Reviewer 1 comments (RC1)

RC1 – This manuscript entitled “Wood density variation in European forest species: drivers and implications for multiscale biomass and carbon assessment in France” presents an analysis of wood density variation across temperate forest species in France. The study makes a valuable contribution to understanding wood density variation and its implications for forest biomass estimation. The authors constructed four linear models to identify the biotic and abiotic factors controlling wood density variability. They also applied the NFI-based and GIS-based models to generate a spatial distribution map of wood density across France. Finally, they evaluated the influence of wood density on biomass estimation at multiple scales, such as the plot, subregion, and country levels. A particularly interesting conclusion is that the choice of method for inferring wood density depends on the spatial scale of interest.

Author’s Answer (AA) –Thank you for this fine assessment and positive feedback.

I have several comments regarding the methods and overall storyline of the study

Specific comments:

RC1 – First, the results from the taxonomic and NFI-based models highlight the importance of tree species in explaining wood density variation. However, this may be partly due to the limited set of climate variables included. The authors only included two variables, mean annual temperature and precipitation. I think only temperature and precipitation may not fully reflect spatial differences in specific environmental conditions. It would be helpful to clarify why only these two climatic variables were selected. Would the inclusion of additional climatic and soil variables (such as soil nutrient availability, soil pH, or atmospheric humidity) alter the identified drivers?

AA – We appreciate the reviewer’s suggestion and tested the inclusion of additional variables in our environmental model, specifically the soil carbon-to-nitrogen ratio and the summer soil water deficit (June–August). The model already incorporated an indicator of soil nutrient availability, originally termed the “trophic level index” which we have renamed in this revised version as the “Soil nutrient availability index” to improve clarity. These additions led to a slight improvement in model fit (AIC decreased from 899,000 to 886,000) and only a marginal increase in explained variance (R^2 from 0.14 to 0.17), which does not alter our conclusion that environmental influence remains limited and tree species is by far the dominant driver of wood density variation in mainland France.

Our decision to focus on mean annual temperature and precipitation was guided by two constraints: (i) climatic and soil datasets must be consistently available across the entire French territory, and (ii) they must represent long-term mean conditions relevant to our objective of analysing spatial variation in average wood density. Many additional variables are either highly correlated with temperature and precipitation or available only at limited spatial scales, thereby restricting their independent explanatory power in this context.

We tested further variables—including soil pH, the annual number of frost days, the annual number of rainy days, and the De Martonne aridity index—but they introduced high collinearity while yielding only marginal improvements in model performance. We therefore added the following clarification (Page 9, lines 228-230):

“Although additional variables—including soil pH, annual number of frost days, annual number of rainy days, and the De Martonne aridity index—were initially tested, they were excluded due to high collinearity and negligible gains in model performance.”

It is also important to emphasise that our study addresses spatial variation in mean wood density, rather than interannual variability as captured by tree-ring analyses. This approach differs fundamentally from dendroclimatological studies, which typically analyse wood density (often maximum latewood density) particularly in environmentally stressful settings such as treeline or latitudinal limits. By construction, our framework minimises the influence of short-term climatic fluctuations and instead highlights long-term spatial patterns.

We have clarified in the *Data & methods* section which climatic and soil variables were tested, and we added discussion to highlight why climate effects appear limited in our spatial framework (Page 24, lines 559-565):

“Beyond this spatial dimension, our study also differs from dendroclimatological research in its temporal focus. Dendroclimatology typically targets interannual variability through tree ring analyses, most often by analysing maximum latewood density (MXD) responses to seasonal or extreme year anomalies, and in climatically stressful settings such as tree line or high latitude environments (Briffa et al., 1998; Hughes et al., 1984). In contrast, our approach investigates spatial variation in wood density averaged over the full lifespan of trees. By construction, this framework integrates long-term climatic fluctuations and reduces the influence of short-term climatic fluctuation. It therefore emphasises persistent spatial patterns in wood density across broad biogeographical gradients rather than year-to-year climatic sensitivity.”

