

Response letter to Reviewer 2

Manuscript: *The global forest diameter spectrum using a machine learning approach*

Manuscript ID: essd-2026-226

We sincerely thank the reviewer for the careful assessment of our manuscript and for the encouraging comments regarding the potential value of the resulting global forest diameter datasets. We particularly appreciate the thoughtful suggestions concerning geographic stratification, predictor selection, variable importance, and computational efficiency.

We agree that these considerations are important for understanding the robustness, generalizability, and interpretability of the modeling framework. Although the primary objective of this data paper is to provide spatially explicit forest diameter products, we recognize that additional sensitivity analyses can offer valuable insight into the methodological choices underlying these products.

In response to the reviewer's suggestions, we conducted complementary sensitivity analyses for all three forest diameter metrics: arithmetic mean diameter (D_{mean}), quadratic mean diameter (D_{qm}), and the coefficient of variation of diameter (D_{cv}). These analyses evaluated alternative geographic stratification strategies, the effect of reducing predictor-set size, ecozone-specific predictor importance, and the relationship between predictive accuracy and computational cost.

For these complementary analyses, we used random-forest models with 250 trees and repeated training–testing partitions, with approximately 90% of observations used for training and 10% for testing in sufficiently large groups. Ecozone–continent groups containing fewer than five usable observations were excluded. The sensitivity analyses were conducted separately from the original production-model workflow.

Our point-by-point responses (in orange -color text) are provided below.

Comment 1. Training models by ecozone versus ecozone × continent

Reviewer comment: My understanding is that the machine learning models are trained separately for each ecozone. These ecozones can span multiple continents. For example, boreal forests occur across both North America and Eurasia. However, forest composition can differ substantially among continents within the same ecozone. For example, spruce may be more dominant in parts of North America, whereas pine and larch may be more important in Eurasia. These differences in species composition could translate into substantial differences in forest structural characteristics, including diameter distributions, as well as differences in the environmental variables that best predict them.

The same issue may potentially apply to temperate and tropical forests. I therefore wonder whether training separate models for each ecozone–continent combination could improve predictive performance compared with the current approach. Is this something the authors considered? If so, could the authors briefly discuss what they found? If not, do the authors think that this would be a potentially useful sensitivity analysis?

Response:

We appreciate the reviewer’s suggestion, which has helped us examine this aspect of our methodology more carefully.

We agree that forests within the same ecological zone may differ considerably across continents in terms of species composition, environmental conditions, and structural characteristics. We therefore conducted an additional sensitivity analysis to explore whether training random-forest (RF) models separately for each ecozone–continent combination could improve predictive performance compared with our original ecozone-based approach. We have now incorporated the methodological details and key findings of this sensitivity analysis into Sections 2.5 and 3.1 of the revised manuscript, respectively. The detailed comparison of ecozone-based and ecozone–continent models are provided in Supplementary Table S1.

We evaluated both modeling strategies for mean diameter (D_{mean}), quadratic mean diameter (D_{qm}), and diameter coefficient of variation (D_{cv}). Given the large global inventory dataset and the computational demands of fitting multiple geographically stratified models, we used 250 trees per RF model. We performed repeated training–testing iterations, evaluating both

approaches on identical held-out observations with matched predictor sets. Combinations with fewer than five complete-case observations were excluded. Model performance was assessed using R^2 , RMSE, and MAE, with results summarized across four broad ecological-zone groups (*please refer to Table S1 below*).

Our results suggest that additional continent-level stratification does not consistently improve predictive performance. For D_{mean} , ecozone–continent models showed improved performance in Boreal forests, with RMSE decreasing from 4.45 to 3.25 cm. However, the original ecozone-based models performed better in Temperate (3.91 versus 5.80 cm), Subtropical (4.63 versus 4.95 cm), and Tropical forests (4.11 versus 4.67 cm). For D_{qm} , the original ecozone-based models achieved lower RMSE across all four broad ecological-zone groups: Boreal (4.94 versus 5.51 cm), Temperate (4.38 versus 6.49 cm), Subtropical (5.27 versus 5.55 cm), and Tropical forests (5.09 versus 5.58 cm). For D_{cv} , the ecozone-based models showed slightly better predictive performance in Boreal (RMSE: 0.10 versus 0.11), Temperate (0.09 versus 0.10), and Subtropical forests (0.11 versus 0.12), while both modeling strategies showed comparable performance in Tropical forests (RMSE: 0.13).

