

Response to Reviewers and Editor

Dear Editor, dear Reviewers,

We thank you for evaluating our manuscript and the constructive comments and suggestions. Based on your feedback we will clarify our methodological approach and refine our discussion, which we believe will result in a more robust version of the manuscript.

Please find our detailed, point-by-point responses to each of the reviewers' and editor's comments below.

On behalf of all co-authors,
Guohua Liu

Reviewer and Editor Comments are in black. Author Responses are in blue.

Response to Reviewer #1

The authors present a new global forest above-ground biomass dataset spanning 1993–2020 at 8.9 km resolution, produced by training a random forest model on ESA-CCI biomass maps using radar backscatter and climate covariates as predictors. Compared to existing long-term products, which are VOD-based at ~25 km, the dataset offers a finer spatial resolution while retaining a near-30-year continuous annual record. This is achieved by exploiting the recently published radar backscatter record of Tao et al. (2023). The product is potentially a useful community resource, and the manuscript includes comparisons against other biomass products and field inventories. However, two important issues should be addressed before publication, which I outline below.

1. The product's central contribution is temporal — interannual to decadal biomass dynamics — but the model structure raises a concern about where that temporal signal originates. Two of the three predictors (MAT_bg, MAP_bg) are static background climate and cannot, by construction, generate any interannual variation; the authors further show that adding dynamic climate does not improve performance (Table 1). The temporal dynamics in the product therefore rest entirely on the radar backscatter input. Radar backscatter is, by its nature, sensitive to soil moisture and vegetation water content as well as to biomass; the authors' wet-season index reduces but does not remove this, and even the selected index retains a non-trivial correlation with soil moisture (Figure S3). Because the product's interannual signal comes only from radar, any residual moisture variability will project directly onto the reported biomass trends. Combined with the low standalone radar skill

(test $R^2 = 0.136$), this raises the concern that part of the temporal signal may reflect moisture dynamics rather than biomass change. To be clear, I am not claiming radar is uninformative — with correlated predictors the variance explained is shared between radar and climate, and the large jump to $R^2 = 0.833$ when static climate is added may reflect climate carrying much of the spatial structure. The concern is specific to the temporal signal that distinguishes this dataset. The authors should (i) report variable importance for the selected model, and (ii) demonstrate that the interannual and decadal trends are driven by genuine radar-sensed biomass change rather than residual moisture variability — for example by relating the temporal anomalies back to the radar signal and to soil moisture directly.

Author Response (AR): We fully agree with the reviewer that the interannual to decadal dynamics in our aboveground biomass (AGB) product are primarily driven by the radar backscatter input, and the static background climate variables generate spatial variation in the model. This is, in fact, by design, since we aim to reconstruct biomass variations based on radar backscatter signal, and not convolute the signal of the predicted biomass with that of climatic variables (which would limit the applicability of the dataset to analyze vegetation responses to climate trends and extremes). The static background variables are not used for temporal predictions directly, but rather to provide context information for the RF model to be able to distinguish different relationships between biomass and the radar backscatter for different climate zones, as a proxy for biomes. We show that by providing this context information, the model is able to better predict biomass across the globe, without overfitting.

We acknowledge that radar backscatter is sensitive to both vegetation structure and soil moisture, and that residual moisture effects may influence the temporal signal of radar backscatter. In order to mitigate this, we carefully evaluated multiple radar backscatter indices and selected the one that minimizes sensitivity to soil moisture (Figure S3). To further address the reviewer's concern and respond to the suggestions, we will include the following two analyses:

1) Variable Importance via Permutation Importance (Fig. R1):

We computed permutation importance for both the Rad-Clim_bg and Rad-Clim_bg-dy models. The results show that radar backscatter is the most important predictor in both models, outperforming background climate and dynamic climate variables (Fig. R1). This confirms that the temporal dynamics in the AGB estimates are predominantly captured through the radar signal, and that background climate variables contribute to the spatial structure. This analysis supports the important role of these variables as context information, also reflected in the large R^2 increase when they are added (Table 1), and the lower information value of the dynamic variables (Fig. R1 right).

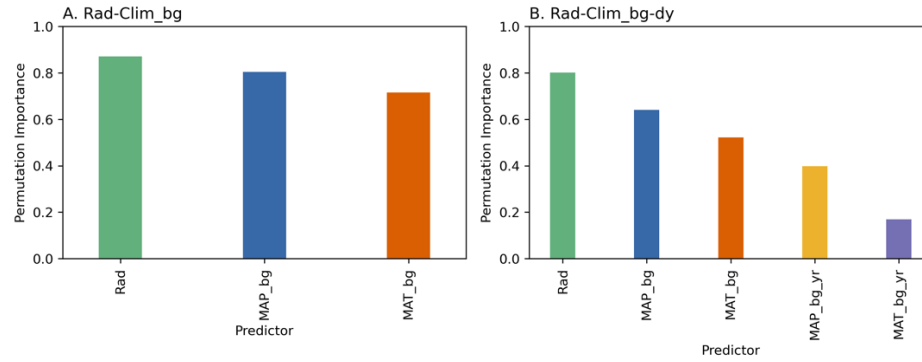


Fig. R1 Permutation importance of predictors for Rad-Clim_bg model (A) and Rad-Clim_bg-dy model (B).