RC1 – Second, I think the inclusion of four distinct models seems somewhat redundant. The environmental model, in particular, appears less critical, as it overlaps significantly with the GIS-based model in terms of predictive capacity using abiotic variables. The GIS-based model, which incorporates both biotic and abiotic factors, could sufficiently illustrate the predictive power of remotely accessible variables (including environmental variables) without the need for a separate environmental model.

AA – In the revised version, we expanded the environmental model to include additional variables (see response to Comment 1). These variables were not all incorporated into the GIS-based model because they are not necessarily remotely accessible, so that their homogeneous spatial availability is not guaranteed. As a result, the environmental model is no longer simply a subset of the GIS-based model, but rather serves a distinct purpose: it allows us to specifically examine the role of environmental drivers in explaining wood density variation.

The GIS-based model, by contrast, integrates both biotic and abiotic predictors that are consistently available across space, thereby illustrating the predictive capacity of remotely accessible variables. We therefore consider the two models complementary: the GIS-based model demonstrates the utility of accessible predictors for large-scale applications, while the environmental model provides a more focused analysis of environmental influences. For this reason, we prefer to retain both, as they now address different questions and contribute distinct insights.

RC1 – Third, given the hierarchical nature of the data (e.g., trees nested within species and plots), the use of ANOVA is not optimal. A generalized linear mixed-effects model (GLMM) would be more appropriate, as it allows variance to be partitioned into random effects (e.g., tree, species, plot) rather than attributing it entirely to fixed effects (e.g., genus, DBH, height). A GLMM would enable all density samples to be analyzed collectively, with nested random effects (e.g., trees within plot and species), thereby improving degrees of freedom and interpretive power.

AA – We agree with the reviewer that ANOVA is not the optimal approach given the hierarchical structure of the data. However, we used it only in the taxonomic model, where our specific objective was to quantify the proportion of variance attributable to botanical class, genus, and species.

With respect to mixed-effects models, it is true that our data include trees nested within plots and species. However, due to the sampling design, most plots contain fewer than two observations on average (for example, in *Quercus robur*, one of the most abundant species in mainland France, the mean number of wood-density measurements is only 1.6 per plot). This sparse replication makes it impossible to specify a random-effects structure such as 'tree within plot,' since the model would be unidentifiable under these conditions

We however tested mixed-effects models (using the *lmer* function in the lme4 package) on the NFI-based model, treating the species as a random effect rather than a fixed effect. This test led to the following conclusions:

- The variance explained and prediction errors were essentially identical to those obtained with linear models.
- Coefficients were very similar across approaches.
- Computation time was substantially longer, making it impractical to perform spatial cross-validation (as requested by Reviewer 2).
- The *lmg* method used to assess variable importance is not available within the mixed-effects framework.

Thus, while mixed-effects models did not alter our conclusions, they introduced significant computational constraints and prevented the use of key methodological tools (spatial cross-validation and variable importance analysis). For these reasons, we chose to keep our modelling framework unchanged.

Finally, one of our main objectives was to precisely quantify species effects, and our dataset was constructed accordingly by retaining only species with sufficient observations. This objective and the associated design support treating species as a fixed effect, making the linear model approach well suited to our purpose.

RC1 – Furthermore, I like this part of "Difference in forest aboveground biomass (AGB) stock depending on the method used to predict wood density", and I additionally suggest extending this analysis to more explicitly quantify the total carbon stock (e.g., in PgC of aboveground carbon) and assess potential over- or underestimation at broad scales under different wood density inference methods.

AA – In the revised version, we now systematically provide estimates of aboveground carbon stock in addition to biomass, using a carbon content of 47.5% for wood. This extension complements the biomass analysis and directly addresses the reviewer's request for a more explicit quantification of carbon.