Overall, ecozone-based models achieved lower RMSE in ten of the twelve response–ecological-zone-group comparisons, whereas ecozone–continent models performed better in only one comparison. We therefore believe that these findings provide reasonable support for retaining our original ecozone-based modeling framework, as further geographic stratification did not consistently improve predictive accuracy. While continent-specific models may better capture regional differences in forest composition and structure, dividing observations into smaller geographic subsets may also reduce the training information available to individual models. This trade-off could help explain the limited improvements observed with finer geographic stratification.

We also acknowledge that this sensitivity analysis was conducted using 250 trees and a complete-case subset of the available observations to maintain computational feasibility. Therefore, we view these findings as an additional assessment of the robustness of our modeling strategy rather than a definitive comparison of all possible geographic modeling configurations. Based on this evidence, we believe that retaining the ecozone-based approach is a reasonable choice for the present study, particularly given its generally comparable or slightly better

predictive performance and its ability to draw upon observations across continents. We have included the results of this sensitivity analysis in *Table S1* below to provide further transparency regarding our modeling decisions.

Table S1. Predictive performance of ecozone-based versus ecozone–continent random-forest models

Response	Ecological-zone group	Modeling strategy	R ²	RMSE	MAE
D _{mean}	Boreal	Ecozone	0.82	4.45	3.07
		Ecozone × continent	0.83	3.25	2.25
D _{mean}	Temperate	Ecozone	0.52	3.91	2.73
		Ecozone × continent	0.45	5.80	3.59
D _{mean}	Subtropical	Ecozone	0.41	4.63	3.26
		Ecozone × continent	0.35	4.95	3.46
D _{mean}	Tropical	Ecozone	0.81	4.11	2.57
		Ecozone × continent	0.76	4.67	2.85
D _{qm}	Boreal	Ecozone	0.83	4.94	3.40
		Ecozone × continent	0.44	5.51	3.42
D _{qm}	Temperate	Ecozone	0.54	4.38	3.06
		Ecozone × continent	0.48	6.49	4.05
D _{qm}	Subtropical	Ecozone	0.43	5.27	3.72
		Ecozone × continent	0.38	5.55	3.91
D _{qm}	Tropical	Ecozone	0.78	5.09	3.18
		Ecozone × continent	0.73	5.58	3.45
D _{cv}	Boreal	Ecozone	0.62	0.10	0.07
		Ecozone × continent	0.42	0.11	0.07

D _{cv}	Temperate	Ecozone	0.47	0.09	0.06
		Ecozone × continent	0.46	0.10	0.07
D _{cv}	Subtropical	Ecozone	0.52	0.11	0.07
		Ecozone × continent	0.51	0.12	0.07
D _{cv}	Tropical	Ecozone	0.47	0.13	0.10
		Ecozone × continent	0.46	0.13	0.10

Comment 2. Number and importance of predictor variables

Reviewer comment: I understand that, for an ESSD paper, the primary emphasis is on producing spatially explicit datasets rather than on explaining the mechanisms underlying spatial patterns. Nevertheless, I wonder whether the authors could provide some additional information on the number and importance of the predictor variables.

In particular, how does model performance change as the number of predictors is reduced? For example, do the best 10 or 20 predictors achieve performance close to that obtained using the full set of 50+ predictors? Conversely, is there a point at which adding additional predictors provides little further improvement in predictive performance? It would also be interesting to know whether this relationship varies spatially or among ecozones.

Relatedly, which predictor variables are most influential, and how does their importance vary spatially? I appreciate that correlated predictors can make variable-importance measures difficult to interpret, but even a broad assessment of predictor importance could provide useful context for interpreting the resulting products, particularly when considered alongside the uncertainty layers.