2) Partial Correlation Analysis (Fig. R2):

To assess whether the observed temporal AGB changes are confounded by soil moisture, we conducted a partial correlation analysis that isolates the relationship between AGB and radar backscatter while controlling for soil moisture variability. The partial correlation between AGB and radar signal remains strong, with a median correlation coefficient of 0.54 across the study region (Fig. R2). In contrast, the partial correlation between AGB and soil moisture is negligible, with a median coefficient of 0.0073. This divergence indicates that the interannual variability in AGB is predominantly aligned with the radar signal rather than soil moisture fluctuations. These results support the interpretation that the temporal dynamics in our product reflect changes in biomass, not residual moisture effects. We also compared the spatial patterns of long-term trends. We calculated a trend for each grid cell in AGB, radar backscatter, and soil moisture, then calculated partial correlations across grid cells. The partial correlation between AGB and radar-backscatter trends, controlling for soil-moisture trends, was 0.122; the partial correlation between AGB and soil-moisture trends, controlling for radar-backscatter trends, was 0.062. Thus, the spatial pattern of AGB trends was more closely associated with the radar-trend pattern than with the soil-moisture-trend pattern.

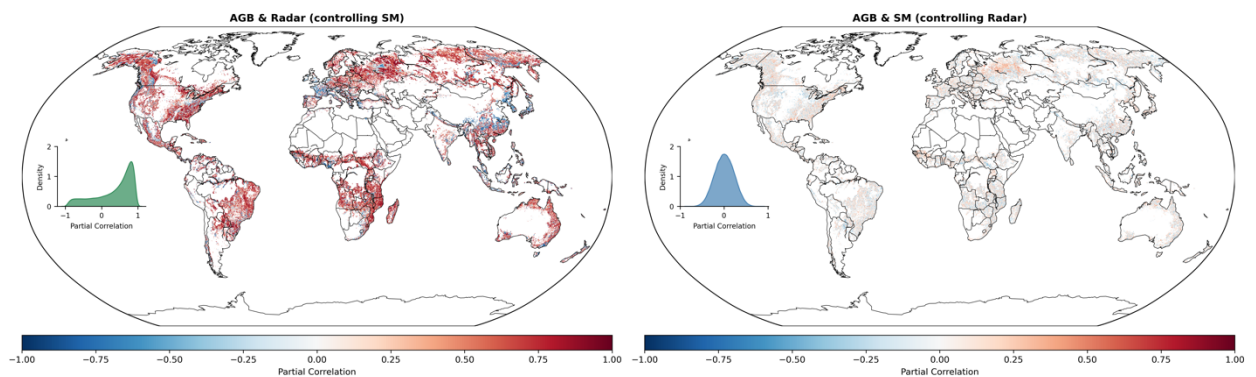


Fig. R2 Spatial distribution of partial correlations between above-ground biomass (AGB), radar backscatter (Radar) and soil moisture (SM). The probability density distribution of these correlations is displayed.

We will update the manuscript accordingly to include these analyses in the results and discussion sections.

2. The reported net sink of 0.44 PgC yr^{-1} is a small residual between much larger, oppositely-signed gross gain ($\sim 0.84 \text{ PgC yr}^{-1}$) and gross loss (~ -0.39 to $-0.74 \text{ PgC yr}^{-1}$) terms, which makes it sensitive to systematic biases in the product. The quoted uncertainty of $\pm 2.5 \text{ PgC}$ (Eq. 1) captures only the spread across cross-validation iterations and does not propagate other potentially important sources — calibration bias (the slope of 0.36 against inventories), offsets at the ERS/ASCAT/QSCAT sensor transitions, and the fixed conversion factors (0.49 carbon fraction; root-to-shoot map) applied globally. It would strengthen the paper if the authors acknowledged more explicitly that the reported uncertainty is a lower bound reflecting model variance only, discussed these additional sources qualitatively even if they cannot be fully quantified, and tempered the net-flux claims accordingly. The 1997 artefact (Figure 3B) is a useful illustration here: it shows that a sensor transition can produce a spurious global biomass swing, so a brief comment on whether the other transitions were checked for similar (smaller) offsets would help reassure readers about the homogeneity of the trend record.

AR: Thank you for the suggestion. We acknowledge that the current quantified uncertainty estimation for this AGB product reflects only the model uncertainty, and it does not account for other sources. To estimate the potential bias from input training ESA CCI biomass data, we are currently conducting an analysis to quantify how the uncertainty reported in the ESA CCI biomass dataset propagates into the AGB product. Furthermore, we will perform a breakpoint test on the AGB time series to assess potential influence from radar sensors (ERS/ASCAT/QSCAT). This analysis will help evaluate whether sensor shifts can introduce spurious trends. We will further expand the uncertainty section to discuss the potential contributions from multiple sources, including input ESA CCI biomass data bias, radar backscatter data (sensor transitions, shortwave band, potential influence from soil moisture), fixed conversion factors, and data processing methods. We will add a quality flag along the dataset to indicate regions and time periods with higher uncertainty, improving transparency and usability.

Minors comments

C-band saturation in dense forest acknowledged (l. 338–341) but underplayed given tropics = $\sim 66\%$ of stock.

AR: We agree that C-band radar saturation in dense tropical forests is a critical limitation, particularly given that tropical regions account for large of the global aboveground biomass carbon stock. We have conducted a detailed analysis of the model performance across biomes and biomass range, and explicitly flag areas where results should be interpreted with caution. These results have been added in detail in the revised Results and Discussion sections.

Test R^2 inconsistent: 0.833 (Table 1) vs 0.883 (l. 213).

