To Reviewer #2:

Thank you for your time and effort to review our manuscript. Each request or comment is repeated below in black, and our responses are in blue. Additional or altered text that has appeared in the manuscript is marked in red.

[General Comment] This paper utilizes an impressive dataset of 48,000 wood density samples from forest in Poland to investigate within and among tree variations in wood density. Taxonomic and landscape factors are correlated with tree density, then a feature selection and random forest approach is used to model spatial variation in wood density using remote sensing metrics.

While I think this dataset and findings are interesting, there is some disconnect between the objectives and analyses, and the analytical approach (and interpretation therein) needs improvement. Moreover, the introduction and discussion are both lacking in depth of narrative development and interpretation and lacks a thorough review of the wide body of literature that has focused on wood density variability.

[Response] Corrected.

[Detailed comments] For this review, however, I will focus primarily on the analytical approach, as I think corrections must be made here first before the interpretation of results in the discussion can be evaluated.

1. Were density samples collected to minimize inclusion of compression or tension wood? Xylem cells can develop strong differences in cell wall thickness (and thus density) due to directional effects of topography/wind speed which could bias density measurements and possibly inflate within-tree variability. It's mentioned that rings were sampled for radial profiles in north and south directions, so it seems likely that some compression or tension wood has been included and possibly influencing results. This needs clarification.

[Response] Thank you for this comment. During the measurement process, each ensity sample was examined for wood defects, including compression and tension wood, knots, resin wood, cracks, abnormal shapes after drying, and other defects. For the presented research samples without any defects have been taken for further analysis due to the standard of small samples density measurement. However, there is still a very interesting base of complete samples with defects influencing the "real" wood density which one can measure with the whole discs without removing and cutting off any wood defects. We have clarified this in the revised manuscript, as follows: "Each disc was cut from north to south to obtain a strip of wood. The samples were divided and numbered into two rays: north and south, starting from the core to the peripheral part. This method allowed for the estimation of variation in the radial density of wood. The number of samples obtained for each disc varied depending on the width of the disc, but each disc typically yielded more than 10 samples along these radial directions. Standardized wood density samples, measuring 2×2×3 cm, were cut from the strips, which were dried in a dryer at temperature of 103 ± 2 ºC to an absolutely dry state. After the samples cooled down in the desiccator, the linear dimensions of the samples were measured using an electronic caliper, and their weight was measured on a laboratory scale. The stereometric density was then calculated from the classical mass/volume formula. During measurement, each density sample was examined for wood defects such as compression and tension wood, knots, resin wood, cracks, abnormal shapes after drying, and other irregularities. In this study, only defect-free samples were selected for further analysis, adhering to the standards for small sample density measurement." (Page: 5; Lines: 232-241)

2. A better description of the predictor data is needed. Line 133 references Table 1, but there are no tables present in the version of this manuscript that I reviewed or in the supplemental. What is the spatial resolution of the remote sensing data or the DEMs used to calculate geomorphons? Why were the specific spectral indices selected and from what satellite products were they computed?

[Response] We apologize for the oversight. Table 1 has been added, and the sources and spatial resolution of covariates are now listed. We have verified that none of these data have reported data quality issues in Central Europe.

Table 1. The predictor covariates used in the random forest model for inter-tree variations in wood density. The original 8-daily values of NDVI and NDWI were aggregated into a median (P50) and a standard deviation (STD) for the entire period.

Variables	Description	Unit	Original resolution	Source
SNDPPT	Weight percentage of the sand particles (0.05–2 mm)	%	250 m	SoilGrids database
NDWI	8-daily Enhanced Vegetation Index (EVI) generated using the	1	0.083º	MOD13A2

	gridded daily surface reflectance product.			
NDVI	8-daily Normalized Difference Vegetation Index (NDVI) generated using the gridded daily surface reflectance product.	1		
Geomorphons	a pattern recognition approach to classification and mapping of landforms from digital elevation models (DEMs)	-	30m	Jasiewicz & Stepinski (2013)

3. A better description of the spatial sampling design is needed. The authors explain different height and age classes, and the total number of samples/trees/plots in Fig 1, and that 30+ samples were collected per tree, but the number of trees in each plot is not provided. This is quite important as it defines the data's hierarchical structure and should drive the analytical approach.

[Response] Thank you so much for your comments. We have added the sentence in the revised manuscript: "Our dataset includes the density of more than 48,000 samples taken from 2,920 trees, and from 391 forest plots in Poland, all carried out in the year 2018 (Figure 1). The number of trees per plot varies, averaging 6.7 ± 3.0 trees." (Page: 3; Lines: 133-134). Following your suggestion, we have performed a generalization linear mixed effects model to assess these data. Please refer to our response to comment #4 for further details.