Finally, retaining only the most influential predictors could potentially reduce computational costs and make the models easier to interpret. Could the authors comment on how they considered the trade-off between model performance, computational complexity, and interpretability when selecting the predictor set?

Response:

We thank the reviewer for the valuable suggestions.

- (a) The reviewer asks how predictive performance changes when the number of predictors is reduced, whether a smaller subset of influential predictors can achieve performance comparable to the complete predictor set, and whether these patterns vary among ecozones.

We agree that examining the sensitivity of model performance to predictor-set size is important for understanding whether the complete predictor set provides meaningful additional predictive information or whether similar accuracy can be achieved using a smaller number of variables.

To address this question, we conducted an additional predictor-reduction analysis for D_{mean} , D_{qm} , and D_{cv} across forest ecozones. Within each training partition, we ranked predictors using random-forest permutation importance and compared models containing the top 10, 20, and 30 predictors with models containing the complete set of available predictors. All predictor-set comparisons within a given ecozone and iteration were evaluated using the same held-out observations. The analysis was repeated across training–testing partitions.

We have also incorporated the predictor-reduction analysis and its principal findings into Sections 2.5 and 3.1 of the revised manuscript, with the corresponding results presented in Supplementary Table S2.

The results (*Table S2 below*) indicate that predictive accuracy generally improved as the number of retained predictors increased, although the magnitude of improvement became relatively small for predictor sets containing approximately 20–30 variables.

- i. For D_{mean} , the mean increase in RMSE relative to the complete predictor set was 0.128 cm using the top 10 predictors, 0.058 cm using the top 20 predictors, and 0.012 cm using the top 30 predictors.

- ii. For D_{qm} , the corresponding mean RMSE differences were 0.168, 0.037, and -0.030 cm, respectively. The slightly negative value for the top 30 predictors indicates that this reduced set achieved marginally lower average RMSE than the complete predictor set across the evaluated ecozone–iteration comparisons.
- iii. For D_{cv} , the mean RMSE differences were 0.0020, 0.0007, and 0.0005 for the top 10, 20, and 30 predictors, respectively.

Table S2. Mean differences in held-out RMSE between reduced predictor sets and the complete predictor set.

Response	Top 10	Top 20	Top 30
D_{mean}	+0.1281	+0.0578	+0.0120
D_{qm}	+0.1678	+0.0366	-0.0300
D_{cv}	+0.0020	+0.0007	+0.0005

Note: Values represent the mean of paired RMSE differences across ecozone–iteration comparisons per response and reduced predictor set. Positive values indicate higher RMSE relative to the complete predictor set; negative values indicate lower RMSE. D_{mean} and D_{qm} values are in cm, while D_{cv} values are dimensionless.

The results suggest that much of the predictive information contained in the complete predictor set can often be retained using approximately 20–30 influential variables. However, the relationship between predictor-set size and predictive accuracy varied among ecozones and among the three diameter metrics. We therefore do not interpret these findings as evidence that a single reduced predictor set would necessarily be optimal across all forest regions. Rather, they indicate that the marginal predictive benefit of additional variables often becomes relatively small once the most influential predictors have been included. This analysis provides additional context for evaluating predictor redundancy and the trade-offs associated with retaining a broader set of environmental and forest-related variables.

- (b) The reviewer asks for additional information on the most influential predictor variables and how their importance varies geographically, while acknowledging that correlations among predictors can complicate interpretation.

Response:

We appreciate this suggestion and agree that an assessment of predictor importance can provide useful context for interpreting the resulting forest diameter products, even though identifying the ecological mechanisms underlying global diameter patterns is not the primary objective of this data paper.

To address this point, we calculated permutation-based variable importance separately for each diameter metric and ecozone, using the random-forest models fitted to the training data. The analysis revealed both commonalities and differences in predictor importance among diameter metrics and forest ecozones.

- i. For D_{mean} and D_{qm} , climatic variables were frequently represented among the most influential predictors. These included precipitation seasonality, temperature seasonality, precipitation of the wettest month, precipitation of the wettest quarter, and other measures of seasonal climatic conditions and water availability. For example, precipitation seasonality appeared among the ten highest-ranked predictors in 10 of the 16 ecozones for D_{mean} and 11 of the 16 ecozones for D_{qm} .
- ii. For D_{cv} , the importance patterns differed somewhat. Alongside climatic predictors, forest diversity metrics were frequently represented among the highest-ranked variables. In particular, tree species richness appeared among the top ten in nine ecozones.

