Responses to Reviewers’ Comments for Manuscript ESSD-2026-102

FineKarstAGB: A 30 m resolution aboveground biomass dataset for Southwest China derived by upscaling plot-level inventory using sub-meter GaoFen satellite data

Addressed Comments for Publication to

Earth System Science Data

by

Yixiang Li, Yongqing Bai, Zhengchao Chen, Caixia Liu, Xuan Yang, Xiaowei Tong, Martin Brandt, Zhaoming Wu, Zeqing Wang, Huaming Gao, Xiaoyi Wang, Sizhuo Li, Xin Xu, and Siyu Liu

Response Letter

Dear Editors,

Please find enclosed the revised version of our submission entitled *"FineKarstAGB: A 30 m resolution aboveground biomass dataset for Southwest China derived by upscaling plot-level inventory using sub-meter GaoFen satellite data"* with manuscript number ESSD-2026-102. We thank the editor, reviewers, and community commenters for their careful evaluation. The revised manuscript, tracked-changes manuscript, supplementary information, and point-by-point responses are provided in this revision package.

We have carefully addressed every comment point by point and revised the manuscript and supplement accordingly. In the following pages, reviewer and community comments are reproduced in light-blue boxes, and our replies are introduced by **Response:**. The corresponding revised or added passages and their locations are reported under **Changes in manuscript:**. In the tracked-changes manuscript, revised and added text is shown in red while line numbers remain black; the clean manuscript contains the same text without revision colouring.

Sincerely,

Yongqing Bai
Corresponding author, on behalf of all authors

Authors' Response to Reviewer 1

General Comment

The authors propose a Canopy Structure-driven Multi-feature Fusion Network (CSMFNet) to produce a 30 m resolution aboveground biomass dataset for Southwest China by upscaling plot-level inventory using sub-meter GaoFen satellite data. Based on independent NFI plot data and UAV LiDAR derived CHM data, they show that the new regional dataset outperforms existing global and national-scale AGB products, which can provide critical support for regional studies. While the proposed CSMFNet seems appealing by fusing the forest spectral information with its horizontal and vertical structural characteristics to achieve high-precision AGB estimation, the major issue lies in the temporal gap between ABG label and remote sensing features. The NFI plot data were collected in 2014-2018 whereas the GaoFen imagery were collected in 2023-2024. I do not think that the filtering procedure based on a random forest model (lines 180-185) can mitigate this issue, because samples used to train and validate the random forest model were associated with uncertainties and lacking of the temporal information.

Response: Thank you for recognizing the value of FineKarstAGB and the advantages of integrating spectral information with the horizontal and vertical structural characteristics of forests. We also appreciate your important concern regarding the temporal gap between the NFI measurements and the GaoFen observations.

We agree that the random-forest procedure cannot directly update the historical NFI measurements to 2024 or completely eliminate temporal uncertainty. Strict plot-by-plot temporal coincidence is rarely attainable in regional wall-to-wall biomass mapping because forest inventories and satellite observations are generally acquired according to different multiyear schedules. In this study, the random-forest procedure was used as a sample quality-screening step to identify plots showing severe disagreement between historical plot AGB and the NDVI, crown coverage, and canopy height observed in the recent imagery. It was not intended to correct the historical NFI measurements to 2024.

To independently assess the temporal consistency of the retained plots, we examined repeated NFI8–NFI9 observations using annualized AGB changes together with stand origin, stand age, Hansen forest-loss records, and persistent changes in Landsat NDVI, NBR, and NDMI. Among the 2,235 matched retained plots, 1,852 plots (82.9%) met the temporal-stability criteria. This result shows that the retained dataset predominantly consists of plots with consistent historical inventory trajectories.

We further quantified the possible influence of the remaining temporal gap using an independent nonlinear growth-slowdown model based on repeated NFI8–NFI9 observations and stand and environmental attributes. The resulting sensitivity scenarios suggested median potential AGB increases of 13.45–

20.23 $\mathrm{Mg\,ha^{-1}}$ over assumed intervals of 6–10 years (Fig. S8). These estimates were used to characterize temporal uncertainty rather than to update the historical NFI labels or the resulting FineKarstAGB map.