4. With hierarchical data such as this, ANOVA is not an appropriate analytical approach. These data should be analyzed with a generalized linear mixed effects model so that the appropriate variance can be partitioned to the random effects (e.g. sample, tree, plot) rather than being fully attributed to the fixed effects (leaf type, family, species, age, dbh, etc). Without a proper hierarchical analysis, the results are challenging to interpret with confidence. Using a GLMM would also allow for all density samples to be pooled into a single analysis (with random effects for sample nested in tree nested in site), implicitly accounting for within tree variation in density, rather than averaging all density measurements per tree and then analyzing at the tree-level. This approach also increases degrees of freedom substantially, thereby increasing interpretive power.

[Response] Following the reviewer's suggestion, we applied a generalized linear model to the sample-level measurements (see Table R1), as well as two generalized linear mixed-effects models with "tree" and "plot" as random effects, respectively (see Table R2). The results indicate that while there is a relationship between wood density and variables such as tree species, families, leaf types, and ages, these factors alone account for only 17% of all the variation in wood density samples. This is consistent with our finding that the intra-tree variability of wood density within individual trees is substantial and cannot be fully explained by species, families, leaf types, or ages. However, when the data are grouped by tree, the model explains 74% of the variation in wood density (Model 2 in Table R2), using all sample values as predicted variable, and "tree" as a random factor. This result confirmed that our approach of linking tree-averaged wood density with tree species, families, leaf types, and ages is feasible. We acknowledge that calculating the tree-level average leads to the loss of intra-tree variation in wood density, which indeed cannot be explained by these factors.

Table R1. The result of generalized linear model for predicting wood density variation among all samples using fixed effects of species, families, leaf type, and ages.

	Model 1 (Sample level)		
	Beta	SE	p-value
Species	0.002	0.001	0.059*
Families	-0.011	0.002	<0.001**
Leaf types	0.082	0.002	<0.001**
Ages	0.001	<0.001	0**
$R^2 = 0.17$			

Table R2. The result of generalized linear mixed effects model for predicting wood density variation using fixed effects of species, families, leaf type, ages, and random effect of tree and plot.

Model 2 (Tree level)				Model 3 (Plot level)			
Fixed effects	Beta	SE	p-value	Fixed effects	Beta	SE	p-value
Species	-0.0001	0.004	0.98	Species	-0.003	0.001	0.001**
Families	-0.013	0.007	0.08*	Families	0.009	0.002	0**
Leaf types	0.111	0.009	0**	Leaf types	0.106	0.003	0**

Ages	0.001	<0.001	0**	Ages	0.0007	<0.001	0**
Random effect		Variance	SD	Random effect		Variance	SD
Intercept: Tree		0.010	0.099	Intercept: Tree		0.06	0.079
$R^2 = 0.74$				$R^2 = 0.43$			

We have added the following text to clarify this in the revised manuscript: "While our approach of linking tree-averaged wood density with tree species, families, leaf types, and ages is validated by the strong explanatory power of the generalized linear mixed-effects model (Table S2), we recognize that this method inherently averages out significant intra-tree variability. This variability, which cannot be fully accounted for by these factors alone, is an important aspect of wood density dynamics that warrants further investigation. Therefore, our findings should be interpreted with the understanding that the tree-level averages, while useful, may not capture the full complexity of wood density variations within individual trees." (Page: 9; Lines: 345-350)

5. It's not clear in the objectives why remote sensing data were used for the random forest modeling when they weren't used for any of the other analyses. Was this an attempt to be able to predict wood density in areas where samples were not collected but remote sensing data is available? If so, the authors need to explicitly state this objective.

[Response] Thanks for your comments. Actually, predicting wood density in areas without samples is not the purpose of this analysis. We have already achieve this globally in another analysis (https://onlinelibrary.wiley.com/doi/full/10.1111/gcb.17224). Here, our aim is to identify the key climatic, edaphic, and biotic factors controlling the spatial variations in observed wood density among trees, and to explore how wood density varies in response to these key factors. Since the climatic, edaphic, and vegetation properties (beyond wood traits) were not recorded during sample collection, we used the high-resolution satellite products and observation-based climate products to provide the local environmental conditions and vegetation state. In addition, we have added the sentence to state this clearly, as follow: "To investigate the key climatic, edaphic or vegetation-related factors influencing the spatial distribution of tree-level wood density, we extracted the relevant predicted variables from high-resolution satellite products and observation-based climate products, based on the latitude and longitude of samples, and employed a feature selection method (Jung and Zscheischler 2013) to identify the most significant predictors." (Page: 6; Lines: 277-280)

6. I don’t understand why a feature selection procedure was used prior to random forest modeling (lines 130-131). Random forest has a built in feature selection process and is, for the most part, robust to a large number of predictors (assuming hyperparameters are appropriately defined). This two-part process may be omitting predictors that RF may have otherwise identified as important.

