess kurtosis of the vertical profiles

214
$$\tau_{KU} = \mu(K)$$

215 where $K = \{\kappa(RH_1), \dots, \kappa(RH_M)\}$ represents the set of excess kurtosis values computed from the RH distribution of each of
216 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
217 any univariate normal distribution, (Fig. S2 a and d panels in the Supplement). Distributions with negative/positive excess
218 kurtosis are said to be platykurtic/leptokurtic. Platykurtic distributions show fewer and/or less extreme outliers than the normal
219 distribution. In this case, the vegetation mass is more concentrated around the VP mean than near the vertical extremes (i.e.
220 the ground and to top canopy height). However, it is important to note that while kurtosis, the fourth moment, does play a role
221 in characterizing the shape of VP, its influence is comparatively smaller than that of the standard deviation, the second moment,
222 and skewness, the third moment. For instance, two distinct VPs may exhibit identical excess kurtosis while displaying
223 markedly disparate distributions in terms of standard deviations.

224

225 2.1.3 Horizontal Diversity Metrics

226 We calculated 5 vertical horizontal diversity indices.

227

228 1. the standard deviation of the canopy heights

229
$$\tau_{CH} = \sigma(CH)$$

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

232

233 2. the standard deviation of the total canopy cover

234
$$\tau_{CC} = \sigma(CC)$$

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

237

238 2.1.4 Combined Vertical and Horizontal and Diversity Metrics

239

240 3. the Shannon-Weaver index

241

$$242 \tau_{SW} = - \sum_{\log p_{\varepsilon\pi o\omega}} p_{\varepsilon\pi o\omega} \log p_{\varepsilon\pi o\omega}$$

243

244

245 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
 246 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
 247 amplified by 10. τ_{SW} measures the uncertainty or disorder inherent to the variable's possible outcomes. $\tau_{SW} = 0$ when all
 248 observations are confined within a single bin, otherwise τ_{SW} is larger than zero. Higher values indicate heterogeneity, while
 249 lower values suggest homogeneity.

250

251 4. Rao's quadratic diversity index

$$252 \tau_{RAO} = \sum_{\varepsilon\pi o\omega} \sum_{\varepsilon'\pi'o'\omega'} p_{\varepsilon\pi o\omega} D_{\varepsilon\pi o\omega}^{\varepsilon'\pi'o'\omega'} p_{\varepsilon'\pi'o'\omega'}$$

253 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
 254 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
 255 bin size on each axis, and the GEDI CC values were amplified by 10. τ_{RAO} ranges from zero, indicating no diversity, to positive
 256 numbers. Differently from τ_{SW} index, τ_{RAO} considers both abundance ($p_{\varepsilon\pi o\omega}$ terms) and dissimilarity in the sampled data
 257 ($D_{\varepsilon\pi o\omega}^{\varepsilon'\pi'o'\omega'}$ term).

258

259 5. convex hull volume

260

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

261 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
 262 CC values were amplified by 10. CVH is the function calculating the convex hull volume on the ensemble $SHT =$
 263 $\{sht_1, \dots, sht_M\}$, with $sht_i = (rh_i^{50}, rh_i^{75}, rh_i^{98}, cc_i)$. Larger volumes indicate increased heterogeneity.

264

265 Table 1. Structural diversity metrics computed in this study.

266

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

267

268 **2.2 Predictor variables**

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

- 272 1. appropriate bands/indices ϕ_α were calculated from the remote sensing raster images;
- 273 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
 274 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
 275 Angular Second Moment (ASM), entropy, and dissimilarity index, respectively (see Appendix A3).
- 276 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
 277 $\Phi_\alpha^\beta = \{\phi_{\alpha,1}^\beta, \dots, \phi_{\alpha,M}^\beta\}$ corresponding to the geographical positions that overlap the image pixels.

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

280

281 **2.2.1 Sentinel-1 radar data**

282 The European Space Agency's (ESA) Sentinel-1 (S1) comprises a constellation of two polar-orbiting satellites, sun-
 283 synchronous orbit with a 12-day repeat cycle, which operate day and night a C-band ($\lambda = 5.5$ cm) Synthetic Aperture Radar
 284 (SAR) to capture data at a spatial resolution of approximately 10 meters. The radar enables the acquirement of imagery

285 regardless of the weather, and the C-band frequency is particularly effective in interacting with fine vegetative elements such
 286 as leaves and branches (Naidoo et al., 2015). In our study, from Sentinel-1 we utilized both backscatter and coherence data.
 287 Backscatter is the portion of the outgoing radar signal that the target redirects directly back towards the radar antenna. The
 288 backscatter characteristics provide crucial insights into the physical properties of forest canopies. For the year 2020, we focused
 289 on the signal dual-polarization VV and VH Sentinel-1A (S1A) and Sentinel-1B (S1B) Ground Range Detected (GRD) data,
 290 acquired in the Interferometric Wide (IW) swath mode, as it predominantly covers land masses (Kellndorfer et al., 2022).
 291 VV(H) is a mode that transmits vertical waves and receives vertical (horizontal) waves to create the SAR image. We selected
 292 data from the descending orbit, which has been shown to exhibit fewer correlations with evapotranspiration (ET) (Mueller et
 293 al., 2022). Sentinel-1 data used in this study were obtained from Google Earth Engine, where they had already undergone some
 294 pre-processing. Preprocessing steps carried out by the Google Earth Engine team include applying the orbit file for geocoding,
 295 removing GRD border noise and thermal noise, and performing radiometric calibration. We performed a radiometric terrain
 296 correction following (Vollrath et al., 2020), as well as the removal of stripes and edges. **Following radiometric terrain**
 297 **correction and the removal of stripes and edge artefacts, we selected all valid Sentinel-1 observations captured over Europe**
 298 **within a six-month window centred on the date of maximum NDVI identified independently for each pixel from the Sentinel-**
 299 **2 dataset (see Section 2.2.3).** We then derived:

- 300 1. the S1 backscatter six-month mean $\phi_{S1VVgs\mu}$ and $\phi_{S1VHgs\mu}$, where the mean is intended to mitigate speckle noise
 301 while emphasizing the vegetation growing season;
- 302 2. the S1 backscatter standard deviation growing season $\phi_{S1VVgs\sigma}$ and $\phi_{S1VHgs\sigma}$;
- 303 3. the S1 backscatter bi-monthly mean $\phi_{S1VVpre\mu}$ and $\phi_{S1VHpre\mu}$ for a window extending two months before the month
 304 before the peak, $\phi_{S1VVact\mu}$ and $\phi_{S1VHact\mu}$ for the period spanning one month before to one month after the peak, and
 305 $\phi_{S1VVpost\mu}$ and $\phi_{S1VHpost\mu}$ for the two months after the month after the peak.