AR: corrected

Response to Reviewer #2

This paper uses Random Forest machine learning trained on CCI-Biomass AGB data with spaceborne K- and C-band scatterometer data and mean climate variables to produce a dataset of annual biomass maps running from 1993 to 2020 at a mapping scale of 8.9 km. This is used to estimate above ground carbon and total vegetation carbon globally and within biomes, and to examine the trends present in the dataset. It is potentially valuable because it is the longest such dataset currently available, and it is unlikely that any longer comparable datasets will be produced as satellite data to do so are not available. Weaknesses of the paper include: (1) The fact that it is just regression-based and no physical reasoning is given as to why such short wavelength radar is able to provide meaningful AGB estimates when it scatters mainly in the upper canopy; (2) The training data from CCI-Biomass are known to contain biases that will presumably transfer into this dataset. This needs a fuller discussion, including indication that the authors know something about these biases; (3) The reference inventory data indicate significant biases in the dataset. The implications of these biases for their estimates of stocks and stock change need to be clarified.

It is generally well written and clear, though with some small English errors, and there is systematic misuse of words like “dynamic” (see below). The figures and tables are generally useful and clear.

Detailed comments follow

The abstract asserts that the estimate of global AGC aligns with other estimates, but it is 20% less than several. It also asserts that this dataset is a benchmark, which implies it can be taken as a valid reference. However, this is not supported by Fig. 1 which shows significant and systematic biases compared to inventory data and by the fact that the training data (CCI-Biomass) is known to exhibit biases, particularly for high and low AGB. It also says the dataset is temporally continuous; it is not, it is annual and this should be stated.

AR: We thank the reviewer for these careful points and have revised the abstract accordingly. We agree the claim that our estimates “align with” other estimates was not fully correct, so that we have removed this sentence. We also agree that “benchmark” implied a validated reference, and we have removed the term “benchmark”, now described simple as product.

We have corrected "temporally continuous" to "annually resolved".

Specifically, we replaced:

"This long-term, temporally continuous, and moderate-resolution dataset provides a new benchmark..."

With:

"This long-term, wall-to-wall, annually resolved, and moderate-resolution dataset provides a new globally consistent product ..."

l. 47-48 The sentence makes no English sense; in addition, and it talks about 10 years when the data period quoted is 16 years.

AR: We have revised the sentence (lines 47–48) for improved clarity, grammar, and accuracy. We also clarified that the “~10 years” refers to studies based on VOD time series, whereas the longer period (e.g., 1992–2018) refers to studies based on radar satellite observations. The sentence has been rewritten to clearly associate each time span with the corresponding data source and references.

l. 58-59 this sentence has something missing or is miswritten.

AR: This sentence has been rewritten for grammatical accuracy and clarity. The corrected version now reads:

"These processes result in a slow accumulation of AGB and a rapid loss following disturbance, along with persistent legacy effects and slow recover."

l. 64 (& 361) What is meant by space for time substitution?

AR: Space-for-time substitution is a method that uses spatial differences across landscapes to infer temporal changes over time. We have added a brief explanation of the space-for-time substitution method in the manuscript.

"Space-for-time substitution is a strategy used to predict ecological responses to changing climate when long-term observations are lacking; prediction is instead based on patterns observed across spatial climatic gradients (Evans et al., 2025)."

l. 80. Where does the 8.9 km resolution come from? L. 358 says that the resolution of the scatterometer data is ~25 km so how is the gain in resolution achieved? Or is it not resolution but instead grid spacing (which implies high correlation between grid cells).

AR: The 8.9 km resolution refers to the spatial resolution of the satellite radar backscatter data produced by Tao et al. (2023), which is the primary dataset used in our study. To clarify the data hierarchy and improve logical flow, we have restructured the subsections as follows: **2.1.1 Satellite radar backscatter data**, and **2.1.2 Land cover data**. This reorganization ensures that the spatial resolution is introduced in the context of the primary remote sensing data, and the rationale for using 8.9 km is now clearly linked to the satellite radar dataset.

We have clarified that the approximately 25 km resolution mentioned in line 358 refers to products from other studies, such as Besnard et al. (2021).

84. It says 7 biomes but only lists 6.

AR: Corrected.

91. Many of 100 m pixels in the CCI dataset in the tropics have $CV > 80\%$, so this exclusion is likely to distort the tropical estimates.

AR: To assess whether excluding CCI pixels with a coefficient of variation (CV) greater than 80% could disproportionately affect tropical estimates, we calculated the CV of the CCI aboveground biomass data for 2017–2020, the period used for model training and testing. Figure R3 shows the global distribution of grid cells above and below the 80% threshold. Grids with $CV \geq 80\%$ account for approximately 3.4% of the tropical area. This indicates that the exclusion affects a relatively small share of the tropical area. We have clarified the exclusion procedure and added this analysis to the revised manuscript.

