Dear Dr. Heim,

Please find below responses to the reviewers' comments. We are grateful for the constructive feedback from you and the reviewers, which has substantially improved the clarity and focus of the paper.

In response to your recommendation to emphasize data presentation over methodological details, we have added two new figures to the main text: one summarizing structural diversity metrics across biogeographic regions, and another showing the spatial distribution of model prediction errors.

We have also addressed the reviewers' comments by: (1) clarifying the rationale for our choice of the eight structural diversity metrics and how they differ from existing GEDI products, (2) explaining the selection of spatial resolutions and their intended applications, (3) providing clearer justification for our choice of predictor datasets, and (4) expanding the discussion of model validation performance, particularly for metrics showing lower predictive accuracy.

We believe these revisions have strengthened the manuscript's focus on the dataset itself while maintaining the methodological transparency expected by the ESSD community.
We look forward to the discussion phase.

Best regards,

Marco Girardello and Gonzalo Oton (on behalf of co-authors)

Editor's comments

Dear Authors

many thanks for your submission of your manuscript to ESSD. As you may know, papers accepted for discussion in ESSD appear directly online for comments and review. Therefore, all papers undergo an access review undertaken by an editor before. Your paper describes an interesting and useful and validated data set. The presentation quality in the paper is high, the methods and the data quality are well described. In the current version of your manuscript the focus is too much on the method for an ESSD data paper. For example, later in the review process, you could add more data visualisations and some more data statistics in the main text. This can be all later adressed in the review process but does at this stage not prevent the manuscript to proceed to review. I look forward to your ESSD discussion,
Best wishes,

Birgit Heim

We would like to thank the editor for the constructive feedback and positive assessment of our manuscript. We are pleased that the dataset, methods description, and data quality validation were found to be of high standard.

We have carefully considered the editor's suggestion to shift the focus more toward the data presentation and away from methodological details. In response, we have revised the manuscript by including two additional figures in the main text:

1. **Structural diversity variables mapped against climate variability (temperature and precipitation SD in space). Lines 449-452 and 460-476.**
2. **A summary of the structural diversity metrics organized by biogeographic regions, providing readers with a clearer overview of spatial patterns and regional variation in the dataset. Lines 478-486 and 496-500.**

These additions enhance the visualization and characterization of the dataset while maintaining the methodological rigor that underpins our work. We believe these revisions better align the manuscript with ESSD's focus on data description and presentation

Response to Reviewer 1's comments

This manuscript presents the results of a case study to produce a continental dataset on vegetation (forest) heterogeneity via associating Sentinel-1 and -2 driven variables with sparsely distributed GEDI-derived structural metrics. The case study per se is not new, as all the underlying data, methods (RF modeling, cross-validation) have been extensively used in a plethora of previous studies at different spatial levels. In this regard, the manuscript can only be considered as a pure data description paper with no technical innovative aspects associated with the underlying case study. There are currently many other modeling approaches via both statistical and deep learning techniques that can be used to increase the performance of the results and their applicability for large-scale analysis. In addition, the fact that the turnover of a number of spectral variables extracted from active and passive remote sensing data are directly associated with 3D structural heterogeneity has been confirmed in the literature for a while. Examples are DOIs. 10.1088/1748-9326/ac5f6d, 10.1016/j.foreco.2023.120987, 10.1186/s13021-023-00228-9, 10.3390/rs14143345, 10.5194/bg-18-1234-2021, 10.1002/eap.2567, 10.1109/TGRS.2022.3156789 and many more.

As already noted to the editor in relation to the open discussion phase, this comment raises general observations about the scope and positioning of the manuscript rather than specific methodological or data-related issues. We therefore respond briefly by clarifying the intended contribution of the study within the context of a data description paper.

We thank the reviewer for their comments and the opportunity to clarify the scope of our work. The reviewer notes that the manuscript does not employ deep learning

methods and combines data from passive optical and SAR satellites in a way they consider not novel. We would like to clarify that our study does not claim novelty in the combined use of optical and SAR data, nor does it focus on the application of deep learning techniques. Rather, the main contribution – as explicitly outlined in the manuscript – lies in the development and adaptation of eight distinct metrics to generate, for the whole of Europe, a comprehensive dataset describing structural diversity at multiple spatial resolutions. To our knowledge, this represents the first attempt to systematically map structural complexity at a quasi-continental scale.
The review also includes DOIs referencing studies that are presented as closely related to our work. After careful examination, we found that some of the provided DOIs appear to be incorrect or point to studies that differ substantially in scope and methodology, particularly in relation to the construction of a dataset describing the structural diversity of European forests. We therefore believe that these references do not directly contradict the scope or contribution of the present study.
Below, we report the DOI verification performed using the DOI Foundation's resolver (https://www.doi.org/), together with the corresponding bibliographic information where available:

- **10.1088/1748-9326/ac5f6d**
  DOI NOT FOUND
- **10.1016/j.foreco.2023.120987**
  McKinney, Caleb M., Ronald E. Masters, Arjun Adhikari, Bijesh Mishra, Omkar Joshi, Chris B. Zou, and Rodney E. Will. Forage quantity and protein concentration changes across a forest–savanna gradient with management implications for white-tailed deer. Forest Ecology and Management 538 (2023): 120987.
- **10.3390/rs14143345**
  Shao, Z., Zhang, X., Zhang, T., Xu, X., & Zeng, T. (2022). RBFA-Net: a rotated balanced feature-aligned network for rotated SAR ship detection and classification. Remote Sensing, 14(14), 3345.
- **10.5194/bg-18-1234-2021**
  DOI NOT FOUND
- **10.1002/eap.2567**
  Grinde, Alexis R., Melissa B. Youngquist, Robert A. Slesak, Stephen R. Kolbe, Josh D. Bednar, Brian J. Palik, and Anthony W. D'Amato. Potential impacts of emerald ash borer and adaptation strategies on wildlife communities in black ash wetlands. Ecological Applications 32(4) (2022): e2567.
- **10.1109/TGRS.2022.3156789**
  DOI NOT FOUND

We hope this clarifies the scope and contribution of our work.

Response to Reviewer 2's comments

The preprint introduces wall-to-wall structural complexity/diversity metrics for Europe which were derived from Sentinel-1, Sentinel-2, ALOS-Palsar and are based on GEDI structural metrics. Both vertical and horizontal, as well as combined structural diversity is addressed. For this, a standard random forest machine learning approach was performed.

There is definitely a gap and the need for spatially-explicit structural complexity assessments for European forests to advance our understanding of the impact of structural complexity or changes of complexity on the carbon cycle and the ecological functions of a forest. Hence, this study is highly appreciated.

We thank Reviewer #2 for the constructive comments. We are pleased that the Reviewer recognises the relevance of a wall-to-wall, spatially explicit dataset on structural diversity for European forests. Below we respond point-by-point, focusing on clarification of design choices, scope, and interpretation, as appropriate for the ESSD public discussion stage

General comments

I feel this study lacks clarity in the choices that were made: e.g. why these particular eight structure metrics? Why Sentinel-1/2 and ALOS-Palsar? I think clarification on a couple of methodological steps and decisions will help to assess whether the metrics and datasets presented here, fill the existing gap of structural diversity metrics for European forests and for which purposes and models the datasets can be used.

We agree with the reviewer that clearer justification of several key methodological choices would improve the transparency and interpretability of the study. While the rationale for the selection of the structural diversity metrics and the choice of Sentinel-1, Sentinel-2, and ALOS-PALSAR as predictor datasets is introduced in the manuscript, we acknowledge that these aspects would benefit from being stated more explicitly. In particular, the sensor combination was chosen to enable the production of spatially and temporally consistent, wall-to-wall estimates of forest structural diversity suitable for large-scale and long-term monitoring. We address these points in detail in the revised version of the manuscript (L77–83). We added the following text:

This sensor combination was specifically selected to enable spatially and temporally consistent, wall-to-wall estimates suitable for large-scale and monitoring, while capturing complementary structural information across different canopy layers: Sentinel-2 multispectral data are sensitive to canopy biochemical and structural properties at the crown surface, including vegetation density and phenological state; Sentinel-1 C-band SAR interacts primarily with the upper canopy and smaller structural components, capturing variations in canopy roughness; and ALOS-PALSAR-2 L-band SAR, with its longer wavelength, exhibits enhanced sensitivity to larger structural elements and sub-canopy structure, providing information on forest vertical complexity.

Specific general comments:

I appreciate this study and the focus on structural diversity of European forests, as well as the integrated use of several openly-available EO-datasets. I would welcome a couple of clarifications on general decisions taken and processing steps. I think the preprint and the usability of the dataset would be enhanced by this. These points do not necessarily correspond to one line or particular paragraph in the preprint, hence this 'specific general comment'-section

1. Existing structural diversity indices or datasets

While this study presents quite a complex assessment of structural diversity, I am missing a paragraph on existing structural diversity indices and datasets. Not all of the existing datasets are wall-to-wall or some only address vertical or horizontal structural diversity but I still think these have to be mentioned and their advantages and disadvantages should be addressed. This would help to identify the current knowledge gap and underline the need for other structural diversity metrics, such as those presented here.

I would expect that at least (but not limited to these): the GEDI L2B products (PAI, FHD, PAVD, …), GEDI L4C WSCI (also z, xy), Forest canopy structural complexity (CSC), are addressed.