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

317 achieved, which correlate strongly with the radar signal and hence, reduce noise levels. Furthermore, we prioritized signal VV
318 polarization to enhance our understanding of the data, as it minimizes vegetation decorrelation effects (Pan et al., 2022)

319

320 **2.2.2 ALOS-PALSAR-2 radar data**

321

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

331

332 **2.2.3 Sentinel-2 optical data**

333 The ESA Sentinel-2 (S2) mission comprises a constellation of two polar-orbiting satellites placed in the same sun-synchronous
334 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
335 at best) allows monitoring of the Earth's surface changes. The Multi-Spectral Instrument (MSI) on board the 2 platforms
336 collects the sunlight reflected from the Earth and supplies high-resolution multispectral imagery with resolutions of 10 and 20
337 meters. Data are acquired at 10 m spatial resolution for Visible (Blue, Green, Red) and Near-Infra-Red (NIR) bands, and at 20
338 m spatial resolution for VNIR-Red Edge (RE1, RE2, RE3, RE4) and Short Wave Infra-Red (SWIR) bands (SWIR1, SWIR2).
339 The Level-2A product provides atmospherically corrected Surface Reflectance (SR) images. In this study we used all the
340 Level-2A images from 2000 to 2021 identified by a scene-level cloud and snow cover smaller than 70% and 5%, respectively,
341 as provided by Google Earth Engine. We then calculated:

- 342 1. the Normalized Difference Vegetation Index

343

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

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

- 346 2. the Normalized Difference Water Index

347

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

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

349 3. the Normalized Difference Red Edge Index

350
$$\phi_{NDRE} = \frac{\rho_{NIR} - \rho_{RE1}}{\rho_{NIR} + \rho_{RE1}}$$

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

353 4. the Modified Soil Adjusted Vegetation Index

354
$$\phi_{MSAVI} = \frac{2 \cdot \rho_{NIR} + 1 - \sqrt{(2 \cdot \rho_{NIR} + 1)^2 - 8 \cdot (\rho_{NIR} - \rho_{Red})}}{2}$$

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

357 5. the Green Normalized Difference Vegetation Index

358
$$\phi_{GNDVI} = \frac{\rho_{NIR} - \rho_{Green}}{\rho_{NIR} + \rho_{Green}}$$

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

361 6. the standard deviation of NDVI

362
$$\phi_{NDVI\sigma} = \sigma(\phi_{NDVI})$$

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

364

365 **2.3 Model training and validation**

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

376 To mitigate the potential for overfitting, we used a backward stepwise selection process that begins with a full model including
377 all available predictors. The algorithm then iteratively removes the least important feature, as determined by its contribution
378 to model performance. The relative importance of predictors was assessed using a permutation procedure (Altmann et al.,
379 2010). At each iteration, the model complexity is reduced by one predictor, and the resulting model is evaluated. We compared

380 each newly simplified model to the immediate predecessor to determine whether there was an improvement in performance or
381 a decrease that was less than 1% worse. The elimination process is halted if the removal of additional predictors causes the
382 model's performance to decrease by more than 1% compared to the previous iteration. At each step, a spatial cross-validation
383 procedure is used to assess the performance of the model. The metric we utilized to assess model performance throughout this
384 process was the coefficient of determination (R^2).

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

- 386 • Random train-validation split: in this approach, the dataset was randomly split, allocating 33% for model validation.
387 Random validation is a common method that provides a quick and often effective means of evaluating model
388 performance on unseen data. However, it has a notable drawback when dealing with spatial data: it disregards the
389 spatial structure inherent in the dataset (i.e. points close to each other are, generally, more similar than points further
390 away). By ignoring this spatial autocorrelation, random validation may inadvertently conceal overfitting issues,
391 leading to an overly optimistic perception of the model's predictive capabilities.
- 392 • 10-Fold Spatial Cross-Validation (Roberts et al., 2016): we implemented a 10-fold spatial cross-validation procedure
393 to address the shortcomings of random validation, thus reducing the overfitting. This more sophisticated method
394 partitions the data into ten spatially distinct subsets, or folds, ensuring that each fold comprises disjointed sets that are
395 geographically separated. The partitioning is achieved by clustering data points according to their spatial coordinates,
396 which preserves the spatial structure and autocorrelation present in the dataset. During the validation process, each
397 fold is used once as a validation set while the remaining folds serve as the training set. This technique provides a more
398 realistic evaluation of the model's performance and its ability to generalize across different spatial regions, thereby
399 offering a safeguard against overfitting and ensuring a more reliable assessment of the model's true predictive power.

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

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

410
411
412

413 3 Results

414

415 3.1 Spatial patterns of forest structural diversity

416

417 The dataset includes spatial grids for eight structural diversity metrics at three different resolutions (10 km, 5km and 1 km).

418 These metrics show a significant variation in structural diversity across the European forests as shown in Fig. 2 (see also Fig.