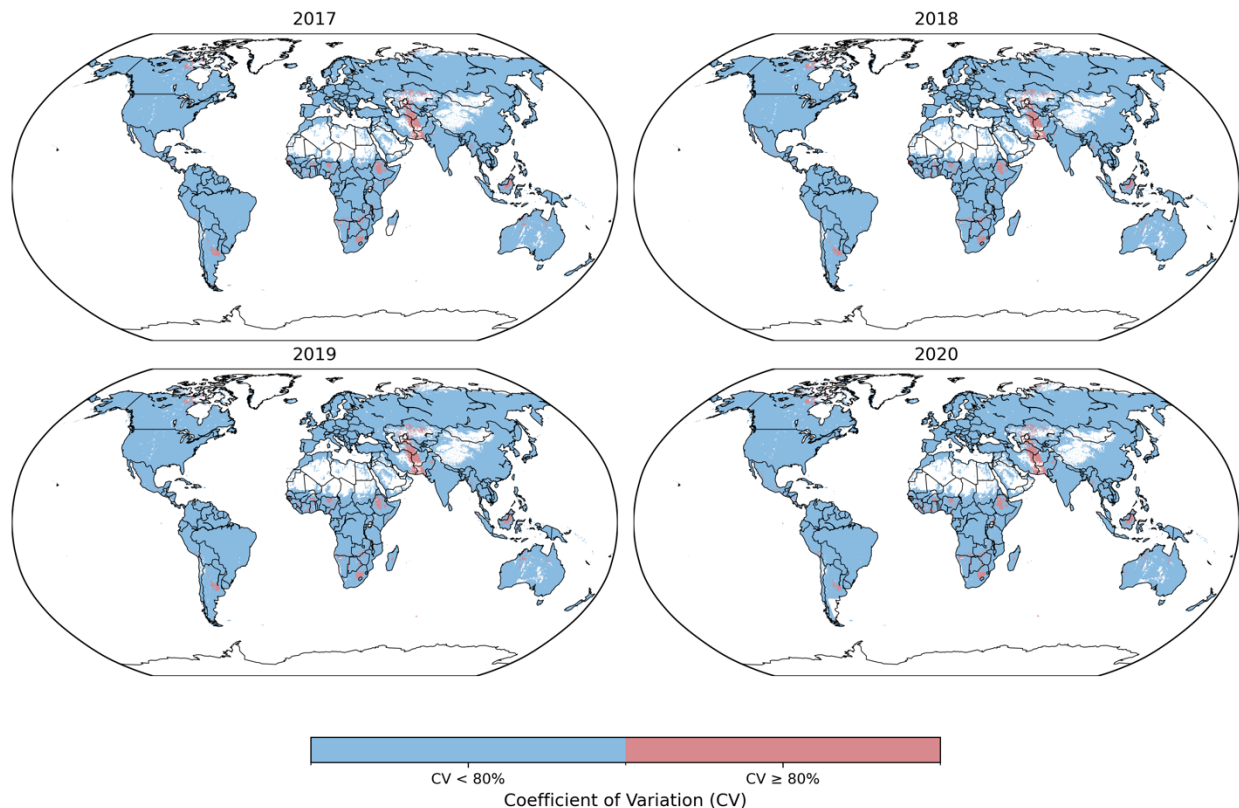


Fig. R3 Spatial distribution of the coefficient of variation (CV) in 100 m ESA CCI aboveground biomass data for 2017–2020. Grid cells with $CV < 80\%$ are shown in blue; those with $CV \geq 80\%$ are shown in red.

96. What is the meaning of biomass_ag_OM, AGB_ini, AGB_fin, Initial AGB? How can one aggregate inventory data to 8.9 km? And how is “sufficient” defined for sampling density.

AR: We thank the reviewer for these questions regarding variable definitions and methodology. We have revised the manuscript to clarify these points as follows:

Variable definitions: The variables biomass_ag_OM, AGB_ini, AGB_fin, and Initial AGB represent aboveground biomass metrics are derived directly from the original inventory datasets. We retained the original variable names to ensure consistency with the source data. To avoid confusion, we have added a new table (Table S2) in the Supplementary Materials that explicitly defines all variables used from the inventory datasets.

Table S2. The above-ground biomass variables and their meanings in each inventory dataset

Source	Variable	Meaning
Anderson-Teixeira et al. (2018)	biomass_ag_OM	Total aboveground live biomass
Brienen et al. (2015)	AGB_ini	Standing above ground biomass at the start of the census interval
Brienen et al. (2015)	AGB_fin	Standing above ground biomass at the end of the census interval
Qie et al. (2017)	Initial AGB	Plot above ground biomass at the initial census
Schepaschenko et al. (2019)	Initial AGB	Plot above ground biomass at the initial census

Aggregation to 8.9 km: To compare field measurements with satellite-derived AGB (which represents an integrated signal over a larger area), we employed a **plot-to-map** aggregation method. Specifically, we averaged the AGB values from all inventory plots located within the boundaries of an 8.9 km pixel. **This approach provides a more reliable spatial representation for comparison than a direct point-to-pixel comparison, which would be susceptible to scale mismatch and spatial heterogeneity.** We have clarified this point accordingly in the manuscript.

Sampling density: We acknowledge that the number of inventory plots within a pixel influences how well the sample represents the true population mean. Given the limited availability of public forest inventory data at the global scale, we established a threshold of 10 plots per pixel as the minimum requirement for a pixel to be considered "sufficiently sampled" in our analysis. This threshold aligns with common practices in large-scale biomass studies where data availability is often a constraint (Santoro et al., 2022). We have added discussion to explicitly acknowledge the uncertainty associated with this plot-to-map comparison and to advise caution when interpreting results from pixels with lower sampling densities.

108. Longer wavelength radar backscatter (L-band and longer wavelengths) is considered useful for estimating AGB, up to a point, due to penetration to the woody parts of the forests, but C- and K band are not since they scatter in the upper canopy. So what is the rationale for estimating AGB from these shorter wavelengths as is done in this paper?

AR: We thank the reviewer for raising this important point regarding the interaction between radar wavelength and forest structure. We fully agree that L-band and longer wavelengths are generally

superior for estimating absolute AGB values due to their greater penetration into the canopy and woody components, whereas C-band signals are more sensitive to the upper canopy (leaves, branches).

However, our rationale for using C-band data in this study is based on the following considerations:

Correlation with AGB through canopy structure: Although C-band does not penetrate as deeply as L-band, its backscatter signal is strongly correlated with canopy structural properties that are closely related to AGB. This makes C-band a viable proxy for biomass, particularly for tracking changes over time.

Long-term temporal coverage: Our study aims to provide a long-term consistent, global radar time series for ecological and climate applications. C-band radar backscatter data have provided near-continuous global observations since 1992, making them the only radar source with the necessary temporal depth. In contrast, L-band records (e.g., ALOS-2) are either too short, discontinuous, or lack the global coverage required for a global analysis. Therefore, C-band radar data could be used to produce acceptable AGB estimates to compensate for the unavailability of longer wavelength data.