We agree with the reviewer that existing structural diversity indices and datasets should be more clearly acknowledged and positioned. Aspects of vertical profile heterogeneity are already addressed in several GEDI-derived products, including GEDI L2B metrics such as Foliage Height Diversity (FHD) and related vertical profile descriptors, as well as the GEDI L4C Waveform Structural Complexity Index (WSCI) and canopy structural complexity (CSC) measures. These datasets have proven highly valuable for characterising forest structure at large scales. We thus expanded the manuscript to explicitly discuss existing GEDI-based structural diversity and complexity products, their scope, and their limitations in relation to the objectives of this study, thereby clarifying the knowledge gap that motivates the development of the proposed metrics. Specifically we added the following text in the Methods section (L121–140):

Several GEDI-derived indices, including Foliage Height Diversity (FHD, L2B) and the Waveform Structural Complexity Index (WSCI, L4C), already capture important aspects of forest structural heterogeneity and have proven highly valuable for large-scale analyses. However, recent work has shown that these indices exhibit strong scaling relationships with top-of-canopy height (RH98) (de Conto et al., 2024). In the context of this study, our objective was therefore not to directly use existing GEDI products as target variables, but to develop structural diversity metrics that explicitly quantify heterogeneity while minimising direct dependence on canopy height. This motivated the selection of eight complementary metrics based on the distributional properties of GEDI relative height profiles and canopy cover, designed to be ecologically interpretable, minimally redundant with each other and with height, and suitable for spatial aggregation and wall-to-wall mapping using multi-sensor satellite data.

1.5. Especially concerning the provided GEDI products (L2B and L4C): Did you assess these products as potential target variables? Why did you choose to calculate 8 new metrics?

We initially explored the use of existing GEDI-derived indices, including products from GEDI L2B (e.g. Foliage Height Diversity), as potential target variables. While these indices capture important aspects of vertical profile heterogeneity and have proven highly valuable for large-scale analyses, they are strongly correlated with top-of-canopy height (RH98).

In the context of this study, our objective was therefore not to directly use existing GEDI-derived variables that are redundant with RH98, but to develop structural diversity metrics that explicitly quantify heterogeneity while reducing direct dependence on canopy height. This motivated the selection of eight complementary metrics based on the distributional properties of GEDI relative height profiles and canopy cover, allowing us to characterise heterogeneity within and among GEDI observations in a way that is interpretable and suitable for spatially continuous, wall-to-wall prediction. The selected metrics explicitly span three complementary dimensions of structural diversity: (i) vertical heterogeneity within individual canopy profiles, (ii) horizontal heterogeneity among GEDI observations within a spatial unit, and (iii) combined multivariate structural diversity. To clarify these design choices, we expanded the Methods section (L139–143) to better explain why these eight metrics were selected, how they relate to existing GEDI products, and which specific knowledge gap they are intended to address. The rationale for this design choice and the description of the selected metrics is provided in the expanded Methods section (L136–140), as detailed above.

References:

de Conto, T., Armston, J., & Dubayah, R. (2024). Characterizing the structural complexity of the Earth's forests with spaceborne lidar. Nature Communications, 15(1), 8116.

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

2. Choice of 8 metrics

I understand that these 8 metrics address vertical, horizontal and combined structural diversity and therefore depict the multidimensional structural diversity. However, I feel the manuscript misses a paragraph that explains why these metrics where chosen, and a little more detail on what the individual metric depicts. I am aware that this is briefly mentioned in 2.1.2, 2.1.3 and 2.1.4 (error in preprint, line 186 the subsection should be 2.1.4) but I think it would be good to elaborate on this to ensure that every reader can follow understand the choice made.

We agree with the reviewer that the manuscript benefits from a clearer and more consolidated explanation of why these eight metrics were selected and what each of

them represents. While the rationale for the metrics was previously introduced across Sections 2.1.2–2.1.4, we acknowledge that this information was fragmented and could be made more accessible to readers.

The eight metrics were deliberately chosen to span three complementary dimensions of structural diversity. Metrics derived from the distributional properties of GEDI relative height profiles characterise vertical heterogeneity within individual canopy profiles, capturing differences in vertical layering and profile shape. Metrics based on variability among GEDI observations within a spatial unit describe horizontal heterogeneity, reflecting spatial variation in canopy height and cover. Finally, multivariate metrics integrate both vertical and horizontal information to represent combined structural diversity within a single framework.

This design provides an interpretable yet comprehensive representation of forest structural diversity while minimising redundancy among metrics and with top-of-canopy height. To address this comment, we added a dedicated paragraph synthesising this rationale and briefly describing the ecological meaning of each metric to Section 2.1 of the Methods (L 141–150):

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

3. 10km, 5km, and 1km spatial resolution

Can you please specify how you chose these 3 spatial resolutions? Can you also address the potential application/uses of structural diversity metrics at these spatial resolutions? Do modelling studies operate at these resolutions? I am just wondering, as the all the inputs have a much higher spatial resolution and also certain potential applications that I can think of (besides modelling) would benefit from a higher spatial resolution, e.g. disturbances (small-scale disturbances are very common), edge effects, forests become more fragmented, …

We thank the Reviewer for this important comment. We agree that the rationale for selecting the three spatial resolutions and their intended applications should be more explicitly articulated in the manuscript.

The choice of 1 km, 5 km, and 10 km resolutions reflects a trade-off between spatial detail, the sampling density of GEDI observations, and the robustness of the derived

structural diversity metrics. Because GEDI provides sparse footprint-level measurements, reliable estimation of structural diversity within a spatial unit requires a sufficient number of GEDI shots. Coarser spatial resolutions therefore improve metric stability, whereas finer resolutions provide greater spatial detail at the cost of higher uncertainty.