419 S3 and Fig. S4 in the Supplement for 5km and 1km resolution datasets).

420

421 An examination of the variability in the 10 km resolution metrics in climate space revealed distinct patterns along temperature

422 and precipitation gradients (Fig. 3). Patterns of variability in metrics describing vertical heterogeneity showed significant

423 differences when comparing the coefficient of variation (τ_{CV}) and skewness (τ_{SK}) against kurtosis (τ_{KU}). The coefficient of

424 variation and skewness primarily exhibited high values at the extremes of the climatic gradient. This is observed in warm and

425 arid climates where total annual precipitation is below ~ 500 mm and Annual Mean Temperature is above ~ 10 °C, as well as

426 in colder climates where Annual Mean Temperature is below ~ 5 °C. Patterns of variability in the kurtosis were more nuanced,

427 consistently showing negative values across the European domain, which suggests a tendency for a platykurtic distribution in

428 the vertical profile of canopies under diverse environmental conditions. The most pronounced negative kurtosis values were

429 observed for the northern part of the temperate climate zone (Fig. 3). By contrast, more heterogeneous patterns occurred in

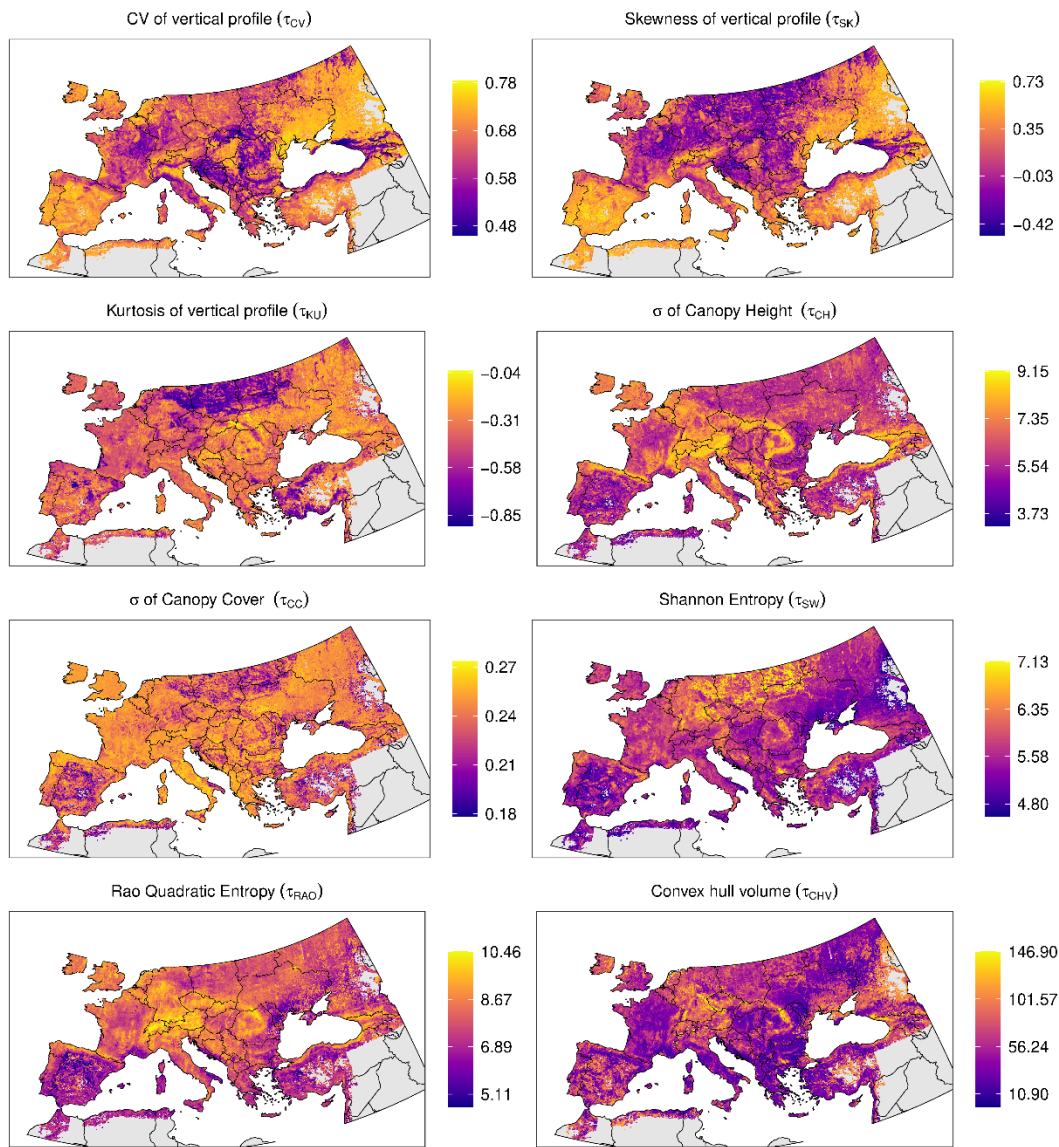
430 other areas such as those with a Mediterranean climate, showing high variability (Fig. 3). Diversity metrics describing

431 structural heterogeneity in horizontal space, as well as combined metrics, (τ_{CH} , τ_{CC} , τ_{SW} , τ_{RAO} , τ_{CVH}) also showed

432 considerable variability along precipitation and temperature gradients. With the exception of the convex hull (τ_{CVH}), all metrics

433 displayed low diversity values in hot and dry climates.