Taken together, the high proportion of temporally stable plots and the quantitative sensitivity analysis support the retained NFI data as a reference for regional AGB modelling while explicitly characterizing the remaining temporal uncertainty. Accordingly, we revised only the relevant explanation of the random-forest screening in Section 2.3.1, added the independent temporal-consistency and growth-slowdown results to Section 5.3.1, and documented the growth-slowdown method and supporting scenarios in Supplementary Text S1 and Fig. S8.

Changes in manuscript: *Section 2.3.1, "National Forest Inventory (NFI) data" (page 7, lines 178–195) [REVISED]*

To identify gross inconsistencies between historical plot AGB and the spectral and structural conditions represented by the recent imagery, we applied an automated cross-source consistency screening workflow.

This procedure was used as sample-compatibility quality control rather than as a direct temporal correction; it removed plots exhibiting severe structural–label incompatibilities while maintaining the representativeness and diversity of the AGB distribution.

Section 5.3.1, "Input-data and plot-matching uncertainties" (page 27, lines 574–590) [ADDED]

To examine the historical stability of the retained plots using information independent of the CSMF-Net inputs, we additionally evaluated the 2,235 retained plots that could be matched between NFI8 and NFI9. In total, 1,852 matched plots (82.9%) showed stable temporal trajectories. This result indicates that the retained dataset was predominantly composed of plots with consistent historical forest conditions. Nevertheless, it does not demonstrate that every retained plot remained unchanged until the acquisition of the GaoFen imagery, and the screening should be interpreted as consistency-oriented quality control rather than a direct temporal correction of the historical NFI AGB values.

To further quantify the remaining temporal uncertainty, an independent nonlinear growth-slowdown analysis based on repeated NFI8–NFI9 observations estimated median potential AGB increases of 13.45–20.23 $\mathrm{Mg\,ha^{-1}}$ *over assumed intervals of 6–10 years (Fig. S8). These scenario-based estimates were not used to update the NFI labels or FineKarstAGB map, but indicate the possible magnitude of the remaining temporal uncertainty.*

Supplementary Text S1, "Growth-slowdown sensitivity analysis" (page 2), and Fig. S8 (page 10) [ADDED]

Comment 1

The authors evaluate the agreement using R^2 and r inconsistently. I suggest using the same measure consistently throughout the paper.

Response: Thank you for this helpful comment. In the original manuscript, Pearson's correlation coefficient (r) was used for the direct comparison between predicted and NFI-observed AGB because this analysis focused on the strength and direction of their linear relationship. In contrast, the coefficient of determination (R^2) was used for the UAV LiDAR CHM assessment because the relationship between AGB and canopy height was described using the nonlinear allometric model $AGB = aH^b$, and R^2 was used to quantify its goodness of fit. Thus, the two statistics were originally selected for different analytical purposes.

Nevertheless, we agree with the reviewer that alternating between r and R^2 may cause unnecessary confusion and reduce the consistency of the evaluation results. We have therefore recalculated and reported R^2 consistently as the primary goodness-of-fit metric for all AGB evaluations conducted in this study, including the comparisons with NFI observations and the structural assessments using UAV LiDAR-derived CHM data. The corresponding values and descriptions have been updated throughout the Abstract, Results, figure annotations and captions, Supplement, and Conclusions. The accompanying MAE, RMSE, and overall bias metrics in the NFI-based evaluation have been retained to characterize the magnitude and systematic pattern of prediction errors.

Changes in manuscript: *Abstract (page 1, lines 27–32) [REVISED]*

Accuracy assessment against independent NFI plot data demonstrated the model's robust performance ($R^2 = 0.69$, RMSE = 28.51 $\mathrm{Mg\,ha^{-1}}$), showing no evidence of saturation in high-biomass regions. Furthermore, a structural consistency assessment using an independent UAV LiDAR-derived canopy height model confirmed that FineKarstAGB maintains a strong relationship with forest vertical structure ($R^2 = 0.54$). Other public datasets showed weak relationships with both NFI observations ($R^2 \leq 0.12$) and UAV LiDAR-derived canopy height ($R^2 < 0.1$).

