Response to Reviewer #1

We sincerely thank Dr. Jonas Lembrechts for the thorough, insightful, and highly constructive review of our manuscript! Below we provide a detailed point-by-point response, including a marked-up revised version of the manuscript (text in green showing the changes made).

General Comment- This is a nice global collaborative study bringing together a mindboggling number of tree DBH-measurements from across the globe, and using these to make 3 km-resolution global maps for a few key metrics.

This will be a highly valuable product for many as the need for more forest structural data increases. It's a bit of a pity that the resolution is not higher though (1 km would have brought it in line with so many other datasets). Perhaps I'm missing a bit of discussion in the paper why a higher resolution was not feasible, despite the number of observations. I assume it has to do with the relatively low R2-values, which indicate the presence of a lot of noise and unexplained variance. Perhaps a discussion of that noise would also be helpful: what makes the data so noisy? Is it disturbance or land use history?

Response- We sincerely thank the reviewer for this insightful suggestion. We agree that the rationale for selecting the 0.027° prediction resolution should be explained more clearly. Although several predictor datasets used in this study were originally available at finer spatial resolutions (e.g., 30 m forest height and 1 km bioclimatic variables), the final dataset was generated on a common 3 km forest grid to provide a globally consistent prediction framework. The selected prediction resolution represents a practical balance between the native resolution of the predictor datasets, the spatial support and heterogeneous global distribution of the forest inventory observations, and the computational demands of generating wall-to-wall global predictions. Although a finer prediction grid could provide greater spatial detail, it would substantially increase the number of prediction cells and computational requirements while not necessarily improving the reliability of predictions, particularly in regions with sparse inventory observations.

We have also expanded the Discussion in the revised version to clarify the sources of unexplained variance in the models. The relatively modest model performance (R^2, RMSE, and

MAE) is expected given the global scope of the study and reflects the substantial ecological heterogeneity encompassed by the dataset. Forest diameter structure is influenced not only by the climate, topography, and soil variables included in this study, but also by factors that are difficult to represent consistently at the global scale, including natural and anthropogenic disturbances (e.g., fire, windthrow, pests, and logging), land-use history, and local edaphic and biotic conditions. Furthermore, differences in inventory protocols, plot sizes, sampling designs, and measurement timing among contributing datasets introduce additional variability. Collectively, these factors contribute to the observed unexplained variance and limit the predictive power of globally consistent environmental predictors alone.

##
##

Comment- L1: I think "using a machine learning approach" would read better.

Response- Corrected. Accordingly, the title has been revised to: "Mapping the global forest diameter spectrum using a machine learning approach."

Comment- Abstract: I wonder if it makes sense to highlight some results as well. For example, do the findings confirm existing literature that temperate forests have the largest mean diameters (Dmean), or is that new information?

Response- We thank the reviewer for this valuable suggestion. We agree that briefly highlighting representative findings in the Abstract helps demonstrate the utility of the dataset. Accordingly, we have revised the Abstract to include a concise summary of the principal spatial patterns revealed by the predicted global maps, including generally larger predicted tree diameters in temperate forests, smaller diameters across boreal forests, and pronounced regional heterogeneity throughout tropical forests. As this manuscript is intended primarily as a data description paper, we have intentionally kept the ecological interpretation brief and focused on illustrating the description of the dataset. A more comprehensive investigation of these ecological patterns and their underlying drivers is currently the subject of ongoing analyses.

Comment- Related, wouldn't it make sense to have one paragraph at the end of the paper to discuss whether the observed patterns are at least roughly in line with the literature, or whether these patterns are novel?

Response: We thank the reviewer for this thoughtful suggestion. We agree that providing brief context for the predicted spatial patterns helps readers better interpret the dataset. Accordingly, we have added a short paragraph to the Discussion relating the broad predicted spatial patterns to previous observations of global forest structure where appropriate (e.g., *Lutz et al., 2018, Nabuurs et al., 2026*). Specifically, we note that the prevalence of monospecific regular stands in Europe, as described by Nabuurs et al. (2026), may contribute to the relatively large mean diameters predicted for temperate oceanic forests. Given the scope of this data descriptor, these comparisons are intended to provide general context for the predicted patterns rather than a comprehensive ecological discussion.

Comment- L514–515: It is not intuitive to me what the difference between a "fixed-area plot" and an "inventory unit" is.