434



435

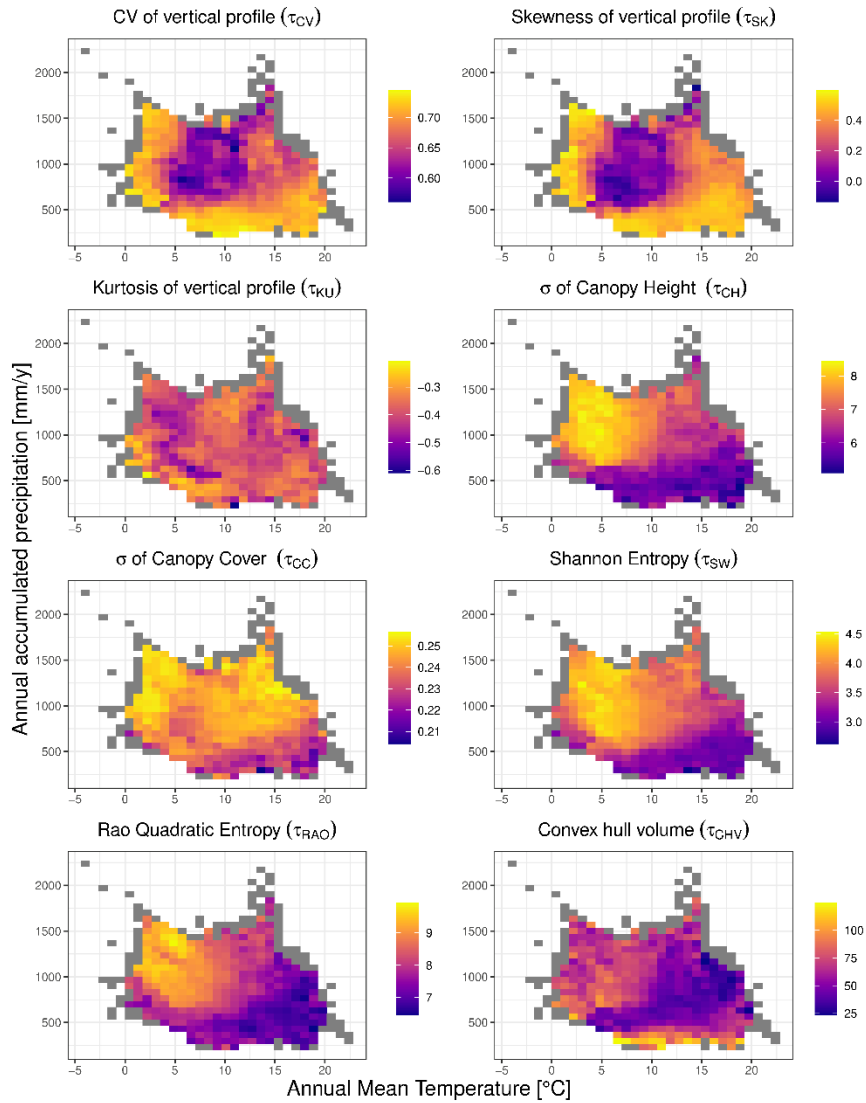
436

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

440

441

442



443

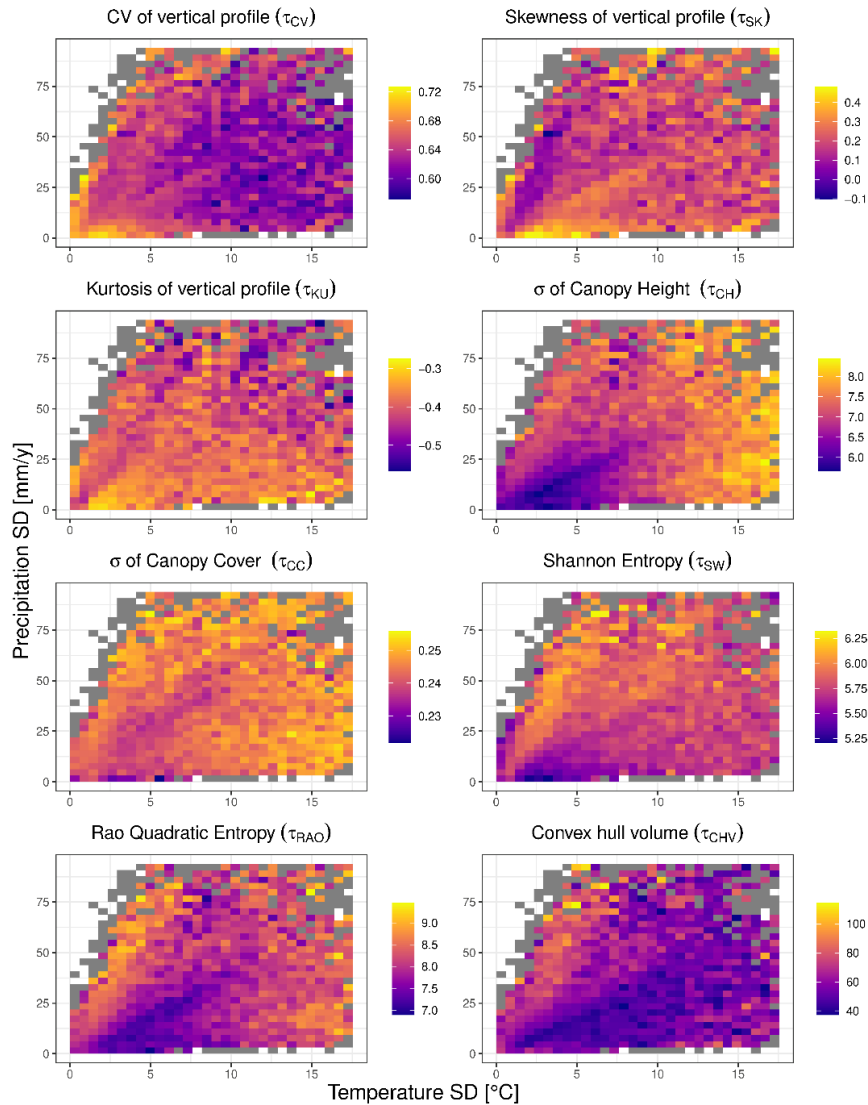
444 Figure 3. **Structural diversity variables mapped against mean climate (temperature and precipitation)**. The results refer to the dataset at 10 km resolution.

445 Coloured bins depict variation in structural diversity, calculated as the average of the structural diversity values falling within each bin. Grey bins indicate

446 those containing fewer than 5 observations, for which the average was not calculated.