Section 4.1, "Evaluation with NFI data" (pages 14–15, lines 353–355 and 375–380) [REVISED]

The CSMF-Net predictions exhibited strong agreement with the independent NFI observations, achieving $R^2 = 0.69$, MAE = 18.64 $\mathrm{Mg\,ha^{-1}}$, RMSE = 28.51 $\mathrm{Mg\,ha^{-1}}$, and an overall bias of 2.38% (Fig. 5a).

The evaluation results for the comparison datasets are presented in Fig. 6. Existing global- and national-scale AGB products generally exhibited weaker relationships with the NFI observations in the study area, with R^2 values ranging from 0.03 to 0.13 and RMSE values ranging from 51.16 to 72.25 $\mathrm{Mg\,ha^{-1}}$. Specifically, the R^2 values were 0.13 for Yang Map, 0.04 for Yan Map, 0.07 for ESA CCI, and 0.03 for Cai Map. The Cai Map also showed a pronounced systematic bias, with an overall bias of 33.66%. Moreover, the fitted linear relationships had consistently low slopes (0.13–0.32), accompanied by substantial underestimation in high-AGB regions ($> 200\ \mathrm{Mg\,ha^{-1}}$), indicating varying degrees of saturation.

Conclusions (page 29, lines 635–636) [REVISED]

The proposed CSMF-Net achieved robust AGB estimation when evaluated against independent NFI plots ($R^2 = 0.69$, RMSE = 28.51 $\mathrm{Mg\,ha^{-1}}$).

Comment 2

The tree crown segmentation dataset is an important input for the model. How accurately were the trees delineated? It is necessary to report accuracies measures such as F1, recall, and precision, as well as how the accuracy assessment was performed. What does R^2 (0.82) refer to? It is not a commonly used metric in individual tree segmentation.

Response: Thank you for this important comment. We agree that the original manuscript did not clearly explain the accuracy assessment and the meaning of the reported R^2. The $R^2 = 0.82$ reported by Xu et al. (2026) refers to the coefficient of determination between the manually labelled and predicted numbers of trees across image-interpretation plots in Guangxi. Thus, it represents plot-level agreement in tree counts.

To provide segmentation-specific accuracy measures across the present study area, we further clarified the existing independent assessment based on 919 image-interpretation plots distributed across different forest types and topographic conditions. These are image-based validation units and are unrelated to the NFI plots used for AGB estimation. All visually identifiable tree crowns within these plots were manually delineated, yielding 9,061 reference crowns. The manually delineated crowns were converted into binary reference masks and compared pixel by pixel with the predicted crown masks. The resulting precision, recall, and F1 score were 0.832, 0.853, and 0.843, respectively.

We have revised Section 2.2.2 to report these accuracy measures and explain how the accuracy assessment was performed.

Changes in manuscript: *Section 2.2.2, "Tree crown segmentation dataset" (page 6, lines 139–150) [REVISED]*
Tree-crown segmentation accuracy was assessed using 919 independent image-interpretation plots, separate from the NFI plots and covering the major forest types and topographic conditions across the study area. Within these plots, 9,061 crowns were manually delineated, and pixel-wise comparison of the binary crown masks yielded a precision of 0.832, a recall of 0.853, and an F1 score of 0.843. The $R^2 = 0.82$ *reported by Xu et al. (2026) refers to the plot-level agreement between labelled and predicted tree counts in Guangxi.*

Comment 3

Section 2.2.3: does the UAV LiDAR data cover the same area of canopy height dataset or just part of it? Please provide more details about the canopy height dataset, such as collection time, point density, location, spatial resolution of CHM, etc.