Response- We thank the reviewer for identifying this point requiring clarification. A fixed-area plot refers to an individual sampling plot with predefined dimensions in which all trees meeting the inventory criteria are measured. In contrast, an inventory unit refers to the complete sampling configuration used by certain national forest inventories, which may consist of multiple fixed-area plots, nested plots, or clustered sampling designs. Thus, the distinction reflects differences in sampling protocols rather than simply plot size. To clarify this, we have corrected the wording in the revised manuscript as follows: "Plot size and sampling designs varied across the contributing datasets due to differences in inventory protocols, ranging from individual fixed-area plots to inventory units comprising multiple fixed-area, nested or clustered subplots."

Comment- L516: How exactly is data expanded?

Response- We appreciate the reviewer's request for clarification. We realize that the term "expanded" could be interpreted as estimating tree diameters beyond the sampled plot, which was not our intention. In this study, no additional trees were inferred outside the sampled plots. Rather, plot-specific expansion factors supplied by the original inventory protocols were used to standardize observed tree counts to an equivalent trees-per-hectare (TPH) basis, accounting for

differences in plot area and sampling design among inventories. Individual tree DBH measurements were not modified during this process. To clarify this, we have revised the wording in the manuscript as follows: ″These expansion factors convert the observed number of trees within each sampled plot to an equivalent density per hectare based on the plot area and sampling design; they do not estimate additional trees outside the sampled plot.″

Comment- L533: Add an example of such a name, perhaps?

Response- We agree that including an example improves the clarity of the naming convention. Accordingly, we have corrected in the revised manuscript to include a representative example of the standardized genus-level identifier assigned to individuals identified only to morphospecies. For example, an individual identified only to the genus *Allophyllum* from dataset 73 is assigned the identifier Allophyllum_sp_73_001, where *Allophyllum* is the genus name, sp denotes an unidentified species within that genus, 73 is the source dataset identifier (DSN), and 001 is a unique numeric suffix. This convention applies only to specimens identified to the genus level but not to species. Taxa lacking a genus-level identification were assigned standardized identifiers based on the lowest available taxonomic rank (or an "Unknown" placeholder when no taxonomic assignment was available), following the same DSN and unique suffix convention.

Comment- L549: I assume the exclusion based on the 10 cm threshold does not apply to the summing of multi-stem diameters? The order of sentences now makes it slightly misleading.

Response- We agree that the original order of the sentences could be interpreted as applying the 10 cm DBH threshold before combining multi-stem individuals. This was not our intention. We have revised the text to clarify that stem diameters were first combined into an equivalent diameter based on summed cross-sectional area, after which the 10 cm DBH threshold was applied to the equivalent tree diameter. This ensures that the inclusion criterion is consistently applied to the whole tree rather than to individual stems.

Comment- L551: how often was disturbance information available? Is there a risk that many observations influenced by disturbance are still included, because the information was not available? If this information was rare, satellite-based proxies of disturbance may have to be used.

Response- We thank the reviewer for this important observation. Disturbance information was available for many of the contributing inventory datasets and was used to exclude recently disturbed plots whenever such information was provided by the original inventory. Following the quality-control, harmonization, and filtering procedures including removing disturbed plots, the compiled database was reduced from 1,787,296 to 1,203,524 forest inventory plots. However, because metadata standards and inventory protocols differed among contributing inventories, disturbance information was not consistently available across all datasets, and consequently some disturbed stands may remain in the final database. We have revised the manuscript to explicitly acknowledge this limitation and clarify that the mapped diameter attributes should be interpreted in the context of the available disturbance information. We also note that future versions of the dataset could explore the incorporation of globally consistent satellite-derived forest disturbance products (e.g., Wang et al., 2026) to improve the identification of disturbed stands where inventory-based disturbance information is unavailable.

Comment- L618: how did you screen for multicollinearity?

Response- Prior to model development, pairwise Pearson and Spearman correlation matrices were used to assess relationships among predictor variables. Predictor pairs with an absolute correlation coefficient ($|r| \geq 0.8$ were considered highly correlated, as this threshold is widely used to identify substantial redundancy while retaining variables that provide complementary ecological information. Variables that could introduce data leakage or circularity, including response variables and related derived products such as plot-level basal area and aboveground biomass, were removed before model fitting. The remaining predictors were retained because the primary modelling approach was Random Forest, which is generally robust to multicollinearity. We have also clarified this in the revised version.