447

448



449

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

453

454

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

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

462

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

477

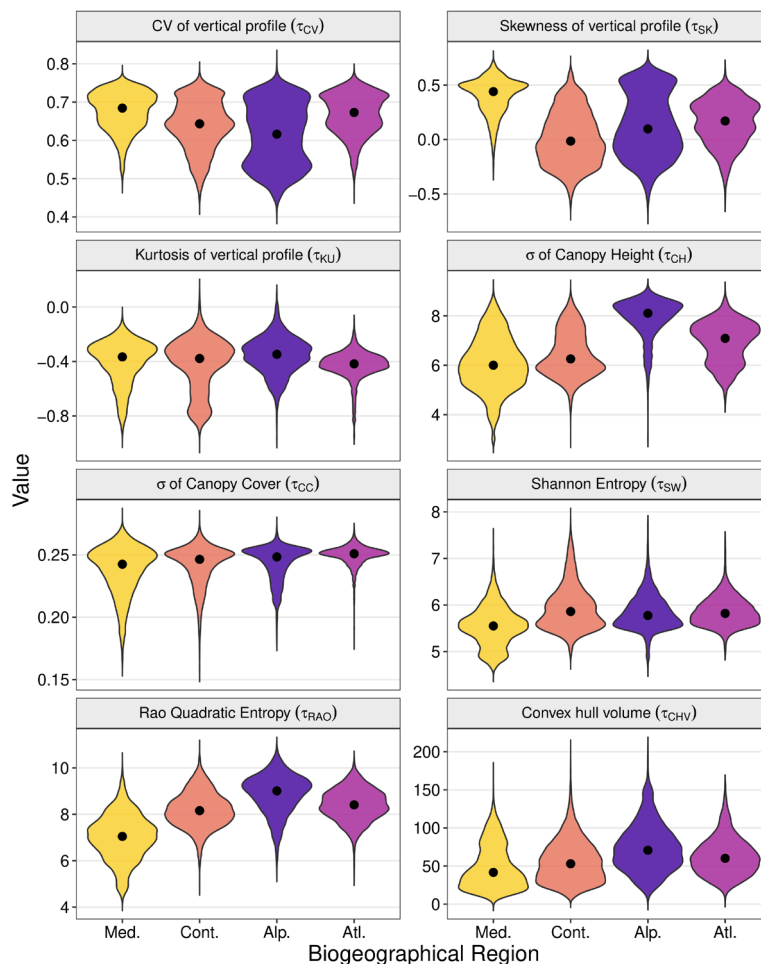
478 An examination of structural diversity patterns by biogeographic region largely reinforced the patterns revealed in the climate
479 space plots (Fig. 5, see Fig. S7 and Fig. S8 in the Supplement for the 5 km and 1 km datasets), while revealing additional
480 topographic effects. Combined horizontal-vertical diversity metrics (Shannon, Rao) were lowest in the Mediterranean region,
481 consistent with the low diversity observed in hot, arid climates, and highest in Continental, Atlantic, and Alpine regions.
482 Conversely, vertical canopy heterogeneity (Skewness τ_{SK}) was highest in the Mediterranean region, reinforcing the pattern of
483 elevated skewness at climatic extremes. Canopy height variability (τ_{CH}) was markedly elevated in the Alpine region ($M_{ed} =$
484 8.11) compared to Mediterranean ($M_{ed} = 6.00$), Continental ($M_{ed} = 6.26$), and Atlantic ($M_{ed} = 7.09$), regions likely
485 reflecting greater topographic heterogeneity. CV of vertical profile remained consistent across all regions (~0.65-0.70), while
486 (τ_{KU}) and σ of canopy cover τ_{CC} , showed minimal biogeographic variation.

487

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

492

493



496

497

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

501

502 3.2 Variable importance and model performance

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

508

509 An examination of the type of predictors selected in the final models highlighted the importance of radar-related predictors,
510 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
511 average proportion of radar-related variables selected across all diversity metrics and resolutions was 0.64, although there was
512 considerable variability. In general, as the resolution of the input dataset increased, the proportion of radar-related variables
513 selected through the feature elimination procedure also increased (Fig. 6A; for the 5 km and 1km datasets see Fig. S11A and
514 Fig. S12A in the Supplement). The diversity variables for which the highest number of radar-related predictors were selected
515 was the convex hull (τ_{CVH}). On the other hand, the one for which the highest number of optical-related predictors were selected,
516 was canopy cover (τ_{CC}).