We have revised the manuscript to clarify this.

110 & 132 Soil moisture, not surface water, is the more problematic confusing effect. There is little direct evidence relating vegetation water to backscatter at this scale. Also important is that the large scale soil moisture products are based on the same scatterometer data that are here being used to estimate AGC. This is likely to be the reason behind the correlations shown in Fig. S3.

AR: We agree that soil moisture can affect the radar backscatter signal. We tested multiple radar backscatter indices and selected the wet-season amplitude index to minimize sensitivity to soil moisture. This radar index is less influenced by soil moisture fluctuations than raw backscatter. We conducted a partial correlation analysis to isolate the relationship between aboveground biomass (AGB) and radar backscatter while controlling for soil moisture (Fig. R2). The results show that the partial correlation between AGB and radar signal remains strong (median $r = 0.54$), whereas the partial correlation with soil moisture is negligible (median $r = 0.007$). This indicates that the radar signal can detect biomass changes, not by residual soil moisture effects. Also, see reply to Reviewer 1 comment 1.

111. Generally agreed? By whom? More importantly, why should this be true? And what effect is being referred to in l. 112?

AR: We have rewritten these sentences.

135 Here we start to talk about pixels instead of grid cells – are these the same? What is the population being talked and for which the SD is calculated?

AR: We have reviewed the terminology throughout the manuscript to use "grid cells" instead of "pixels", which refer to the spatial unit (the 8.9 km spatial unit of our analysis). Regarding the

standard deviation (SD), we calculated the SD for the temporal population of radar backscatter observations. Specifically, for each individual grid cell, we calculated the SD of the radar backscatter values across the entire time series (1993–2020).

157. The CCI dataset has known biases: how are these accounted for?

AR: We acknowledge that the ESA CCI biomass dataset contains biases and uncertainties, particularly in dense tropical forests. These biases could potentially propagate into our AGB product.

To mitigate this, we implemented a quality control filter: we excluded grid cells where the coefficient of variation of the AGB estimates exceeded 80%. This threshold was applied to remove data points with high uncertainty and potential retrieval errors, thereby reducing the influence of biased observations on our analysis. Additionally, we are currently conducting a detailed analysis to quantify how the known bias in CCI propagates into our AGB product (see Reply to Reviewer 1 comment 2). This potential bias inherited from CCI biomass is discussed in the revised manuscript.

Para 3.1. How well does a model perform that is based solely on climate? How much does the scatterometer data add?

AR: To quantify the specific contribution of the radar backscatter data and climate variables, we performed a variable importance analysis using permutation importance. The results show that radar backscatter is the most important predictor for models. This indicates that the radar backscatter data captures significant variability in AGB that is not explained by climate variables.

Table 1 It's naïve to include the BIC to 7 significant figures.

AR: We thank the reviewer for pointing this out. We have revised Table 1 to report the BIC values to a more appropriate number.

216. This para is the wrong way round: the dataset underestimates high AGB values and overestimates low ones. This probably reflects biases introduced from CCI which also shows such effects. Name the 2 datasets used for comparison.

AR: We have corrected the statement. We acknowledge that this bias pattern is likely inherited from the ESA CCI biomass data used in the model training. Additionally, we have explicitly named the two datasets used for comparison in the revised manuscript.

224. How is it possible for there to be significant IAV in forests in arid and semi-arid regions? This effect is highly likely to be because the scatterometer data are particularly sensitive to soil moisture in such regions, and this is a soil moisture, not AGB, effect.

AR: As our answer before about the possible soil moisture influence on radar backscatter data, we conducted a partial correlation analysis (Fig. R2) to isolate the relationship between AGB, radar backscatter and soil moisture. The results show a substantially higher correlation between AGB and radar backscatter than that observed between AGB and soil moisture. This indicates

that the temporal variability in radar backscatter remains closely aligned with AGB changes, supporting the interpretation that the observed IAV reflects biomass dynamics, not soil moisture effects. A potential reason for such high IAV might be the fact that these regions are prone to fires, which can induce high IAV in AGC.

Fig. 1 The inventory data are different between the 3 plots. Why is this and how does this affect the results. The keys quote a bias, but in what quantity? What are the solid lines (I assume best linear fits)?

AR: We thank the reviewer for these careful observations. We have revised Figure 1 and its caption to address these points as follows:

Differences in inventory data: The differences between the three plots were initially caused by the varying spatial coverage of the three AGB datasets, which resulted in different numbers of common grid cells being available for comparison. To ensure a fair and consistent comparison, we have updated the figure so that all three datasets are compared using the exact same set of grid cells.

Bias units: We have added the specific units (Mg ha^{-1}) to the bias values reported in the figure keys.

Solid lines: We have clarified in the figure caption that the solid lines represent the best linear fits (linear regression) between the estimated and reference AGB values.

Fig.2 What does the SD represent in 2C? The error bars shown are not credible as indicators of the true uncertainties in these estimates.

AR: In Figure 2C, the bar plot represents the mean aboveground biomass carbon stock for each biome averaged over the period 1993–2020. The bars shown in the figure represent the standard deviation of the biomass carbon stock over time (i.e., interannual variability) between 1993 and 2020, rather than the confidence interval or total uncertainty of the mean estimate.

We have revised the figure caption to explicitly state that these bars represent the temporal standard deviation. Also note that we will provide a more detailed uncertainty assessment.

260 What is meant by temporally consistent?

AR: “temporally consistent” here indicates that the AGB estimates maintain a consistent level and scale from year to year, rather than a single year or a set of discrete years. We have rephrased ‘temporally consistent’ to “annually resolved”.