We also added further detail on the potential over- or underestimation of carbon stocks at broad scales when different wood density inference methods are applied. As suggested by the second reviewer, we notably compared our biomass estimates with ESA-CCI product, and discussed the comparison (Page 27, lines 653-665):

"To evaluate whether current remote sensing biomass estimates for mainland France are biased by the absence of explicit consideration of spatial variation in wood density, we compared forest biomass estimates derived from our "NFI based model" with the European Space Agency Biomass Climate Change Initiative map for 2020 (Santoro and Cartus, 2025). At the national scale, the comparison indicates good overall agreement (a 4.7% difference in total biomass across mainland France, corresponding to 113 million tons, or 54 million tons of carbon), whereas at finer spatial scales, such as biogeographical subregions, notable discrepancies emerged (see Supplementary Figure 7). Importantly, these differences do not follow the direction that would be expected if they were primarily driven by unaccounted wood density variation. For example, ESA biomass estimates are higher than NFI estimates in the Mediterranean region, despite this region being characterised by comparatively high wood density. If wood density were the dominant source of bias, neglecting this high wood density should have resulted in lower biomass values from the remote-sensing product. The fact that the opposite pattern is observed indicates that wood density effects alone cannot explain the discrepancies. This suggests that other factors—such as differences in allometric assumptions, forest structure representation, or remote-sensing signal interpretation—likely contribute to the divergence in biomass estimates and merit further investigation."

RC1 – Finally, it remains unclear how the NFI-based model, which relies on species identity, can be upscaled to generate a large-scale wood density map. Species-level information is often unavailable at broad scales, and the model's dependency on such data may limit its practical application. The authors should clarify how this limitation is addressed or propose alternative approaches for spatial extrapolation.

AA – We agree with the reviewer that the NFI-based model, which relies on detailed species-level information, cannot be directly upscaled to generate large-scale wood density maps, as such data are not broadly available. We therefore consider the NFI-based model as a reference framework: it provides a benchmark against which other approaches can be compared, but is not intended for direct extrapolation.

To address the need for spatial prediction, we tested a GIS-based model that integrates remotely accessible biotic and abiotic variables. This model is specifically designed to be scalable across large spatial domains, in contrast to the NFI-based model. In the revised manuscript, we have clarified this distinction and added text to explain the complementary roles of the two approaches. For example for the NFI-based model (Pages 9-10, lines 237-240):

"By integrating data at both the individual tree and stand scales, this model was designed to maximise precision in explaining wood density variation and to serve as a benchmark for comparison with other approaches. It is not, however, intended for direct spatial extrapolation, as the underlying data are restricted to NFI plots (e.g., detailed tree measurements) and cannot be generalised to broader scales or finer resolutions."

And for the GIS-based model (Page 10 lines 258-260):

"This model was designed to assess the feasibility of reconstructing wood density variation using data available at broad scales, without relying on detailed tree- or stand-level descriptors such as those provided by NFI data and incorporated in the "NFI-based model"."

Finally, we expanded our discussion on the potential of remote sensing techniques to infer species identity at broad scales. Such advances could, in the future, help bridge the gap between species-dependent models and large-scale applications (Page 27-28, lines 673-683):

"In parallel, advances are emerging in the automated identification of tree species, spanning scales from individual trees to entire landscapes, with several initiatives carried out across metropolitan France. The recently released *PureForest* dataset exemplifies how large-scale airborne laser scanning (ALS) and high-resolution aerial imagery, combined with deep learning techniques, can provide benchmark references for robust species classification across dozens of taxa, thereby enabling scalable mapping efforts (Gaydon and Roche, 2025). Complementary studies based on multispectral satellite time series and advanced machine learning architectures have demonstrated promising classification performance in complex forest environments, highlighting the importance of dense temporal and spectral information for broad-scale species discrimination (Mouret et al., 2025). More broadly, multi-source approaches that integrate spectral, structural, and textural features derived from optical imagery and LiDAR data continue to progress, paving the way toward more accurate, robust and transferable species maps. Such methodological progress has the potential to improve biomass estimation by enabling finer representation of both tree volume and wood density in forest models (Karasiak et al., 2020)."