Response: Thank you for this comment. We agree that the relationship between the UAV LiDAR data and the canopy height dataset should be described more explicitly. To address this concern, the UAV LiDAR data cover selected areas within Guangxi Province rather than the full spatial extent of the canopy height dataset. They comprise 18 UAV LiDAR datasets acquired in 2020, with a combined CHM coverage of approximately 1,208 ha. These datasets are distributed across Guilin, Fangchenggang, and Wuzhou. The LiDAR point clouds have an average density exceeding 10 points m^{-2}, and the resulting CHMs have a spatial resolution of 0.05 m.

In contrast, the canopy height dataset used as an input for AGB estimation is a wall-to-wall product covering all four provinces in the study area. As described by Tong et al. (2025), the dataset was generated at a spatial resolution of 0.8 m using a U-Net model with a ResNet-50 backbone trained with airborne LiDAR-derived CHMs and sub-meter GaoFen imagery. Therefore, the UAV LiDAR data described above constitute an additional independent validation dataset and cover only a portion of the regional canopy height dataset. We have accordingly expanded Sections 2.2.3 and 2.3.2 to clarify the coverage, role, and characteristics of the two datasets.

Changes in manuscript: *Section 2.2.3, "Canopy height dataset" (page 6, lines 151–161) [REVISED]*

This study used a wall-to-wall canopy-height dataset with a spatial resolution of 0.8 m covering the entire study area. As described by Tong et al. (2025), the dataset was generated using a U-Net model with a ResNet-50 backbone, with sub-meter GaoFen imagery as input and airborne LiDAR-derived CHMs as reference data. The dataset achieved an MAE of 1.94 m and an RMSE of 2.76 m on an independent validation dataset, with no evident saturation in the higher canopy-height range. These errors were substantially lower than those reported by Tong et al. (2025) for the canopy-height products of Liu et al. (2022), Lang et al. (2023), and Potapov et al. (2021) (MAE ≥ 5.47 *m; RMSE* ≥ 6.98 *m).*

Section 2.3.2, "UAV LiDAR data" (pages 7–8, lines 197–206) [REVISED]

The evaluation CHM data used in this study were obtained from LIDARNET (http://lidar.pku.edu.cn) and covered typical forest scenes in Guangxi Province. The data comprise 18 UAV LiDAR datasets acquired in 2020, covering approximately 1,208 ha across Guilin ($n = 10$)*, Fangchenggang* ($n = 6$)*, and Wuzhou* ($n = 2$)*. The LiDAR point clouds were acquired using Riegl VUX-1, with an average point-cloud density exceeding 10* $\mathrm{pts\,m^{-2}}$*. The resulting CHMs have a spatial resolution of 0.05 m. Their spatial distribution and a representative example are shown in Fig. S4.*

Comment 4

Section 2.3.2: consider adding a map to show the coverage of the UAV LiDAR. When were the UAV LiDAR data collected? What's the spatial resolution of UAV LiDAR derived CHM?

Response: Thank you for this helpful suggestion. We agree that a spatial display of the UAV LiDAR data would provide a clearer understanding of their distribution and coverage. We have therefore added a new supplementary figure showing (a) the locations of the 18 UAV LiDAR datasets across Guangxi Province, (b) the GaoFen image corresponding to a representative UAV LiDAR site, and (c) the associated UAV LiDAR-derived CHM. The UAV LiDAR data were acquired in 2020 and cover approximately 1,208 ha in total, while the resulting CHMs have a spatial resolution of 0.05 m. These details have also been added to Section 2.3.2, as described in our response to Comment 3.