272. I don’t know if the authors are looking at a different plot, but 100-300 is not consistent with Fig. 1A.

AR: We have revised the value range to consistent with the values shown in the Figure 1A.

275. What are these structural differences and why is bias inevitable? What would be the sources of any bias, and would it be in the inventory data or the estimator used in this paper?

AR: The structural differences refer to the mismatch between the spatial scale and nature of the two data sources. Inventory data are discrete point measurements (plots) that capture local forest structure at specific locations, whereas radar-based AGB estimates represent an area-averaged signal over the entire 8.9 km grid cell. Within a single grid cell, forest structure can vary substantially due to natural heterogeneity, land-use mosaics, or disturbance history, meaning that the averaged plot measurements may not fully represent the true structural mean of the pixel.

Bias is inherent in plot-to-map comparison because (a) several plots (even >10) cannot fully capture the spatial variability of forest structure within a grid cell, and (b) the radar backscatter signal integrates over the entire grid cell, including areas not sampled by inventory plots. This scale and representativeness mismatch makes the one-to-one comparison challenge. We have added a discussion of these structural differences and potential bias to the revised manuscript.

280 the paper should remind the reader that CCI is used to train the algorithm used. The para. mentions their estimates being “slightly” smaller than some others, but in several cases they are significantly lower (20-25%). The last sentence is not very sensible; inventories don’t take place in wetlands, barren land etc.

AR: We agree that the term “slightly” understated the differences between our estimates and those reported in some other studies. We have removed it and now describe the differences more accurately. We have also clarified that CCI biomass product was used to train the model. We revised the last sentence to explain that differences in estimated biomass carbon may be partly caused by differences in the forest area size used in the respective analyses. We now calculate the forest area in our study and compare it with the area used for biomass-carbon estimation in the inventory-based studies.

Table 2. Where does the value for 1990 come from? It’s not in the period covered by the dataset.

AR: In the original manuscript, we used the 1993 value as a proxy for 1990, but we agree this is misleading and should not be included. We have now removed the 1990 from Table 2 to ensure accuracy.

295 delete dynamics

AR: done, thanks.

333. I what sense is the dataset robust?

AR: We agree that describing the dataset as “robust” might be vague and potentially overstated. We have removed this subjective term and rephrased the sentence to provide a more precise description of the dataset.

“Despite the dataset being supported by multiple independent inputs and quality-control procedures, several sources of uncertainty remain.”

337. Earlier the paper says Ku-band scatt data are used, not just C-band. Which is correct?

AR: We have corrected it.

344. What is meant by “spatial heterogeneity in calibration data” for CCI -Biomass? Which calibration data are being referred to, since CCI-Biomass does not use in situ data for calibration?

AR: We used CCI AGB to train and calibrate our model. By "spatial heterogeneity in calibration data," we refer to the inherent spatial variability and uneven distribution of the CCI Biomass estimates, which can introduce uncertainty into our model calibration. We have clarified the sentence and will further provide a more detailed uncertainty assessment and output uncertainty flags.

342. What is meant by uncertainty throughout this paragraph? For example, what actual quantity is displayed in Fig. 5? If this is just precision then it tells us little about error or bias and is misleading as regards the overall uncertainty in the dataset.

AR: Thanks for this clarification. We agree that the term "uncertainty" was used loosely and have revised the manuscript to define it precisely. Figure 5 specifically displays the model prediction uncertainty (precision), which quantifies the variability of the model predictions. We clarified the caption of Figure 5. We acknowledge that this represents only the component of uncertainty related to model precision and does not include sources such as calibration bias in the training data (e.g., ESA CCI biomass), sensor transition offsets, fixed global conversion factors, or C-band saturation in dense forests. We are quantifying these potential uncertainties and will expand the discussion to ensure that users are aware of the uncertainty for this product (also see the reply to reviewer 1 comment 2).

347. should say backscatter. What the forest structure predictors referred to here?

AR: Corrected. The forest structure predictors have been removed.

356. If the resolution of the data is 25 km, then it is wrong to say the spatial resolution of the dataset is 8.9 km. This is the interpolated grid spacing but the underlying resolution is unchanged.

AR: The radar backscatter dataset used (Tao et al., 2023) in our study has a native spatial resolution of 8.9 km, which is the spatial scale of the input data used to train our model. The approximately 25 km resolution mentioned in the manuscript refers to products from other studies, not to our dataset. We have rephrased the sentence to make this distinction clear and avoid implying that 25 km is the resolution of our product.

362. What is meant by annually consistent?

AR: "Annually consistent" here means that the product provides an estimate for every year from 1993 to 2020 with no temporal gaps, i.e., a continuous annual time series rather than a set of discrete snapshots.

368 get rid of dynamical in 4.3; it is just annual

AR: removed.

369. these are not “direct” observations of biomass changes but very indirect estimates using data that are fairly insensitive to woody biomass.

AR: We have revised the sentence.

372 these are not continuous temporal observations but just annual estimates

AR: modified.

Some general remarks on terminology:

The word “dynamic” is misused: most of its occurrences in the paper should be replaced by “annual”.

The word “continuous” is used in several places and it is often unclear whether this means spatially or temporally continuous, e.g. l.39, l.46, but many others. But it mainly serves no purpose: by definition a map will be spatially continuous, and temporally continuous really means annual in this paper.

The paper sometimes talks about grid cells and sometimes about pixels and it’s not clear if these are synonymous; use of a single term would be better (unless these indicate different things).

In many places the paper needs to be specific and say “scatterometer” rather than just “radar”, e.g., l. 89, 178, but many others.

AR: We thank the reviewer for these careful and constructive suggestions regarding terminology. We have reviewed the entire manuscript and made the corrections to ensure precision and consistency.