The 10 km resolution was identified as the most robust scale for continental-scale analyses and large-area monitoring, and is broadly compatible with the spatial aggregation typically used in regional ecosystem modelling and model–data integration studies. The 5 km and 1 km products were additionally provided to support applications requiring finer spatial detail, such as regional ecological analyses and biodiversity assessments.

We acknowledge that certain applications, including the analysis of small-scale disturbances, edge effects, and forest fragmentation, would benefit from even finer spatial resolution. However, at such scales the limited density of GEDI observations substantially constrains the reliable estimation of structural diversity metrics over large areas. For this reason, we opted to provide multi-resolution products that allow users to select the spatial grain most appropriate for their specific application, while explicitly recognising the associated trade-offs in uncertainty and metric robustness.

To address this comment, we added text to the Methods section clarifying the rationale for the selected spatial resolutions and explicitly linking each resolution to its intended applications and limitations (L 174–183):

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

Specific comments

4. line 74/75: give sources/citations from other studies.

We thank the Reviewer for this suggestion. We agree that additional references should be provided to support this statement. Appropriate citations to previous studies have been added in the revised manuscript.

5. Figure 1: difference between solid edge boxes and dashed boxes is not clear. I am not sure what the difference between 'data that were directly utilised' and 'raw, original data' means pratically

We revised Figure 1 caption and legend to improve interpretability.

6. line 128: What is M?

We thank the Reviewer for pointing this out. M refers to the number of valid GEDI shots within each spatial unit. This has now been explicitly defined in the revised manuscript (L. 164).

7. line 132: What are 'our calculations' referring to? The calculation of the eight individual structural metrics?

Yes, "our calculations" refers to the computation of the eight structural diversity metrics derived from GEDI observations. We have rephrased the sentence to make this explicit. The revised text now reads:

We computed the eight structural diversity metrics at three spatial resolutions (1 km, 5 km, and 10 km) using grids in the Lambert Azimuthal Equal Area (LAEA) projection. (L 174 – 175 in the revised version)

8. line 128 to 136: I do not understand this paragraph. Is this the GEDI shot selection? Is the reference to pixels or areas? What is the median value based on? Same question for the Z-score?

We thank the Reviewer for highlighting that this paragraph was unclear. The purpose of this section is to describe the selection, filtering, and aggregation of GEDI shots within spatial analysis units prior to the computation of the structural diversity metrics.

The reference is to spatial grid cells (analysis areas defined at 1 km, 5 km, or 10 km resolution), rather than individual GEDI footprints. For each grid cell, all valid GEDI shots overlapping that area during the study period were collected and used to compute the structural diversity metrics.

To ensure robust metric estimation, grid cells with insufficient GEDI sampling were excluded. The minimum sampling threshold was defined based on the median number of valid GEDI shots across all grid cells at a given resolution. In addition, extreme metric values were filtered using a z-score criterion ($|z| > 3$), applied to the distribution of metric values across grid cells, to remove outliers.

We agree that the original wording did not sufficiently clarify these steps. We have therefore rewritten this paragraph in the revised manuscript to explicitly describe the spatial units, GEDI shot selection, sampling thresholds, and outlier filtering criteria. Specifically we added the following text (L 163-173):

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

9. Line 152-153: definition not clear. What is the expected CV of VP? I think 'latter' is not correct in this sentence anything. What is central tendency?

We agree with the Reviewer that the definition in the original text was unclear and that the wording was imprecise. In this context, the coefficient of variation (CV) of a vertical profile is defined as the ratio between the standard deviation and the mean of the relative height distribution for a given GEDI shot. The reference to "expected value" was therefore ambiguous.

We have revised the text to explicitly define the coefficient of variation and to clarify that higher values of τ_{CV} indicate greater vertical dispersion relative to the mean profile height, and hence greater vertical heterogeneity. Specifically, the following text was added to the manuscript (Lines 199–201):

The coefficient of variation $cv(RH)$ *quantifies the the extent of vertical variability in relation within a vertical profile as the ratio between the standard deviation and the mean of the relative height distribution. A higher* τ_{CV} *therefore indicates greater vertical dispersion relative to mean profile height and hence greater vertical heterogeneity*

10. Line 158, reference to Figure S2: be more specific in the main text and also in the Figure S2 caption. There are so many panels in this figure with different foci; please clarify which panel refers to what

We agree that the references to Figure S2 are currently too general. We note that while all panels in Figure S2 display the four moments (μ, σ, γ, κ) for each vertical profile shown, the panels differ in which parameter varies. We have now revised both the main text (line 207) and the Figure S2 caption to explicitly direct readers to the panels that best illustrate each concept: panel a shows the reference normal distribution (γ = 0, κ = 0), while panels d demonstrate variation in skewness (γ). Specifically, the following text was added to Fig S2 caption:

Examples of different unimodal vertical profiles and their relative moments μ (mean), σ (standard deviation), γ (skewness), and κ (excess kurtosis).. All panels display the four moments for each profile shown; panels differ in which parameter is varied to demonstrate its effect on vertical profile shape. Panel a shows a normal distribution (γ = 0, κ = 0) as reference. Panels b-c demonstrate variation in μ and σ, which determine the coefficient of variation. Panels d demonstrate variation in γ (skewness). Panels e demonstrate variation in κ (kurtosis).