Changes in manuscript: *Supplementary Fig. S4 [ADDED]*

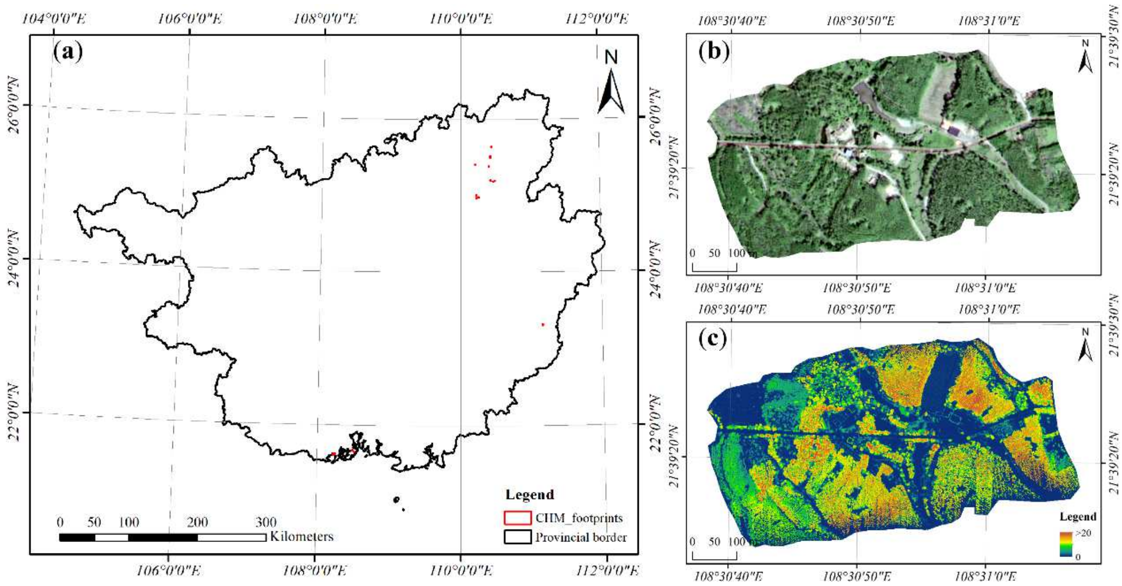


Figure S4. Spatial distribution and example of the UAV LiDAR data used in this study. (a) Distribution of the 18 UAV LiDAR datasets across Guangxi Province, including Guilin ($n = 10$), Fangchenggang ($n = 6$), and Wuzhou ($n = 2$). (b) GaoFen imagery of a representative UAV LiDAR site. (c) Corresponding UAV LiDAR-derived canopy-height model. Panels (b) and (c) show the same spatial extent. The UAV LiDAR data cover approximately 1,208 ha in total, and the CHMs have a spatial resolution of 0.05 m.

Comment 5

The spatial resolution of ESA CCI product is different from that of the product in this study. How did the authors address this scale mismatch between the ESA CCI pixels and NFI plots?

Response: Thank you for this important comment. For the NFI-based evaluation, the ESA CCI product was retained at its native 100 m resolution, and the value of the ESA CCI pixel containing the center of each NFI plot was extracted. This approach establishes a direct spatial correspondence between each plot and the original ESA CCI observation without introducing interpolated values. Nevertheless, because the ESA CCI pixel represents a larger spatial support than an individual NFI plot, we have clarified that mixed-pixel effects and spatial-scale mismatch may affect the plot-level comparison.

For the pixel-wise comparison in the Discussion, ESA CCI was originally resampled from 100 m to 30 m to align with the FineKarstAGB grid. We appreciate the reviewer's comment, which made us recognize that, although this operation aligns the two raster grids, it does not increase the effective spatial resolution of ESA CCI and may amplify apparent local differences. We therefore revised the analysis by aggregating FineKarstAGB to the native 100 m ESA CCI grid using area-weighted averaging. The pixel-wise differences and associated statistics were then recalculated using spatially overlapping valid pixels at the 100 m scale. The recalculated difference map retained the same overall spatial pattern, while the mean difference and standard deviation changed only slightly, from 6.40 and 57.17 $\mathrm{Mg\,ha^{-1}}$

to 8.09 and 54.52 $\mathrm{Mg\,ha^{-1}}$, respectively. Therefore, this scale adjustment did not affect the original interpretation or the main conclusions. The corresponding statistical results and Fig. 13c have been updated accordingly.