A more detailed description of the candidate predictor variables is needed. For instance, what is the spatial resolution of the remote sensing data used in the GIS-based model? Which specific spectral indices were selected, and from which satellite products were they derived?

AA – In the revised manuscript, we have extended the appendix (Supplementary Table S1) that provides details on all predictor variables, including how they were measured or calculated. This addition provides the necessary detail to clarify the data sources and strengthen the transparency of our modelling approach.

Reviewer 2 (Simon Besnard) comments (RC2)

General assessment

RC2 – The manuscript by Cuny et al. offers a valuable contribution by quantifying interspecific and intraspecific variation in wood density across France using the XyloDensMap dataset (≈110,000 increment cores). The authors use a clear modelling hierarchy: Taxonomic, Environmental, NFI-based, and GIS-based modelling framework, to disentangle ecological and structural drivers of wood density and to evaluate implications for biomass estimation across multiple spatial scales. The manuscript is well-structured, well-written, and analytically thorough. The figures are clean and intuitive, and the work is highly relevant for both forest ecology and the remote-sensing biomass community.

The core findings are robust and important: species identity dominates total variance (≈78.5%), intraspecific variation exists but remains difficult to model, and neglecting wood-density variation has minimal effect at the national scale but produces substantial biases (up to

30%) at subregional and plot scales. For these reasons, the study constitutes a significant step forward in improving estimates of biomass and carbon stocks.

Author's Answer (AA) – Thank you for this assessment and positive feedback.

There are, however, a couple of things that could be addressed to strengthen the MS. Below, I provide some comments for revision:

1. Cross-validation strategy requires clarification and extension

RC2 – It is currently unclear whether cross-validation was performed through random sampling of individuals or whether spatial autocorrelation was accounted for. Because both wood density and environmental variables exhibit strong spatial structure, random train-test splits tend to inflate performance metrics.

I strongly recommend explicitly assessing spatial generalisation capabilities. This could include:

- Spatial block cross-validation (training and testing in different regions, e.g., biogeographical subregions or k-means clusters in geographic space), and/or

- Feature-space cross-validation (e.g., K-nearest-neighbour splits in predictor space), which helps assess model robustness across underrepresented combinations of climate, stand structure, and species.

AA – In the first version of the manuscript, cross-validation was indeed performed through random sampling of individuals (80% for model training; 20% for model testing). In the revised version, we followed the reviewer's suggestion and implemented spatial block cross-validation. Specifically, we used the *group_vfold_cv* function of the *rsample* package in R, partitioning the dataset into five folds according to biogeographical subregions, and repeated the procedure 100 times to ensure robustness.

We explained it in the "Model evaluation" section of Data & methods (Page 11, lines 276-286):

"We applied repeated spatial k fold cross validation to evaluate model performance and predictive ability (Roberts et al., 2017). In this approach, the dataset is partitioned into k spatial blocks; models are fitted on k-1 blocks and validated on the remaining block, with the procedure repeated until each block has served once for validation. In our case, biogeographical subregions served as spatial blocks. Biogeographical subregions—86 classes in total—are defined by the French NFI as "sufficiently large geographical areas within which the combination of values taken by the factors determining forest production or the distribution of forest habitats is original" (Cavaignac, 2009; https://inventaire-forestier.ign.fr/spip.php?rubrique267). Because they differ systematically in key factors shaping forest characteristics (e.g., soil type, climate), this design minimises autocorrelation between training and testing subsets. We used five folds (i.e., 4/5 of the biogeographical subregions were used for training and 1/5 for testing, repeated five times), and repeated the entire procedure 100 times. In total, each model was evaluated across 500 iterations (5 folds × 100 repetitions). Implementation was carried out using the group_vfold_cv function of the rsample package (Frick et al., 2025)."

We have added a new table comparing the performance of each model under spatial block cross-validation, and the final models were subsequently built using the entire dataset. This addition clarifies our approach and explicitly addresses the issue of spatial autocorrelation in model evaluation.