Comment- L639: values of what? Of all predictor layers?

Response- Corrected. The manuscript has been revised to clarify that values from "all predictor layers" were extracted at the geographic coordinates of each forest inventory plot and subsequently used as explanatory variables for model development.

Comment- L639: I'm a bit worried about the idea of extracting values only at the centroid of the pixel. Wouldn't the mean of the pixel be more representative? What if the centroid of the pixel just has a forest gap?

Response- We thank the reviewer for this thoughtful comment. We agree that extracting predictor values directly from a single fine-resolution pixel could potentially be influenced by localized features such as forest gaps. However, this was not the case in our workflow. The original 30 m forest height product was first aggregated to 1 km resolution and subsequently to 3 km resolution by averaging the corresponding 1 km pixels. During predictor extraction, the forest inventory plot coordinates were used only to identify the corresponding 3 km grid cell, and the extracted value therefore represents the mean forest height of the entire 3 km cell rather than the value of an individual fine-resolution pixel. Consequently, localized features such as small canopy gaps are unlikely to disproportionately influence the extracted predictor values. We have revised the text accordingly.

Comment- Table 2: could you make a clearer distinction between the categories?

Response- Corrected through improved alignment and table formatting

Comment- L677: do you have references that argue for modelling separately per ecozone? Intuitively, I would say that by reducing the environmental range in your dataset and the number of datapoints, you reduce the explanatory power of your model. Also, these machine learning models deal well with non-linearity, so why is the splitting necessary?

Response: We thank the reviewer for this insightful comment. We agree that machine-learning algorithms such as Random Forest and XGBoost are well suited to modelling complex nonlinear relationships. However, our rationale for developing ecozone-specific models was not solely to address nonlinearity, but also to account for the substantial ecological heterogeneity and spatial non-stationarity among global forest ecosystems. Forest diameter structure is influenced by

region-specific combinations of climate, species composition, disturbance history, management practices, and environmental conditions, resulting in predictor–response relationships that vary among ecozones. During model development, we also evaluated a global Random Forest model. However, the global model did not adequately capture ecozone-specific differences in forest diameter structure. Although partitioning the data reduces the number of observations available for each individual model, it also reduces ecological heterogeneity within each modelling domain, allowing the models to better capture region-specific ecological relationships while reducing extrapolation across contrasting environmental conditions. Similar regionalized modelling approaches have been adopted in previous global forest studies that stratified modelling by ecological regions or biomes (e.g., *Crowther et al., 2015; Huang et al., 2021; Mo et al., 2023*).

Comment- L693–695: separately across each model technique, or averaged across techniques?

Response- We thank the reviewer for requesting this clarification. The reported performance statistics (R^2 and RMSE) were calculated separately for each modelling technique (OLS, RF, and XGBoost).

Comment- L727: and which function for the OLS?

Response- To improve the reproducibility of the modelling workflow, we have corrected in the revised manuscript to explicitly specify the R function used for the ordinary least squares model. Specifically, the OLS model was implemented using the *lm()* function in R.

Comment- Figure 2: I'm surprised to see such high values—a lot of green colors, e.g. in Europe, with values far above the mean and standard deviation as reported in Table 1.

Response- We thank the reviewer for this thoughtful observation. The apparent discrepancy arises because Table 1 and Figure 2 summarize different aspects of the dataset and therefore are not directly comparable. Table 1 reports the summary statistics (mean and standard deviation) of the observed forest inventory measurements across plots, whereas Figure 2 presents the wall-to-wall spatial predictions generated by the Random Forest models for all forested grid cells globally. Consequently, Figure 2 illustrates the spatial distribution of model-predicted values across all forested grid cells rather than the statistical distribution of the observed inventory measurements summarized in Table 1. Therefore, the two summaries are complementary and are