Response to Editor

The manuscript presents a potentially valuable global annual forest above-ground biomass (AGB) record for 1993-2020. A record of this duration could become a useful community resource. However, I agree with both referees that substantial revision and additional validation are required before the product can be considered reliable for interpreting interannual-to-decadal biomass change. The map-level agreement reported for biomass stocks does not, by itself, establish the fidelity of the temporal signal, which is the main novelty of the dataset.

My main editorial requirements, complementing the detailed referee reports, are as follows.

1. Evaluate spatial and temporal generalization separately: The selected model obtains most of its map-level predictive skill after the addition of static background climate variables, whereas all year-to-year variability must come from the radar record. The current performance statistics therefore mix spatial stock prediction with temporal-change prediction. Moreover, the public training code constructs cross-validation groups by combining year with 0.5-degree spatial blocks. This permits observations from the same spatial block in different years to occur in different folds; the 2020 holdout is temporally, but not spatially, independent. Please revise or clarify the evaluation so that spatial blocks are held out across all years, and add temporal and sensor-era holdouts designed specifically to test annual anomalies and trends. At minimum, report performance for anomalies and/or first differences, by biome and biomass range, in addition to performance for biomass levels.

AR: We sincerely thank the editor for the insights and constructive comments. We agree that the original cross-validation design does not ensure spatial independence across years, but rather performs a joint space-time cross-validation. We considered it useful for assessing joint space-time generalization, which is relevant to applying the model across locations and years. But we agree that this design does not ensure spatial independence across years. In response, we have restructured the evaluation framework to add spatial block cross-validation to assess spatial generalization.

Assessment of spatial and temporal generalization in model evaluation:

We will compare our results with spatial-block-only cross-validation, in which entire 0.5° spatial blocks are held out across all years, ensuring that no observations from a given region are used in training. This approach provides a more realistic test of spatial generalization. We will compare the model's performance under this spatial-block scheme with that of the original spatiotemporal block design and report both results in the revised manuscript. Regarding temporal and sensor-era holdouts: due to the temporal span of our training data (2017–2019) and the high correlation between consecutive years, we were unable to construct meaningful temporal holdouts without introducing severe overfitting. Similarly, sensor-era holdouts are not feasible given the training data from the same sensor periods.

Evaluation of model performance across biomes and biomass ranges:

We have added an evaluation of model performance across biome and biomass ranges. For each biome (e.g., tropical rainforest, temperate forest, boreal forest) and biomass range, we will report key performance metrics including R^2 , bias, and RMSE. This analysis will provide critical insight into the model's ability to capture biomass dynamics across a range of ecosystem types and carbon stocks.

We hope these revisions will strengthen the evaluation framework and provide a more rigorous assessment of the robustness of our results. We will update the Methods, Results, and Discussion sections accordingly.

2. Demonstrate temporal homogeneity and isolate the biomass signal: Please provide explicit diagnostics for all ERS, QSCAT, and ASCAT overlap and transition periods, not only the visible 1997 anomaly. These should include breakpoint or change-point tests, estimated offsets, any corrections applied, and a clear quality flag or exclusion rule for affected years. The authors should also show that the reconstructed anomalies and trends remain after accounting for soil moisture and vegetation-water effects. Useful tests could include comparisons in demonstrably stable forest, areas with independently documented disturbances or recovery, and anomaly/partial-correlation analyses against radar and soil-moisture series. The central criterion is whether the radar-derived temporal signal represents biomass change rather than hydrological variability or sensor discontinuities.

AR: We thank the editor for this important and constructive feedback. We fully agree that ensuring temporal homogeneity across sensor transitions and isolating the true biomass signal from confounding factors such as soil moisture and sensor discontinuities are critical for the credibility of our product. In response, we will provide new diagnostics and analyses to address these concerns.

Assessment of sensor influence (including breakpoint analysis and quality flags):

We acknowledge the importance of assessing temporal homogeneity across sensor transitions. In response, we are currently conducting breakpoint detection analyses for the global AGB time series. These analyses aim to identify potential shifts in the AGB time series and estimate the magnitude of any offsets. Based on the results, we will define quality flags to indicate years affected by sensor transitions (e.g., ± 1 year around detected breakpoints) and apply exclusion rules for trend and anomaly analysis in those periods. These quality flags will be included in the dataset metadata and clearly described in the revised manuscript. While the full analysis is ongoing, we will report the detected breakpoints, estimated offsets, and associated uncertainty in the final version.

Isolation of the biomass signal from non-biomass influences:

To ensure that the temporal signal reflects biomass change rather than hydrological variability, we have conducted a partial correlation analysis between AGB, radar backscatter, and soil moisture across the global domain. The results (Fig. R2) show a strong partial correlation between AGB and radar backscatter (median $r = 0.54$), while the partial correlation with soil moisture is negligible (median $r = 0.007$). This indicates that the temporal variability in the radar signal is primarily linked to biomass dynamics, not residual soil moisture effects.

3. Clarify the quantities called gross gain, gross loss, and net sink: As presently described, gains and losses appear to be aggregated after pixels are classified by their full-period positive or negative trend. If so, these quantities are a decomposition by a posteriori trend class, not independent estimates of annual gross biomass gains and losses. Please define the calculation precisely, discuss the selection implied by this classification, and use terminology that cannot be confused with gross carbon fluxes. The manuscript should also

reconcile the reported global AGC rates of 0.29 and 0.44 PgC yr⁻¹ by stating the estimator, period, and spatial mask used for each. Causal statements attributing regional changes to CO₂ fertilization, management, drought, fire, or land-use change should be framed as hypotheses unless attribution is directly tested.