11. Line 168: reference to Figure S2: same comment as above. Be more specific in Figure S2, which panel/s illustrate what (skewness, kurtosis?); this will increase understandability

We agree with the Reviewer's suggestion. We have revised the main text (line 207) and Figure S2 caption to specify which panels best illustrate each distributional property. Panel a shows the normal distribution reference (κ = 0), while panels e demonstrate variation in kurtosis (κ). As detailed in our response to comment 10, the Figure S2 caption has been enhanced to clarify which panels demonstrate which distributional properties. This clarification improves readability by directing readers to the panels where each parameter varies.This clarification improves readability by directing readers to the panels where each parameter varies.

12. Figure 4: PCA (panel D) displayed but not mentioned in text. Why was a PCA performed? What does the result show us?

We thank the Reviewer for this comment. While the PCA (Fig. 4D) is mentioned in the manuscript in the context of intercorrelation among metrics, we agree that its purpose and interpretation were not explained explicitly.

In the revised manuscript, we have expanded the Results section (Lines 488–491) to clarify that the PCA was used as an exploratory tool to assess the degree of intercorrelation among the eight predicted structural diversity metrics. We now explicitly describe how the distribution of metrics in the PCA biplot, characterised by wide angular separation among vectors, indicates low to modest correlations and supports the complementary nature of the vertical, horizontal, and combined metrics. We further note that similar patterns are observed across spatial resolutions (Fig. S7, Fig. S9D, and Fig. S10D in the Supplement). Specifically, we added the following text:

A Principal Component Analysis (PCA) biplot (Fig. 4D) was used to explore the degree of intercorrelation among the eight predicted structural diversity metrics. The PCA shows that the metrics are distributed across the principal component space with generally wide angular separation among vectors, indicating low to modest correlations. These patterns are consistent across spatial resolutions (Fig. S7, Fig. S9D and Fig. S10D in the Supplement).

13. Table B1, B2: Model validation paragraph in main text (starting line 415). Some of the validation results are quite poor (both for the random and cross-validation method); e.g. σ of Canopy Cover (τCC) 0.16 (random validation). I think this should be discussed more. Are the provided datasets on figshare those that showed these poor validation results? This is not

clear. I am not sure if maps/datasets with such a low validation score should be included/used, e.g. from modellers as input to their models. Or what do you think? Maybe I am misunderstanding this. I think this is not highlighted/marked sufficiently in the text.

We thank the Reviewer for this important comment. We agree that the range of model performance across metrics, including those with lower validation scores, should be more clearly discussed and highlighted in the manuscript.

In the revised manuscript, we have expanded the Results section (Lines 540–545) to explicitly discuss the observed variability in model performance across metrics and spatial resolutions. In particular, we now highlight that metrics describing horizontal variability, such as the standard deviation of canopy cover (τ_CC) and convex hull volume (τ_CVH), exhibit lower predictive performance, especially at finer spatial resolutions. We further explain that these metrics rely more directly on variability among individual GEDI observations within spatial units, which is less consistently preserved when locally derived optical and SAR predictors are aggregated to the grid-cell level. Specifically, we added the following text to the Results section (LINES 540-545):

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

All datasets provided on Figshare correspond exactly to the reported validation results. To further improve transparency for downstream users, we have also updated the Data Availability section (Lines 624-626) to explicitly state that all eight structural diversity metrics at all three spatial resolutions are provided, and that users are encouraged to consult the validation results (Tables B1 and B2) when assessing the suitability of individual metrics for specific applications. Specifically, we added the following text to the Data Availability section:

All eight structural diversity metrics at all three spatial resolutions are provided; users are encouraged to consult the validation results (Tables B1 and B2) to assess the suitability of individual metrics for specific applications.

14. Table S1: GEDI points; I think conventionally it is GEDI shots or GEDI footprints but not points.

We agree with the Reviewer that "GEDI shots" or "GEDI footprints" is more appropriate terminology than "points". We have now corrected this throughout Table S1.

Technical corrections

15. line 186: the subsection should be 2.1.4

We thank the Reviewer for noting this error. The subsection numbering has been corrected to 2.1.4 in the revised manuscript.

Response to Reviewer 3's comments

This work aims to present a new dataset on the structural diversity of forest canopies for the European region, built by combining available datasets using a machine learning approach. The paper is generally well written, with a good introduction and a clear description of the methods used to create the dataset. The proposed dataset is state of the art: it makes the best use of available information in a single product that will be of great utility for the scientific community in ecology, climatology, and hydrology. In particular, it provides rich, fine-scale information on structural diversity, which is strongly linked to a series of processes relevant to ecosystem functioning and to the climatological and hydrological feedbacks ecosystems can produce. In the context of land surface models and large-scale hydrological models, such a dataset would contribute to more realistic simulations of vegetation dynamics components.