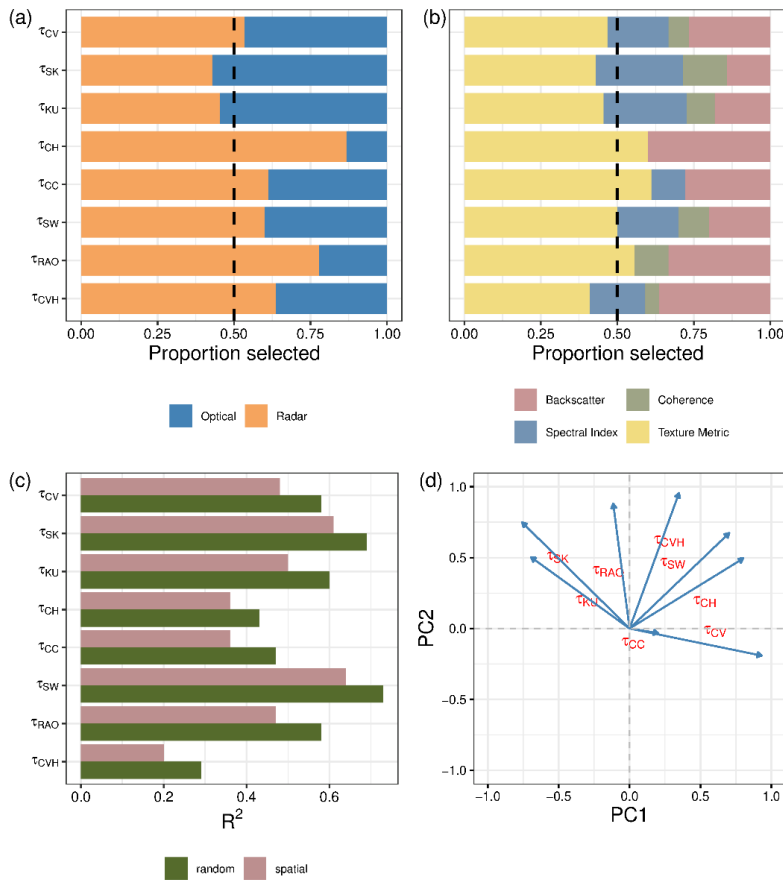
517

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

523

524 Model validation revealed that random cross-validation consistently outperformed spatial cross-validation across all
525 resolutions. **At 10 km resolution, the model for the Shannon** the model for Shannon index τ_{SW} achieved the highest scores,
526 with 0.73 in random validation and 0.64 in spatial validation (Fig 6C and Tables B1 and B2). Conversely, the model with
527 convex hull (τ_{CVH}) as a variable showed the lowest performance, scoring 0.29 in random cross-validation and 0.20 in spatial
528 cross-validation. **This difference likely reflects the contrasting statistical properties of the metrics: Shannon entropy integrates**
529 **information across the full distribution of GEDI observations within a spatial unit and is therefore more robust to sampling**
530 **variability, whereas convex hull-based metrics are more sensitive to outliers and local sampling density.** The best-performing
531 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
532 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
533 and B2; Fig. S11C and Fig. S12C in the Supplement).

534



535

536 Figure 6. Results of the random forest modelling exercise at 10 km resolution. Panels display the variable selection frequencies (A and B)
 537 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
 538 Principal Component Analysis (PCA) conducted on the **estimated structural diversity** metrics at this resolution.

539

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

546

547 An examination of the standard deviation of **model outputs** revealed generally increasing trend of prediction uncertainty across
 548 resolutions (Fig. S13, Fig. S14 and Fig. S15 in the Supplement), except in Rao (τ_{RAO}) and convex hull (τ_{CVH}). **Overall,**
 549 **uncertainty was generally low** across the spatial domain of interest, reflecting limited variability within the ensemble. Notable

550 exceptions occur in the Mediterranean region for the convex hull, kurtosis, and standard deviation of canopy height metrics.
551 Further variability is observed in Eastern Europe, particularly for the convex hull, skewness, kurtosis, Shannon index, and
552 standard deviation of canopy height.

553

554 **4. Discussion**

555

556 **4.1. Model-based predictions of structural diversity**

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

563

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

576

577 Our findings revealed that model performance differed according to the spatial resolutions and diversity metrics, with several
578 models achieving R^2 values indicative of moderate to strong predictive accuracy, particularly at coarser spatial resolutions
579 (Appendix B, Tables B1 and B2). This variation highlights the critical role of resolution in model performance, indicating that,
580 depending on the application of interest, coarser resolutions may optimize the utility of the models. As expected, spatial cross-
581 validation consistently yielded lower R^2 values than random train-validation random validation across most metrics and
582 resolutions. This outcome reflects the challenges inherent on machine learning methods (Meyer and Pebesma, 2021) of

583 predicting outcomes in areas geographically distinct from the training data. Nevertheless, the decrease was generally modest,
584 affirming the broad applicability of our models beyond the training domain.

585

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

593

594 **4.2. Potential applications**

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

597

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

605

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

612

613 Thirdly, the dataset could be used for improving Earth system models. Historically, plant canopy structure has not been
614 adequately represented in these models (Atkins et al., 2018; Schneider et al., 2020). This lack of detailed representation can
615 lead to significant errors in predicting energy balance, carbon cycling, and ecosystem responses to environmental changes
616 (Duveiller et al., 2023). Integrating structural diversity into these models has the potential to enhance the accuracy of

617 simulations by incorporating more realistic representations of light interception, photosynthetic rates, and energy fluxes. In
618 particular, these applications are directly relevant to contemporary Earth system modelling frameworks such as CMIP6 and
619 forthcoming CMIP7 simulations, which underpin IPCC climate assessments and projections.

620

621 **5. Data availability**

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

627 **6. Code availability**

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

630

631 **7. Conclusions**

632

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

639

640

641

642

643

644

645

646

647

648

649

650

651 **Appendices**

652

653 **Appendix A: Supplementary Methods**

654

655 **A1 Statistical indicators**

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

658

659

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

660

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

661

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

662

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

662

664

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

665

666 **A2 Binning in cartesian 4d space**

667

668 $p_{\varepsilon\pi o\omega}$ indicates the fraction of the GEDI shots falling in the bin identified by the indices $(\varepsilon, \pi, o, \omega)$ in the 4D cartesian space
 669 defined on the basis $(e^\varepsilon, e^\pi, e^o, e^\omega)$, see Figure S2, with

670

$$\sum_{\varepsilon\pi o\omega} p_{\varepsilon\pi o\omega} = 1 \quad (A2)$$

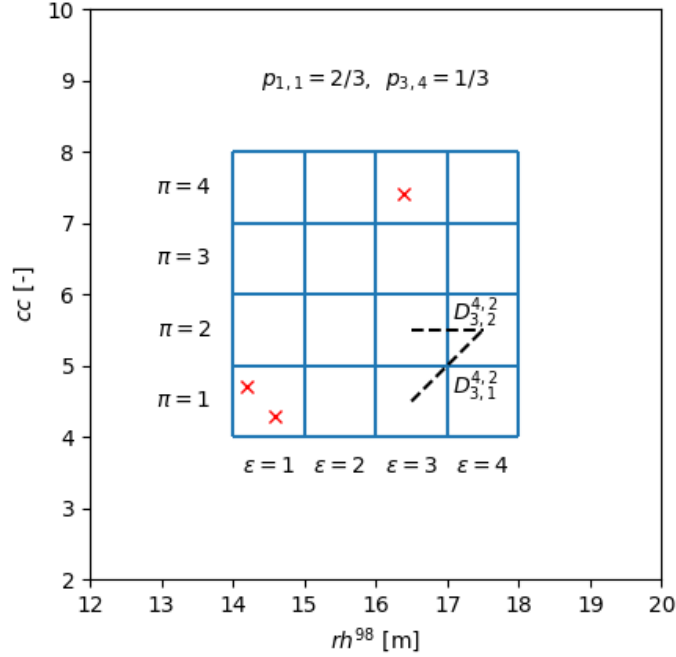
671

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

672

indicates the cartesian distance between $(\varepsilon, \pi, o, \omega)$ and $(\varepsilon', \pi', o', \omega')$ bin.