[Response] Thanks for your comment. We performed a feature selection prior to random forest modelling for following reasons:

(a) The total of the extracted climatic, edaphic or vegetation-related covariates variables from high-resolution products exceed 2,000, and computational limits prohibit the use of all covariates in the random forest modelling;

(b) In data-driven modelling, choosing the right predictor based on prior experience or knowledge can bring some biases. Therefore, we aimed to identify the most informative predictors directly for building the random forest model;

(c) The feature selection method used in this analysis can return multiple optimal solutions if they exist, which is important for interpreting the results (Jung and Zscheischler, 2013).

7. More information is needed on the random forest modeling methods, which package/library was used. How were hyperparameters selected or tuned? Were data centered or scaled in any way? Are random forests run at the sample-, tree-, or plot-level? If at tree- or plot-level how were density measurements aggregated? If averaged, were the individual density measurements normally distributed or was a median reported?

[Response] Thank you for pointing it out. We have added the information of the random forest modelling method, as follow: “The R package ‘randomForest’ was used in this analysis to build a random forest model with 500 trees. The tree-level wood density, calculated as the average value of all wood density samples within each tree, was randomly partitioned into training and testing subsets, with 80% of the measurements allocated to the training set and the remaining 20% reserved for testing.” (Page: 6; Lines: 283-287)

Additionally, As the reviewer mentioned, we calculated the average of sample-level wood density to represent the tree-level wood density. Firstly, to determine if the average value is a good representation of population, we compared the mean and median values within each individual trees. As shown in Figure R2a, the difference between mean and median values are minor (slope: 1.02, $R^2 = 0.99$). This suggests that the wood density samples within individual tree tend to follow a normal distribution, and the tree-level wood density can be effectively represented by the average of the sample-level values. Secondly, we examined whether the intra-tree variability of wood density differs among tree species. Figure R2b shows that, despite varying tree diameters, the magnitude of intra-tree variability is similar across tree species, with no significant differences observed.

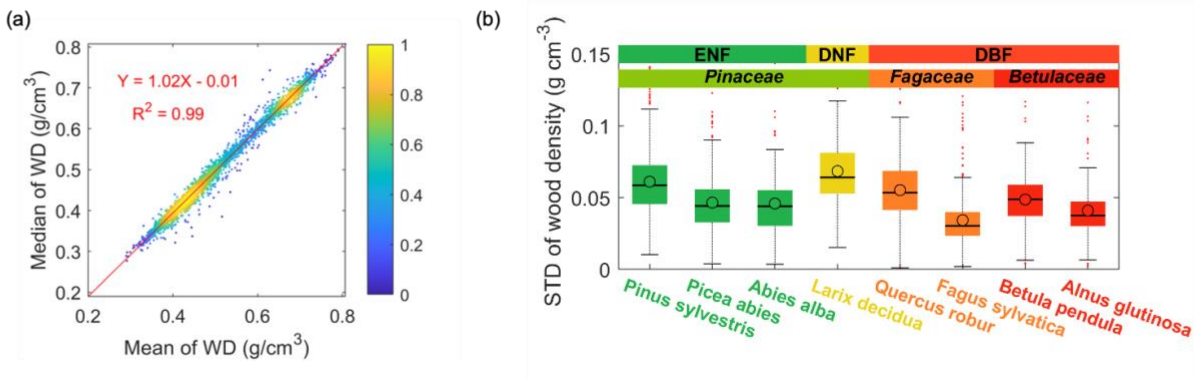

Figure R1. (a) Comparison of the mean and median values of wood density within the individual trees. (b) The standard deviation (STD) of the sample-level wood density within the individual trees. Boxplots indicate the median, mean, minimum, maximum, and the 25th and 75th percentiles of wood density for eight tree species.

We also added the following text: "Tree-level wood density was calculated by the average of all samples within each individual trees. This method is used because there was no significant difference between the mean and median values of the samples (Figure S1a), indicating that wood density within an individual tree typically follows a normal distribution. Furthermore, the magnitude of intra-tree variability is consistent across eight tree species (Figure S1b)." (Page: 6; Lines: 274-277)

8. Out of bag estimates from random forest models are not "predictions," (e.g. lines 193, 265, 324) they report a variance explained only for the model fit to the data i.e. OOB estimates are for variance explained on data subsets withheld when each tree is defined. If the authors would like to report model "predictions" then the model needs to be tuned to a subset (70-80%) of the data (preferably with entire sites/plots omitted) and tested on another subset that has been entirely withheld from model tuning process. This also means that the 91% reported accuracy is likely overestimated in terms of predicting

wood density estimates from "new" or withheld data. In simpler language: as reported, the random forest model explains how well the model captures variation in the data, but that has little bearing on the ability of the model to predict (i.e. extrapolate) density estimates using spectral data from trees/areas not included in the dataset used to tune the model.