The fact that radar-based (as opposed to optical-based) predictors emerge as important predictors highlights the value of multi-source data, which this work leverages effectively. The finding that coarser resolution yields good, or even better, predictive accuracy than fine resolution is also interesting.

We thank the Reviewer for the positive and encouraging assessment of the dataset and the manuscript. We appreciate the Reviewer's recognition of the potential value of the proposed structural diversity metrics for a broad range of applications, which is the intended purpose of our dataset. Below we provide a point-by-point response, clarifying the rationale behind our design choices, the study's scope, and the interpretation of the results, in line with the ESSD public discussion stage.

Overall, I find this work worthy of publication after minor revisions—see Specific Comments and the following note.

Specific comments
73-74 – I am not comfortable with the use of the term "predict", which I associate to the weather forecast or climate projections realms. Perhaps "estimate"? Although I realise such a setup may be employed for future data too.

We thank the Reviewer for this helpful suggestion. We agree that the term "predict" can be ambiguous and may be interpreted as implying temporal forecasting. In this study, the modelling framework is used to generate spatially explicit estimates of forest structural diversity from Earth observation data rather than forecasts in time.

We have therefore replaced "predict" with "estimate" (or "derive", where appropriate) throughout the manuscript when referring to the mapped outputs, while retaining "predictive modelling" only when describing the methodological framework. This

change clarifies that the models are used for spatial estimation based on contemporaneous EO data, while remaining applicable to future datasets if desired.

82 – “A list of the metrics is reported in Table 1” this is a repetition with later L. 115. At this stage you could simply refer to Section 2.1 as opposed to the Table 1 directly.

We thank you the reviewer for this comment. We have now deleted the sentence “A list of the metrics is reported in Table 1”

87 – repetition. Remove “(Fig 1)”.

94-95 – repetition. Remove “also with tree cover exceeding 30%.”

We thank the Reviewer for noting these repetitions. We will remove the redundant references and change the text as suggested.

92-98 – I understand that you employed more stringent criteria than those of FAO, but could you provide insight on your choices? E.g. 30% vs. 10% tree cover.

We thank the Reviewer for this comment. We adopted a more stringent tree cover threshold (30%) than the FAO definition to focus the analysis on areas with clearly developed forest canopies and to reduce potential noise from sparsely treed or transitional land-cover types.

To clarify this choice, we have added the following explanation in the revised manuscript (Lines 109–112): “This threshold was intentionally adopted to focus on landscapes with well-developed and spatially continuous forest canopies. This more conservative threshold reduces the influence of sparsely treed or transitional land-cover types and improves the robustness and interpretability of GEDI-based structural diversity metrics.”

104 /Figure 1. “The process culminates with yellow boxes” – Make sure you refer to the right color (I don’t see yellow color in the figure).

We thank the Reviewer for pointing this out. We will correct the color reference in the figure caption to ensure consistency with the actual figure.

126 – Why filter? To attenuate noise or other?

We thank the Reviewer for this comment. Filtering was applied to remove low-quality and unreliable GEDI observations and to reduce noise in the input data, thereby ensuring robust estimation of the structural diversity metrics.
We have clarified the purpose of the filtering step in the revised manuscript (Lines 160–162) and added the following text: ***“****The GEDI data were downloaded from Google Earth Engine after applying a filtering procedure to remove low-quality and unreliable observations and to reduce noise in the input data, based on standard GEDI quality flags and thresholds (Table S1 in the Supplement)**.”*** **Readers are referred to Table S1 in the Supplement for a detailed description of the applied quality criteria**

158 – Correct typo: “an unimodal” to “a unimodal”.

Lines 176 / 186 – The sections here have the same number – 2.1.3! Also, if you make a distinction between vertical, horizontal, and v. and h. combined, then L. 177 should go to 2 diversity indices, and the following section with the remaining 3 combined indices, in accordance with Table 1.

189 – Watch for the top index in the capital sigma, it is missing.

We thank the Reviewer for identifying these issues. The typos, subsection numbering, and missing index notation will be corrected in the revised manuscript.

213 – “The variables used as ML predictors were calculated from Sentinel-1, Sentinel-2, and ALOS-Palsar-2 observed data”. The predictors sources are presented very swiftly at this point, as if these data sets are the obvious choice. Can the authors provide a context as to why these observed data are fit for this purpose?

We have now provided more context as to why these observed data are fit for purpose. Specifically we added the following text (Lines 269-271):

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

222 – Same as the comment above. At this line you report that 47 predictors were derived. If you could provide some information on how you got to this set of predictors, the trade-offs you had to deal with.

We thank the Reviewer for this comment. We agree that the rationale behind the selection of the predictor datasets and the derivation of the final set of 47 predictors required clearer explanation. To address this, we have expanded the description in the Methods section to explicitly motivate the choice of Sentinel-1, Sentinel-2, and ALOS-PALSAR-2 as complementary data sources for characterising forest structure. In particular, we now clarify that these sensors were selected to capture complementary structural information through optical reflectance, C- and L-band SAR backscatter, and associated textural properties. We also clarify that the predictor set was designed to balance information content and redundancy while enabling spatially consistent, wall-to-wall estimation. Specifically, we added the following text in Section 2.2 (Lines 269–271):

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

In addition, broader motivation for the choice of sensors and their suitability for large-scale structural diversity mapping is now explicitly introduced in the Introduction (Lines 77–83).