These findings suggest that the relative predictive contributions of climatic conditions and forest diversity characteristics differ among diameter metrics. Climatic predictors appear particularly prominent for the two measures of average tree diameter, whereas forest diversity metrics are frequently influential for variation in diameter structure. At the same time, predictor rankings varied among ecozones, suggesting geographic

differences in the relative predictive importance of these variables within the fitted models.

We agree with the reviewer that correlated predictors complicate the interpretation of permutation importance. In particular, environmental variables representing related climatic gradients may share predictive information, and their individual importance values may therefore depend on the presence of other correlated predictors. Accordingly, we interpret these results as indicators of relative predictive contribution within the fitted models, rather than as evidence of direct ecological causation or independent mechanistic effects.

We believe that this additional analysis provides useful descriptive context for the global diameter products and helps clarify how the predictive information used by the models varies among forest regions.

- (c) The reviewer asks whether retaining fewer influential predictors could reduce computational costs and improve interpretability, and how the authors considered these trade-offs when selecting the predictor set.

Response:

We thank the reviewer for emphasizing this practical consideration. We agree that predictor selection should balance predictive performance, computational efficiency, and interpretability, particularly in global modeling applications involving large spatial datasets.

We have also briefly discussed the computational benefits of predictor reduction and the associated trade-offs in the revised manuscript. The detailed model-fitting time comparisons are provided in Table S3 of this response letter.

To evaluate the computational implications of predictor reduction, we recorded random-forest model-fitting times for the complete and reduced predictor sets under the same sensitivity-analysis settings. The results (as shown in Table S3 below) indicate that reducing the number of predictors can substantially decrease model-fitting time. Relative

to the complete predictor set, retaining the top 10 predictors reduced average fitting time by approximately 55.7%, 54.2%, and 51.8% for D_{mean} , D_{qm} , and D_{cv} , respectively. Using the top 20 predictors reduced average fitting time by approximately 40.8%, 40.6%, and 40.7%, respectively. Retaining the top 30 predictors produced smaller but still meaningful reductions of approximately 32.8%, 25.5%, and 23.1%.

Table S3. Mean reduction in model-fitting time relative to the complete predictor set.

Predictor set	D_{mean}	D_{qm}	D_{cv}
Top 10	55.7%	54.2%	51.8%
Top 20	40.8%	40.6%	40.7%
Top 30	32.8%	25.5%	23.1%

Note: Percentages represent average paired reductions in random-forest fitting time. They do not include the additional time required for initial predictor ranking, spatial prediction, or uncertainty estimation.

These findings demonstrate that reduced predictor sets can offer substantial computational benefits while often maintaining predictive performance close to that obtained using the complete predictor set.

From an interpretability perspective, models containing fewer variables can also facilitate a more concise description of the environmental and forest characteristics contributing to predictions. Nevertheless, a reduced predictor set may omit information that is useful in particular ecozones, even when the average change in predictive accuracy is small.

Our original modeling strategy prioritized predictive performance and broad environmental representation across geographically diverse forest systems. The additional sensitivity analyses indicate that reduced predictor sets may provide an attractive alternative for computationally constrained applications, but we believe that they do not

demonstrate that one universally reduced set would be preferable across all ecozones and diameter metrics. We therefore consider the complete predictor framework a defensible choice for the original global mapping application, while recognizing the potential value of more parsimonious models for future applications and model refinement.

Overall, the additional analyses have allowed us to further evaluate the robustness of our modeling framework and clarify the trade-offs associated with geographic stratification and predictor selection. We have incorporated the relevant methodological details, key findings, and their implications into the revised manuscript, with supporting results provided in the Supplementary Information where appropriate. We sincerely appreciate the reviewer's constructive suggestions, which have helped us strengthen the transparency and interpretation of the global forest diameter products.

Sincerely,

Ankita Mitra and Jingjing Liang