Changes in manuscript: *Section 5.1, "Comparison with other AGB datasets", and Fig. 13 (pages 22–24, lines 475–524; Fig. 13 on page 25) [REVISED]*

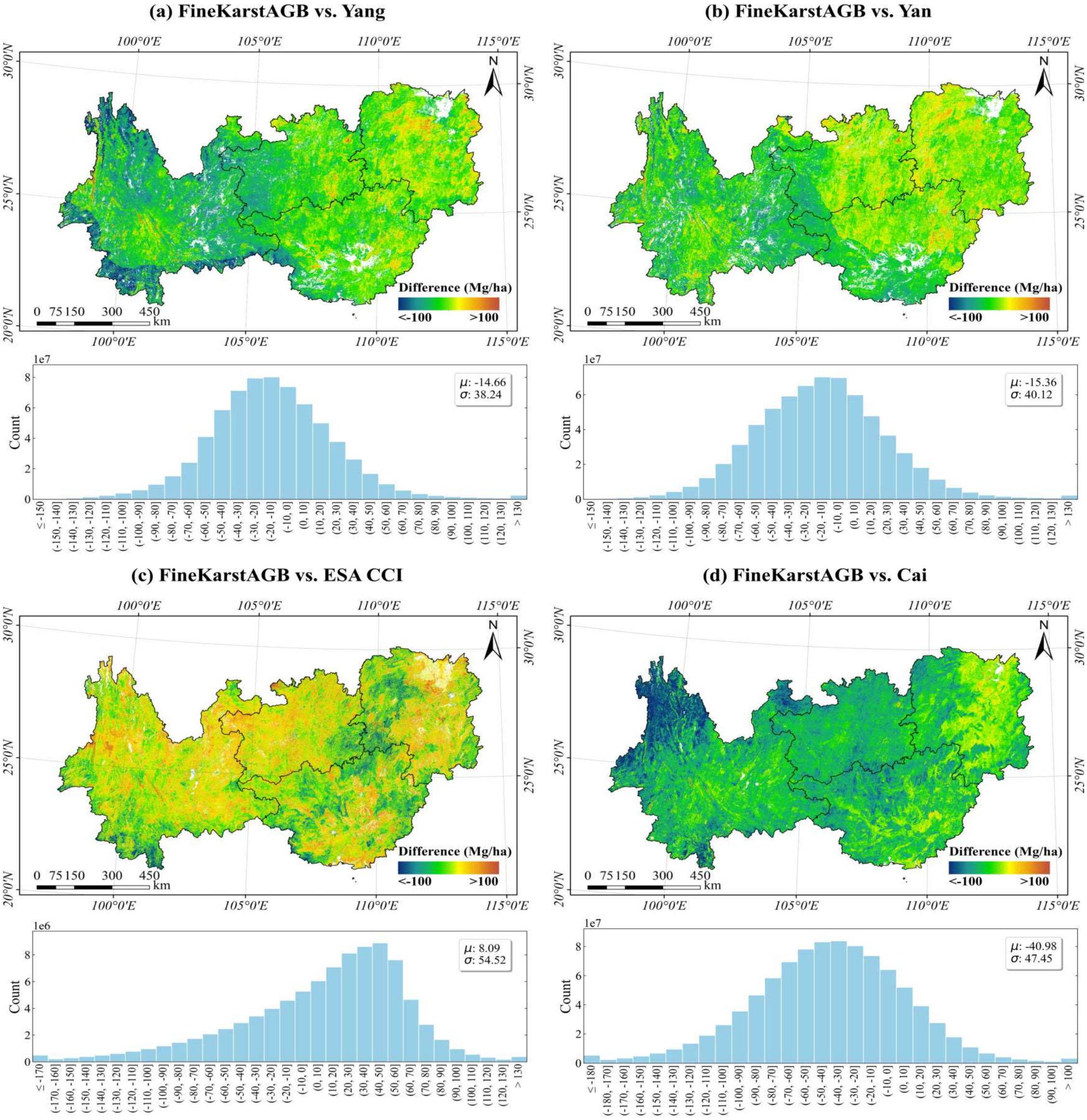


Figure 13. Spatial distribution and statistics of AGB differences between FineKarstAGB and the comparison datasets: (a) Yang et al. (2023), (b) Yan et al. (2023a), (c) ESA CCI, and (d) Cai et al. (2024b). Pairwise differences were calculated after spatial alignment using only pixels with valid estimates in both products. FineKarstAGB was aggregated to the native 100 m grid for comparison with ESA CCI. μ and σ denote the mean and standard deviation of the AGB differences, respectively.