This sensor combination was specifically selected to enable spatially and temporally consistent, wall-to-wall 77 estimates suitable for large-scale and monitoring, while capturing complementary structural information across different 78 canopy layers: Sentinel-2 multispectral data are sensitive to canopy biochemical and structural properties at the crown surface, 79 including vegetation density and phenological state. Sentinel-1 C-band SAR interacts primarily with the upper canopy and 80 smaller structural components, capturing variations in canopy roughness ALOS-PALSAR-2 L-band SAR, owing to its longer 81 wavelength, exhibits enhanced sensitivity to larger structural elements and sub-canopy features, providing information on 82 forest vertical complexity.

239 – not sure I understand: "We selected all the valid images captured over Europe within a six-month window, centred around the day of maximum NDVI from the Sentinel-2 dataset" – so, of the entire dataset, you picked the day of maximum NDVI and took 3 months before and after it?

We thank the Reviewer for highlighting this ambiguity. The six-month window is defined independently for each pixel, based on the pixel-specific date of maximum NDVI derived from the Sentinel-2 dataset, rather than a single peak date across the entire study area. For each pixel, observations were selected within a window spanning three months before and three months after its local NDVI maximum.

We have rephrased the corresponding sentence in the Methods section to explicitly clarify this pixel-wise temporal selection procedure (Lines 296-299).

Following radiometric terrain 296 correction and the removal of stripes and edge artefacts, we selected all valid Sentinel-1 observations captured over Europe 297 within a six-month window centred on the date of maximum NDVI identified independently for each pixel from the Sentinel-298 2 dataset (see Section 2.2.3).

327-341 – provided that, as the authors write, the 2nd approach is less prone to overfitting and more adequate for gridded data where spatial autocorrelation exists, why didn't the authors just go with the 2nd approach?

We thank the Reviewer for this insightful question. We agree that spatial cross-validation is better suited for gridded spatial data because it accounts for spatial autocorrelation and provides a more conservative assessment of model generalisation.

In this study, we therefore prioritise spatial cross-validation when evaluating model robustness and transferability. However, we retained random validation as a complementary approach to facilitate comparison with previous studies that rely on standard random splits and to characterise overall model behaviour under commonly used machine-learning evaluation settings.

We have clarified this rationale in the Methods section (Lines 400–405), explicitly stating the distinct purposes of the two validation strategies and emphasising the role of spatial cross-validation for assessing model applicability. Specifically we added the following text:

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

References

Wadoux, A. M. C., Heuvelink, G. B., De Bruin, S., & Brus, D. J. (2021). Spatial cross-validation is not the right way to evaluate map accuracy. Ecological Modelling, 457, 109692.

355 – Figure 2 (but valid also for Figure 3 and annex figures). While the layout is clear and well organised, I suggest the colorbar to have: larger tick label font size, and a color palette with discrete bins, not continuous gradient. Also, possibly ticks with even values.

We thank the Reviewer for this helpful suggestion. In the revised figures, we have increased the tick label font size of the colour bars to improve readability.

Regarding the use of a discrete colour palette, we opted to retain a continuous gradient. This choice reflects the continuous nature of the underlying variables and avoids introducing artificial thresholds that could imply categorical interpretation. We therefore felt that a continuous colour scale better supports the intended visual interpretation of the results.

Where applicable, we ensured that colour bar ticks are clearly legible and consistently scaled across figures.

360 – Figure 3 is a very interesting figure unveiling insightful information on the relationship between precipitation and temperature in the different metrics. I wonder if there is a chance to point a region or two in the figure's climate space. For instance, at L. 375, you refer to the characteristics for the Mediterranean region, it would help interpretation to find the region in Figure 3 if that is possible.

We thank the reviewer for this helpful suggestion. To improve the interpretability of regional patterns discussed in the text, we have added a new figure (Figure 4) in which results are separated by biogeographic region. This allows regions such as the Mediterranean to be directly identified in climate space, complementing the general overview provided by Figure 3. The manuscript has been updated accordingly.

379-380 –It may seem obvious but I find it interesting that low Precip. and relatively high mean temperature is associated with the lowest levels of diversity.

384-385 – This is also interesting, that results are insensitive to the grain size.

We thank the reviewer for these observations. We agree that these patterns are interesting and highlight robust relationships between climate conditions, structural diversity, and spatial scale, as already discussed in the manuscript.

413 – the Figure 4d graph introduces PCA. I would introduce it in the manuscript text providing insight on why it is there and what it tells us.

We agree that the PCA should be more clearly introduced in the manuscript. Although Figure 4D is currently referenced in the context of intercorrelation among the structural diversity metrics, the purpose of the PCA and the interpretation of its results are not described explicitly. The PCA was used to assess correlations among the predicted metrics and to illustrate their complementarity across vertical, horizontal, and combined dimensions. In the revised manuscript, we will explicitly introduce the PCA in the main text and provide guidance on how to interpret Figure 4D.