673



674

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

677

678 A3 Predictor calculation

679 Starting with appropriate bands/indices (step 2 of the workflow in the main text), the four scalars $\phi_{\alpha,i}^{\beta}$, where $\beta \in$
 680 $\{SM, ASM, ENT, DISS\}$, are calculated from the cluster of 7×7 pixels $\phi_{\alpha,i}(p, q)$ overlapping the footprint of the GEDI shot
 681 i , where p and q represent the pixel indices within the window. In details, we calculated:

682 1. the spatial mean (SM)

683

$$\phi_{\alpha,i}^{SM} = \mu(\phi_{\alpha,i}(p, q)) \quad (A3)$$

684 which is performed to compensate for potential footprint geolocation inaccuracies, and reduce the presence of noise, and three
 685 texture metrics. Texture metrics provide spatial content information (Nichol and Sarker 2011), and are highly effective in
 686 capturing the pixels heterogeneity. Defining $\bar{\phi}_{\alpha,i}(p, q)$ as the grey-levels matrix, which is calculated from $\phi_{\alpha,i}(p, q)$ by
 687 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-
 688 levels co-occurrence matrix (GLCM), with dimension 256×256 (Haralick et al. 1973):

$$689 \quad C_{\alpha,i}(m, n) = \sum_{p=1}^7 \sum_{q=1}^6 1, \text{ if } \bar{\phi}_{\alpha,i}(p, q) = m \text{ and } \bar{\phi}_{\alpha,i}(p, q + 1) = n; 0, \text{ otherwise}$$

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

$$691 \quad p_{\alpha,i}(m, n) = \frac{C_{\alpha,i}(m,n)}{\sum_{p=0}^{255} \sum_{q=0}^{255} C_{\alpha,i}(p,q)} \quad (A4)$$

692 we calculated:

693

694 2. the angular second moment (ASM)

$$695 \quad \phi_{\alpha,i}^{ASM} = - \sum_{m=0}^{255} \sum_{n=0}^{255} [p_{\alpha,i}(m, n)]^2 \quad (A5)$$

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

698 3. the entropy

$$699 \quad \phi_{\alpha,i}^{ENT} = - \sum_{m=0}^{255} \sum_{n=0}^{255} p_{\alpha,i}(m, n) \log p_{\alpha,i}(m, n) \quad (A6)$$

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

702 4. the dissimilarity index

$$703 \quad \phi_{\alpha,i}^{DISS} = \sum_{m=0}^{255} \sum_{n=0}^{255} p_{\alpha,i}(m, n) |m - n| \log p_{\alpha,i}(m, n) \quad (A7)$$

704 Dissimilarity measures the complexity and the nature of grey-level transitions among neighbouring pixels (Connors et al.
705 1984). It quantifies image contrast, with higher dissimilarity values reflecting pronounced differences among neighbouring
706 pixel values.

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

709

710

711

712

713

714

715

716

717

718

719

720

721

722 **Appendix B: Model validation results**

723

724 Table B1. Results of the random validation procedure conducted for the forest structural metrics at three spatial resolutions:
 725 1x1km, 5x5km, and 10x10km. The validation outcomes are presented in terms of the coefficient of determination (R^2), which
 726 quantifies the proportion of the variance in the dependent variable that is predictable from the independent variables.

727

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

728

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

732

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

733

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

738

739 **Competing interests** The authors declare no competing financial interests

740

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

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768 **References**

- 769 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
770 than species diversity enhances aboveground carbon storage in secondary subtropical forests in Eastern China, *Biogeosciences*,
771 13, 4627–4635, <https://doi.org/10.5194/bg-13-4627-2016>, 2016.
- 772 Altmann, A., Toloşi, L., Sander, O., and Lengauer, T.: Permutation importance: a corrected feature importance measure,
773 *Bioinformatics*, 26, 1340–1347, <https://doi.org/10.1093/bioinformatics/btq134>, 2010.
- 774 Aponte, C., Kasel, S., Nitschke, C. R., Tanase, M. A., Vickers, H., Parker, L., Fedrigo, M., Kohout, M., Ruiz-Benito, P.,
775 Zavala, M. A., and Bennett, L. T.: Structural diversity underpins carbon storage in Australian temperate forests, *Global Ecology*
776 *and Biogeography*, 29, 789–802, <https://doi.org/10.1111/geb.13038>, 2020.
- 777 Aragoneses, E., García, M., Ruiz-Benito, P., and Chuvieco, E.: Mapping forest canopy fuel parameters at European scale using
778 spaceborne LiDAR and satellite data, *Remote Sens Environ*, 303, 114005, <https://doi.org/10.1016/j.rse.2024.114005>, 2024.
- 779 Atkins, J. W., Fahey, R. T., Hardiman, B. S., and Gough, C. M.: Forest Canopy Structural Complexity and Light Absorption
780 Relationships at the Subcontinental Scale, *J Geophys Res Biogeosci*, 123, 1387–1405, <https://doi.org/10.1002/2017JG004256>,
781 2018.
- 782 Bae, S., Levick, S. R., Heidrich, L., Magdon, P., Leutner, B. F., Wöllauer, S., Serebryanyk, A., Nauss, T., Krzystek, P.,
783 Gossner, M. M., Schall, P., Heibl, C., Bäessler, C., Doerfler, I., Schulze, E.-D., Krah, F.-S., Culmsee, H., Jung, K., Heurich,
784 M., Fischer, M., Seibold, S., Thorn, S., Gerlach, T., Hothorn, T., Weisser, W. W., and Müller, J.: Radar vision in the mapping
785 of forest biodiversity from space, *Nat Commun*, 10, 4757, <https://doi.org/10.1038/s41467-019-12737-x>, 2019.
- 786 Breiman, L.: Random forests, *Mach Learn*, 45, 5–32, 2001.
- 787 Bruggisser, M., Dorigo, W., Dostálová, A., Hollaus, M., Navacchi, C., Schläffer, S., and Pfeifer, N.: Potential of Sentinel-1 C-
788 Band Time Series to Derive Structural Parameters of Temperate Deciduous Forests, *Remote Sens (Basel)*, 13, 798,
789 <https://doi.org/10.3390/rs13040798>, 2021.
- 790 Cartus, O., Santoro, M., Wegmuller, U., Labriere, N., and Chave, J.: Sentinel-1 Coherence for Mapping Above-Ground
791 Biomass in Semiarid Forest Areas, *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5,
792 <https://doi.org/10.1109/LGRS.2021.3071949>, 2022.
- 793 Coops, N. C., Tompalski, P., Goodbody, T. R. H., Queinnec, M., Luther, J. E., Bolton, D. K., White, J. C., Wulder, M. A., van
794 Lier, O. R., and Hermosilla, T.: Modelling lidar-derived estimates of forest attributes over space and time: A review of
795 approaches and future trends, *Remote Sens Environ*, 260, 112477, <https://doi.org/10.1016/j.rse.2021.112477>, 2021.
- 796 Coverdale, T. C. and Davies, A. B.: Unravelling the relationship between plant diversity and vegetation structural complexity:
797 A review and theoretical framework, *Journal of Ecology*, 111, 1378–1395, <https://doi.org/10.1111/1365-2745.14068>, 2023.
- 798 Crockett, E. T. H., Atkins, J. W., Guo, Q., Sun, G., Potter, K. M., Ollinger, S., Silva, C. A., Tang, H., Woodall, C. W.,
799 Holgerson, J., and Xiao, J.: Structural and species diversity explain aboveground carbon storage in forests across the United
800 States: Evidence from GEDI and forest inventory data, *Remote Sens Environ*, 295, 113703,
801 <https://doi.org/10.1016/j.rse.2023.113703>, 2023.