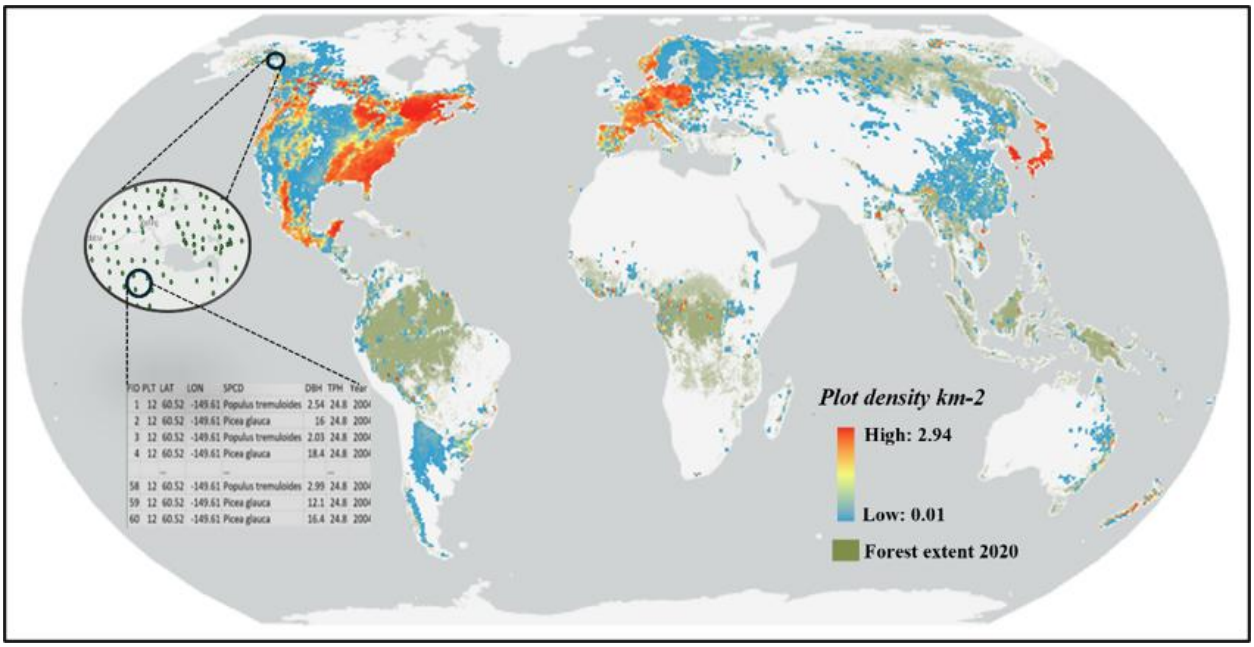
not expected to match directly. Localized regions containing mature forests or particularly favorable growing conditions may exhibit relatively high predicted mean diameters without being inconsistent with the global summary statistics reported in Table 1. Furthermore, the mean ± standard deviation describes the central tendency and variability of the observations rather than their full range. The underlying inventory data include substantially larger observed mean diameters across several ecozones (maximum observed Dmean of 152.6 cm), whereas the maximum predicted value shown in Figure 2 is 77.6 cm, indicating that the spatial predictions remain well within the range of the observed data used for model development.

Comment- Figure 2: do we see abrupt changes between ecozones, stemming from the running of separate models?

Response- We thank the reviewer for this insightful question. Separate models were developed for each ecozone to account for regional ecological heterogeneity and were subsequently combined to generate continuous global prediction maps. Visual inspection of the aggregated predictions did not reveal obvious artificial discontinuities along ecozone boundaries. Although independent models were fitted within each ecozone, neighboring ecozones often share similar environmental gradients and predictor values, resulting in relatively smooth transitions across boundaries. Where localized transitions occur, they are consistent with underlying ecological gradients and differences in forest structure rather than apparent modelling artefacts.

Comment- Finally, I would love to see a map of data coverage (i.e. observation points) in the main text.

Response- We agree that visualizing the spatial distribution of inventory observations provides valuable context for interpreting the prediction maps and understanding regional variation in sampling density. Accordingly, we have added a new figure (**Figure 2**) to the main text showing the global distribution and density of the approximately 1.2 million forest inventory plots used in this study. We believe that this figure improves transparency regarding the spatial coverage of the underlying observations and helps users identify regions with relatively dense and sparse sampling.



Global distribution and density of forest inventory sample plots used to develop the forest stand attribute maps. Geographic distribution of 1,203,524 sample plots, with plot density (plots per 1 km²) shown as a heat map over the global forest extent (green). The map is presented in the Robinson projection (WGS84 datum). Insets: Enlarged view of sample plot locations in south-central Alaska and an example of the raw tree-level inventory data.