Comment 6

According to Table 1, there is significant temporal gap between the AGB datasets and NFI plots. How reliably could the comparison below evaluate the performances of different products?

Response: Thank you for raising this important concern. We acknowledge that the temporal gap between the 2014–2018 NFI observations and the AGB products may affect their absolute agreement. We would nevertheless like to clarify that the original evaluation compared the most recent available versions of the products against the same historical NFI reference framework. The comparison therefore provides a consistent assessment of their agreement with the available plot observations.

To determine whether temporal mismatch was sufficient to explain the observed differences, we conducted an additional sensitivity analysis for the Yan, ESA CCI, and Cai products using their annual estimates (Supplementary Fig. S5) and per-plot median estimates for 2015–2018. To isolate the influence of product year from changes in data availability, all annual and median estimates for each product were evaluated using a fixed product-specific set of plots. Using the 2015–2018 median estimates reduced RMSE from 50.95 to 45.90 $\mathrm{Mg\,ha^{-1}}$ for Yan, from 71.09 to 64.54 $\mathrm{Mg\,ha^{-1}}$ for ESA CCI, and from 73.68 to 66.66 $\mathrm{Mg\,ha^{-1}}$ for Cai. The corresponding MAE reductions were 3.99, 5.77, and 5.52 $\mathrm{Mg\,ha^{-1}}$, respectively. Thus, temporal mismatch contributed to the absolute errors. However, RMSE and MAE remained 45.90–66.66 $\mathrm{Mg\,ha^{-1}}$ and 33.24–52.02 $\mathrm{Mg\,ha^{-1}}$, respectively, when estimates representing the NFI period were used.

These analyses indicate that temporal mismatch can explain the observed differences to some extent, but does not fully account for them. Accordingly, we now report the fixed-support annual and 2015–2018 median sensitivity results in Section 4.1 of the main manuscript, with Supplementary Fig. S5 providing the supporting annual RMSE and MAE trajectories. We also expanded Section 5.1 to discuss the influence of temporal mismatch on the cross-product comparison.

Changes in manuscript: *Section 4.1, "Evaluation with NFI data" (pages 16–17, lines 382–387) [ADDED]*

To isolate the influence of product year from changes in data availability, the temporal-sensitivity analysis evaluated annual and 2015–2018 median estimates using a fixed product-specific set of NFI plots (Fig. S5). Using the 2015–2018 per-plot median estimates reduced RMSE from 50.95 to 45.90 $\mathrm{Mg\,ha^{-1}}$ *for Yan Map, from 71.09 to 64.54* $\mathrm{Mg\,ha^{-1}}$ *for ESA CCI, and from 73.68 to 66.66* $\mathrm{Mg\,ha^{-1}}$ *for Cai Map; the corresponding MAE reductions were 3.99, 5.77, and 5.52* $\mathrm{Mg\,ha^{-1}}$*, respectively. Temporal mismatch therefore affected the absolute evaluation metrics but did not eliminate the broad differences among the products.*

Section 5.1, "Comparison with other AGB datasets" (page 24, lines 521–524) [ADDED]

The temporal difference between the NFI observations and the evaluated products represents a source of uncertainty in the plot-level comparison. The additional sensitivity analysis showed that using product estimates corresponding more closely to the NFI observation period improved their agreement with the NFI observations, but did not eliminate the differences among the products.