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

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

809 Duveiller, G., Pickering, M., Muñoz-Sabater, J., Caporaso, L., Boussetta, S., Balsamo, G., and Cescatti, A.: Getting the leaves
810 right matters for estimating temperature extremes, *Geosci Model Dev*, 16, 7357–7373, [https://doi.org/10.5194/gmd-16-7357-](https://doi.org/10.5194/gmd-16-7357-2023)
811 2023, 2023.

812 Ehbrecht, M., Seidel, D., Annighöfer, P., Kreft, H., Köhler, M., Zemp, D. C., Puettmann, K., Nilus, R., Babweteera, F., Willim,
813 K., Stiers, M., Soto, D., Boehmer, H. J., Fisichelli, N., Burnett, M., Juday, G., Stephens, S. L., and Ammer, C.: Global patterns
814 and climatic controls of forest structural complexity, *Nat Commun*, 12, 519, <https://doi.org/10.1038/s41467-020-20767-z>,
815 2021.

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

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

819 Forzieri, G., Girardello, M., Ceccherini, G., Spinoni, J., Feyen, L., Hartmann, H., Beck, P. S. A., Camps-Valls, G., Chirici, G.,
820 Mauri, A., and Cescatti, A.: Emergent vulnerability to climate-driven disturbances in European forests, *Nat Commun*, 12,
821 1081, <https://doi.org/10.1038/s41467-021-21399-7>, 2021.

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

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

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

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

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

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

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

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

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

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

844 Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz,
845 S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., and Townshend, J. R. G.: High-Resolution
846 Global Maps of 21st-Century Forest Cover Change, *Science (1979)*, 342, 850–853, <https://doi.org/10.1126/science.1244693>,
847 2013.

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

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

853 Lang, N., Jetz, W., Schindler, K., and Wegner, J. D.: A high-resolution canopy height model of the Earth, *Nat Ecol Evol*, 7,
854 1778–1789, <https://doi.org/10.1038/s41559-023-02206-6>, 2023.

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

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

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

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

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

869 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.:
870 The coordinated impact of forest internal structural complexity and tree species diversity on forest productivity across forest
871 biomes, *Fundamental Research*, <https://doi.org/10.1016/j.fmre.2022.10.005>, 2022.

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

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

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

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

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

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

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

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

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

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

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

910 Rocchini, D., Thouverai, E., Marcantonio, M., Iannacito, M., Da Re, D., Torresani, M., Bacaro, G., Bazzichetto, M., Bernardi,
911 A., Foody, G. M., Furrer, R., Kleijn, D., Larsen, S., Lenoir, J., Malavasi, M., Marchetto, E., Messori, F., Montagni, A., Moudrý,
912 V., Naimi, B., Ricotta, C., Rossini, M., Santi, F., Santos, M. J., Schaeppman, M. E., Schneider, F. D., Schuh, L., Silvestri, S.,
913 Šímová, P., Skidmore, A. K., Tattoni, C., Tordoni, E., Vicario, S., Zannini, P., and Wegmann, M.: rasterdiv—An Information
914 Theory tailored R package for measuring ecosystem heterogeneity from space: To the origin and back, *Methods Ecol Evol*,
915 12, 1093–1102, <https://doi.org/10.1111/2041-210X.13583>, 2021.

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

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

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

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

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

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

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

934 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
935 processes in China, *J Hydrol (Amst)*, 561, 187–199, <https://doi.org/10.1016/j.jhydrol.2018.04.003>, 2018.

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

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

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

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

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

948 Wang, C., Zhang, W., Ji, Y., Marino, A., Li, C., Wang, L., Zhao, H., and Wang, M.: Estimation of Aboveground Biomass for
949 Different Forest Types Using Data from Sentinel-1, Sentinel-2, ALOS PALSAR-2, and GEDI, *Forests*, 15, 215,
950 <https://doi.org/10.3390/f15010215>, 2024.

951 Zhai, L., Will, R. E., and Zhang, B.: Structural diversity is better associated with
952 forest productivity than species or functional diversity, *Ecology*, 105, <https://doi.org/10.1002/ecy.4269>, 2024.