In line with comment made by reviewer 2 we responded to

In the revised manuscript, we have expanded the Results section (Lines 488–491) to clarify that the PCA was used as an exploratory tool to assess the degree of intercorrelation among the eight predicted structural diversity metrics. We now explicitly describe how the distribution of metrics in the PCA biplot, characterised by wide angular separation among vectors, indicates low to modest correlations and supports the complementary nature of the vertical, horizontal, and combined metrics. We further note that similar patterns are observed across spatial resolutions (Fig. S7, Fig. S9D, and Fig. S10D in the Supplement). Specifically, we added the following text:

A Principal Component Analysis (PCA) biplot (Fig. 4D) was used to explore the degree of intercorrelation among the eight predicted structural diversity metrics. The PCA shows that the metrics are distributed across the principal component space with generally wide angular separation among vectors, indicating low to modest correlations. These patterns are consistent across spatial resolutions (Fig. S7, Fig. S9D and Fig. S10D in the Supplement).

416-417 – So the model with Shannon index achieved highest scores, while the one with convex hull the lowest. Is there an explanation?

We thank the Reviewer for this question. We have now clarified this point in the Results section. The difference in performance likely reflects the contrasting statistical properties of the metrics: Shannon entropy integrates information across the full distribution of GEDI observations within a spatial unit and is therefore more robust to sampling variability, whereas convex hull–based metrics are more sensitive to outliers and local sampling density. This explanation has been added to the Results section (Lines 524–533):

Model validation revealed that random cross-validation consistently outperformed spatial cross-validation across all resolutions. At 10 km resolution, the model for the Shannon The 10 km resolution analysis, the model for Shannon index τ_SW achieved the highest scores, with 0.73 in random validation and 0.64 in spatial validation (Fig 64C and Tables B1 and B2). Conversely, the model with convex hull (τ_CVH) as a variable showed the lowest performance, scoring 0.29 in random cross-validation and 0.20 in spatial cross-validation. This difference likely reflects the contrasting statistical properties of the metrics: Shannon entropy integrates information across the full distribution of GEDI observations within a spatial unit and is therefore more robust to sampling variability, whereas convex hull–based metrics are more sensitive to outliers and local sampling density.The best-performing models at 5 km and 1 km differed from those at 10 km,. These trends varied across resolutions, with skewness models 〚(τ〛_SK) yielding the best results at both 5 km and 1 km, while canopy cover variability (τ_CC) was the worstlowest-performing at 1 km and convex hull (τ_CVH)) at 5 km (Tables B1 and B2; Fig. S119C and Fig. S120C in the Supplement).

445-446 – This indicates well the novelty of this work! "indices based solely on optical data fail to capture crucial aspects of structural heterogeneity".

We thank the Reviewer for highlighting this aspect of novelty.

470 – Well crafted the "Potential applications"! At the third point, the one on Earth System Models, I would specify that these constitute the CMIP6 and soon to the CMIP7 set of simulation models contributing to the IPCC reports and their important guidance on present and future climate.

We thank the Reviewer for this helpful suggestion. We have revised the Discussion to explicitly link the proposed applications to contemporary Earth system modelling frameworks, including CMIP6 and forthcoming CMIP7 simulations, which contribute to IPCC climate assessments. This clarification has been added to Section 4.2 (Lines 617–619).

Final note on clarifying the approach in the introduction.

As a reader I struggled at first to understand the use of predictors in this work. I think the authors could make an effort in the text to frame their work and clarify that combining different data sources forces to deal with gaps that need to be dealt with.

So on predictors, I would suggest adding a basic clarification, something like:

> Predictors are used to bridge the gap between sparse GEDI LiDAR observations and the need for continuous, large-scale forest structure maps. They act as observable proxies - derived from optical (Sentinel-2), radar (Sentinel-1, ALOS), and texture metrics - that contribute to describe canopy height, cover, and complexity. By feeding these predictors into a machine learning model, the study extrapolates GEDI-derived structural metrics across the

entire domain, enabling wall-to-wall mapping and regular updates using available satellite data.

We thank the Reviewer for this very helpful suggestion. We agree that the role of predictor variables benefits from clearer framing, particularly for readers less familiar with machine-learning–based upscaling of LiDAR observations.

In the revised manuscript, we have added a concise clarification in the Introduction (Lines 64–66) explaining that predictor variables are used to bridge the gap between sparse GEDI LiDAR observations and the need for continuous, wall-to-wall maps of forest structural diversity. The added text clarifies how predictors act as observable proxies for canopy structure and enable the spatial extrapolation of GEDI-derived metrics within a machine-learning framework. Specifically we added the following text:

In this context, predictor variables derived from complementary satellite observations are used to bridge the gap between sparse GEDI LiDAR measurements and the need for continuous, wall-to-wall maps of forest structural diversity. These predictors act as observable proxies for canopy height, cover, and structural complexity, enabling the spatial extrapolation of GEDI-derived structural metrics across Europe within a machine-learning framework.