2. Section 4.4 ("Why intraspecific variation…") underexplores climate effects

RC2 – The authors use mean annual climate variables, but intra-annual or interannual climate anomalies often show stronger relationships with wood formation. Given the dendrochronological nature of ring-width measurements used in the predictors, it would be interesting to explore:

- seasonal climatic anomalies,

- drought severity indices (e.g. SPEI),

- lagged climate effects, and/or

- extreme-year metrics.

Even if such variables are unavailable for the entire dataset, performing a sensitivity analysis on a subset (and reporting the result) would help clarify whether the weak climate signal is due to model formulation or intrinsic biological limits. This is especially important because the current discussion suggests ecological constraints, whereas part of the limitation may arise from the coarse temporal structure of the climate variables used.

AA – In the revised version, we tested the inclusion of additional environmental variables in our model (including the summer soil water deficit (June-August) and the carbon-to-nitrogen ratio). These additions only marginally increased the explained variance (R^2 rising from 0.14 to 0.17), which does not alter our conclusion that climate exerts a limited influence on wood density variation in mainland France.

As noted in our response to Reviewer 1, we initially tested additional variables—including soil pH, annual number of frost days, annual number of rainy days, and the De Martonne aridity index—but these introduced substantial collinearity without improving model performance. We have clarified this point in the revised text (Page 9, lines 228-230):

"Although additional variables—including soil pH, annual number of frost days, annual number of rainy days, and the De Martonne aridity index—were initially tested, they were excluded due to high collinearity and negligible gains in model performance."

We acknowledge that part of this limitation may stem from the temporal resolution of the climate variables used. Our study focuses on spatial variation in mean wood density, rather than interannual variability within trees. This approach differs fundamentally from dendroclimatological studies, which typically analyze wood density (or more particularly maximum latewood density) responses to seasonal or extreme-year anomalies in stressful environments (e.g., tree line or latitudinal limits). By construction, our framework mitigates the influence of short-term climatic fluctuations and emphasizes long-term spatial patterns.

We have clarified this distinction in the revised manuscript and added discussion to explain why intra-annual or interannual climate anomalies, while highly relevant in dendrochronological contexts, are less informative for our objective of mapping mean wood density across France (Page 24, lines 559-565):

"Beyond this spatial dimension, our study also differs from dendroclimatological research in its temporal focus. Dendroclimatology typically targets interannual variability through tree ring analyses, most often by analysing maximum latewood density (MXD) responses to seasonal or extreme year anomalies, and in climatically stressful settings such as tree line or high latitude environments (Briffa et al., 1998; Hughes et al., 1984). In contrast, our approach investigates spatial variation in wood density averaged over the full lifespan of trees. By construction, this framework integrates long-term climatic fluctuations and reduces the influence of short-term climatic fluctuation. It therefore emphasises persistent spatial patterns in wood density across broad biogeographical gradients rather than year-to-year climatic sensitivity."

3. Benchmarking against external biomass datasets (e.g. ESA-CCI biomass)

RC2 – The manuscript demonstrates how choices in wood-density modelling affect biomass at different scales. However, it would be extremely valuable to compare the resulting biomass maps (from either the NFI- or GIS-based model) with ESA-CCI Biomass products. This is particularly relevant because the GIS-based model shows species-dependent biases (Fig. 5) and ESA-CCI biomass is known to have systematic regional biases, especially in high-biomass areas. A comparison of spatial residuals, biome-stratified biases, or scatterplots at NFI plot scale could directly link the wood-density-driven biases identified here with well-documented remote-sensing limitations.