[Response] Thank you for your comments. Firstly, we completely agree with the reviewer on the importance of assessing the performance of model in predicting data not used during training. To assess this, we indeed partitioned the wood density data into training and testing subsets, randomly allocating 80% of measurements for training and the remaining 20% for testing. We have clarified this in the revised manuscript, and we provided the scatter plots comparing predictions and observations for both training and testing subsets (Figure R2). Secondly, the out of bag (OOB) error serves as a cross-validation accuracy metric. It helps avoid overfitting and assesses the prediction performance of the random forest model for samples not seen during the training. Therefore, we believe that the OOB estimates are also important and should be shown in the manuscript.

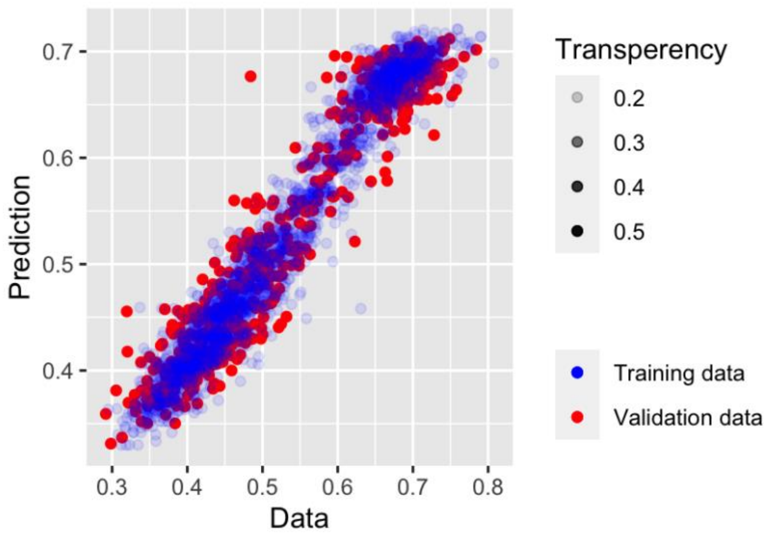

Figure R2. Comparison between predictions from the random forest model and observations for both training (blue dots) and testing (red dots) subsets.

9. Technically, spatial data inherently violate model assumptions of data independence for random forests. While it is increasingly common to use RF for modeling spatial data, the authors need to acknowledge potential biases in model performance and partial

effects of predictors due to spatial autocorrelation. There are several packages available to run random forests on spatial data such as the 'spatialRF' package in R.

[Response] Thanks for your comments. We agree with the reviewer that spatial autocorrelation can lead to a positively overestimation of the prediction skill of machine learning model (Ploton et al., 2020). Thus, a spatial blocked cross validation is indeed important for models used for prediction (as what we did in Yang et al. 2024).

However, in this analysis, the random forest model was designed to find the influencing factors for spatial variations in observed wood density, rather than to predict wood density. Thus, the risk of extrapolation is not a primary concern in this context. In this case, we extracted predictor variables from very high-resolution satellite or observation-based products. If spatial autocorrelation exists in these climatic, edaphic and vegetation variables, and if similar spatial autocorrelation is observed in the wood density measurement, it can help identify key factors influencing the spatial pattern of wood density.

[Minor comments] Line 17 – Typo: "representing" should be "represented"

[Response] Corrected.

Line 19 – Define or describe geomorphons at first mention rather than in following sentences.

[Response] We have rewritten this part as follows: "Geomorphons, which are landform elements derived from digital elevation models (DEM) and soil sand context provided insights into wood density variations. Lower wood density values were linked to landforms characterized by low geomorphons, such as summit, ridge, or shoulder. Conversely, higher wood density was found in landforms with high geomorphons, including valley, depression, or hollow areas." (Page: 1; Lines: 19-22)

Lines 64 and 76 – This study looks at variation *among* trees (i.e. comparison among many) rather than between trees – "between" implies a comparison between two individuals.

[Response] Corrected

Line 80 – What does it mean that trees were aged 5 years?

[Response] In this study, we used the tree samples from trees older than five years. This is because the younger trees were not selected during the sampling process, as they would need to be cut down for inclusion in the study. We have clarified this in the revised manuscript: "Since the sampling process involves cut down trees, only those older than 5 years were included in this study. These trees, representing a range of species, had their relevant information such as latitude, longitude, age, and species type recorded." (Page: 3; Lines: 134-136)

References

Jung, M., & Zscheischler, J. (2013). A guided hybrid genetic algorithm for feature selection with expensive cost functions. Procedia Computer Science, 18, 2337-2346.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., ... & Pélissier, R. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nature Communications, 11(1), 4540.

Yang, H., Wang, S., Son, R., Lee, H., Benson, V., Zhang, W., ... & Carvalhais, N. (2024). Global patterns of tree wood density. Global Change Biology, 30(3), e17224.