AR: We have carefully checked each terminology through the manuscript, and revised the Methods section to clearly define how gross gain and gross loss are computed. The revised manuscript now clarified the grid classification, aggregation procedure, and limitations of this approach. We have clarified the source of the two reported rates (0.29 and 0.44 PgC yr⁻¹). Finally, we revised the discussion of potential drivers to present them as hypotheses where causal attribution was not directly tested.

4. Correct the treatment of resolution and uncertainty: Please distinguish the 8.9 km grid spacing from the effective spatial resolution supported by the native scatterometer footprints and the merged radar product, and revise claims about detecting small disturbances accordingly. Equation (1), Figure 5, the error bars, and the quoted ± 2.5 PgC must be made mutually consistent: the equation appears to describe variability in fold-level mean error, whereas Figure 5 is spatially varying. The revised paper and data release should state exactly what each uncertainty quantity represents and should provide per-cell uncertainty and quality information. Uncertainty or sensitivity in stock and trend estimates should address, as far as possible, training-product bias, biomass-range compression evident against inventories, sensor harmonization, forest masks and exclusions, carbon fraction, and the root-to-shoot conversion. Claims that the product is a "benchmark" should be tempered unless these uncertainties are adequately characterized.

AR: Thank you very much for the detailed and constructive feedback.

Clarification of spatial resolution

The 8.9 km resolution used in this study refers to the spatial resolution of the long-term radar backscatter dataset produced by Tao et al. (2023), which we used as the primary input. This dataset results from merging and spatially aggregating raw radar scatterometer observations (ERS, ASCAT, QSCAT), and the 8.9 km represents the effective output grid spacing and the spatial scale of our analysis. We have revised the manuscript to clearly state that 8.9 km is the spatial resolution of the input radar product and the scale of our study, and we have tempered claims about detecting small-scale disturbances.

Uncertainty

In the revision, we expand the uncertainty discussion to explicitly address multiple sources of uncertainty.

- Model uncertainty: The uncertainty (Eq. 1) reflects the standard deviation of model prediction error across cross-validation folds, i.e., model uncertainty. It is not a full

representation of total uncertainty. We now clarify this in both the Methods and Discussion sections.

- Uncertainty from ESA CCI biomass bias: We will include an analysis to quantify how the CCI biomass bias can propagate on AGB. The estimated impact will be included in the quality flag and dataset metadata.
- Uncertainty from radar backscatter data: We are performing a breakpoint analysis on the AGB time series to assess potential influence of different sensors of the radar backscatter data. We will discuss uncertainties related to sensor harmonization, C-band limitations, and residual sensitivity to soil moisture.
- Uncertainty from fixed carbon fraction: We acknowledge that using a global constant (0.49) to convert biomass to carbon ignores species- and tissue-specific variation, which may introduce uncertainty in carbon estimates. This is now discussed in the manuscript.

We will include all these sources in the revised discussion, and this quantified uncertainty will be added as the quality flag along with the AGB dataset. We have tempered claims that the product is a “benchmark” to be “reference”.

5. Make the archived product and code independently usable and reproducible: The current Zenodo record provides a single compressed NetCDF with only a brief landing-page description, while the GitHub repository contains six workflow scripts and a short overview; the scripts use project-specific paths and no reproducible software environment is supplied. Before acceptance, please provide a data user guide and complete NetCDF metadata, including variable definitions, units, coordinates/CRS, cell area, forest-mask definition, fill values, valid ranges, uncertainty fields, and sensor/year quality flags. The code release should include an open-source license, pinned environment or dependency file, configurable paths, exact commands and processing order, input dataset versions, model settings and random seeds, and an archived release tied to the manuscript DOI or an immutable commit. If model files or intermediate inputs are required to reproduce the published NetCDF, they should be archived or their generation documented unambiguously.

AR: We sincerely thank you for these detailed and constructive comments regarding data and code reproducibility. We agree that providing a complete, independently usable archive is essential for the community. In response to these suggestions, we have undertaken the following major updates to both the Zenodo data record and the GitHub repository:

Data user guide and metadata: We have added a comprehensive Data User Guide to the Zenodo record. This document includes detailed variable definitions, units, coordinate reference systems (CRS), cell area, forest-mask definitions and descriptions of uncertainty fields and sensor/year quality flags. The NetCDF metadata has also been updated.

Code usability and environment: We have included the following in the GitHub repository:

Reproducible Environment: We have added an environment.yml and a requirements.txt file to pin all specific package versions, ensuring the software environment can be exactly reproduced.

Reproducibility and Archiving: We have documented the exact model settings and random seeds used in the training process. We have also created a new archived release on GitHub tied to a specific commit hash (and the manuscript DOI), which includes all scripts necessary to reproduce the results.

We thank the reviewer for prompting these improvements, which we believe make the product and workflow independently usable and reproducible.

Please also complete a careful consistency and language revision, including the discrepancies already identified by the referees (for example, test $R^2 = 0.833$ versus 0.883; six versus seven biome groups; C-band-only language despite the merged C- and Ku-band record; and terminology such as "continuous," "dynamic," "pixel," and "grid cell").

AR: We have conducted a thorough consistency and language revision throughout the manuscript to ensure accuracy, clarity, and scientific precision.

I would be willing to reconsider a substantially revised manuscript that addresses both referee reports and the points above. The decisive issue is not whether the model reproduces broad spatial biomass gradients, but whether the authors can establish a homogeneous, uncertainty-characterized, and independently usable annual record of biomass change.

AR: We thank the editor for the detailed and constructive feedback, and for the opportunity to submit a substantially revised manuscript. We address concerns raised by the reviewers and the editor, and carefully revised every section to ensure clarity, accuracy, and transparency in the revised manuscript.