AA – Thank you for the suggestion. In this revised version, we have incorporated a comparison between our biomass estimates derived from the NFI-based model and the ESA-CCI product, and we discuss the outcomes of this comparison (Page 27, lines 653-665):

"To evaluate whether current remote sensing biomass estimates for mainland France are biased by the absence of explicit consideration of spatial variation in wood density, we compared forest biomass estimates derived from our "NFI based model" with the European Space Agency Biomass Climate Change Initiative map for 2020 (Santoro and Cartus, 2025). At the national scale, the comparison indicates good overall agreement (a 4.7% difference in total biomass across mainland France, corresponding to 113 million tons, or 54 million tons of carbon), whereas at finer spatial scales, such as biogeographical subregions, notable discrepancies emerged (see Supplementary Figure 7). Importantly, these differences do not follow the direction that would be expected if they were primarily driven by unaccounted wood density variation. For example, ESA biomass estimates are higher than NFI estimates in the Mediterranean region, despite this region being characterised by comparatively high wood density. If wood density were the dominant source of bias, neglecting this high wood density should have resulted in lower biomass values from the remote-sensing product. The fact that the opposite pattern is observed indicates that wood density effects alone cannot explain the discrepancies. This suggests that other factors—such as differences in allometric assumptions, forest structure representation, or remote-sensing signal interpretation—likely contribute to the divergence in biomass estimates and merit further investigation."

We included this section only in the discussion, as it does not constitute the core of the paper but rather a preliminary analysis that opens the way for more rigorous and detailed comparisons between NFI-based forest biomass estimates and remote sensing products.

4. GIS-based model: Could species-level information be used via the ForestPaths European tree genus map?

RC2 – Recent remote-sensing products, such as the ForestPaths European tree genus map (https://zenodo.org/records/13341104), provide spatially explicit genus-level classifications across Europe. Given that the genus explains ~30% of total variance (Section 3.1.1), integrating this map into the GIS-based model could improve predictive performance and reduce the regression-to-the-mean bias documented in Fig. 5. I encourage the authors to test whether adding ForestPaths genus information improves GIS-based predictions.

AA – In our study, the "GIS-based model" relied on the French *BD Forêt* map, which provides species-level forest classification (32 classes in total, with some species not distinguished). We acknowledge, however, that genus-level maps such as *ForestPaths* can be particularly useful in contexts where species-level maps like *BD Forêt* are not available. To evaluate the potential of the *ForestPaths* product, we developed a second "GIS-based model" using the *ForestPaths* European tree genus map in place of the French *BD Forêt* map. The results were very similar, suggesting that *ForestPaths* represents a promising first step toward incorporating species-specific variation in wood density into remote sensing biomass estimates. This addition has been included in the discussion (Pages 26-27, lines 639-645):

"Because the *BD Forêt* map is restricted to mainland France, we explored the potential for broader scale application by testing the *ForestPaths* European tree genus map (early access version), which provides genus level classification at 10 m resolution across Europe for the year 2020 (De Keersmaecker et al., 2025). In this test, we substituted the *ForestPaths* map for the *BD Forêt* map in the "GIS based model". The model yielded very similar performance (R^2 = 0.29 at the tree level and R^2 = 0.44 at the NFI plot level with *ForestPaths*, compared to 0.31 and 0.48 with *BD Forêt*), indicating that products such as the *ForestPaths* genus map may be highly valuable for capturing spatial variation in wood density at both fine resolution and broad spatial scales."

Again, we included this section only in the discussion, as it does not constitute the core of the paper but rather a preliminary analysis that opens interesting perspectives.

Concluding remarks

RC2 – This is an excellent, well-written manuscript. The dataset and modelling framework represent a significant contribution to European forest research. The key findings are robust:

- Species identity dominates wood-density variation.

- Intraspecific variation exists but is structurally difficult to predict.

- The effect of wood density on biomass is negligible at the national scale but substantial at the subregional and local scales.

- GIS-based predictors recover a significant portion of variation even without species identity.

- Addressing the points above, particularly the treatment of cross-validation, the more in-depth analysis of intraspecific drivers, and benchmarking against external datasets, will significantly enhance the robustness and impact of the study.

AA – We sincerely thank the reviewer again for this very positive assessment and encouraging feedback. We agree that the points raised are important, and we have incorporated the suggested improvements regarding spatial cross-validation, the consideration of additional environmental drivers, and benchmarking against external datasets. We believe these additions have strengthened the robustness and impact of the study, and we hope that the revised version will meet the reviewer’s expectations.