Supplementary Fig. S5 (page 7) [ADDED]

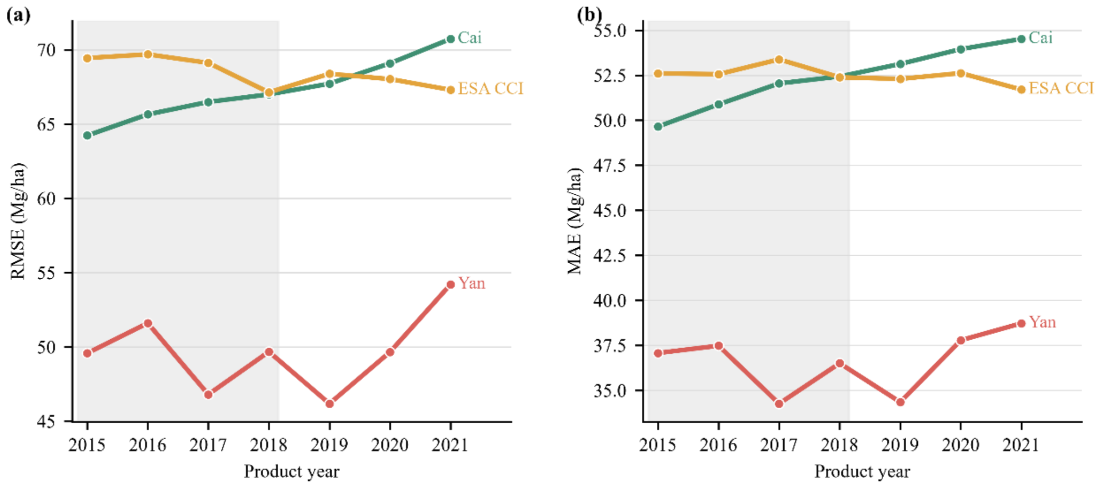


Figure S5. Sensitivity of the cross-product evaluation to product year. Annual (a) RMSE and (b) MAE from 2015 to 2021. For each product, all annual estimates were evaluated using the same product-specific set of NFI test plots with valid estimates throughout the period. The grey bands indicate the common 2015–2018 interval overlapping most of the NFI observation period. Although both metrics varied among product years, the broad differences among the products persisted throughout 2015–2021.

Comment 7

Fig 5: it is difficult to judge whether the difference between products should be attributed to model performance or temporal gap.

Response: Thank you for this important comment. We understand that this concern relates primarily to Fig. 5c. The purpose of this panel was to compare how the evaluated AGB datasets respond to the biomass gradient represented by the NFI AGB intervals, rather than to attribute the differences directly to a single source.

The comparison datasets exhibited different responses across the NFI AGB intervals. The average changes in median estimates between adjacent intervals were 9.5 $\mathrm{Mg\,ha^{-1}}$ for Yang, 13.3 $\mathrm{Mg\,ha^{-1}}$ for Yan, and 4.5 $\mathrm{Mg\,ha^{-1}}$ for Cai, which were markedly smaller than that of FineKarstAGB (32.1 $\mathrm{Mg\,ha^{-1}}$), indicating weaker responses to the NFI-observed biomass gradient. Yan and Cai also showed decreases between individual adjacent intervals. ESA CCI exhibited a comparatively stronger interval response (25.1 $\mathrm{Mg\,ha^{-1}}$), but its interquartile ranges remained wide across all intervals (85.5–101.5 $\mathrm{Mg\,ha^{-1}}$), indicating substantial dispersion within the same NFI AGB classes.

Temporal mismatch may affect the absolute agreement and increase the dispersion of product estimates. However, as described in our response to Comment 6, temporal alignment improved the agreement of the annual products with the NFI observations but did not eliminate the differences among them (Supplementary Fig. S5).

Taken together, we consider that temporal mismatch can explain only a limited part of the contrasting gradient patterns shown in Fig. 5c. Although temporal mismatch may affect the absolute agreement and within-interval dispersion, it is unlikely to account for the markedly different responses of the comparison datasets across the NFI AGB intervals. We have clarified this interpretation in the revised Results and Discussion.

Changes in manuscript: *Section 4.1, "Evaluation with NFI data" (pages 14–15, lines 365–369) [ADDED]*
The average changes in median estimates between adjacent NFI AGB intervals were 9.5 $\mathrm{Mg\,ha^{-1}}$ *for Yang Map, 13.3* $\mathrm{Mg\,ha^{-1}}$ *for Yan Map, and 4.5* $\mathrm{Mg\,ha^{-1}}$ *for Cai Map, compared with 32.1* $\mathrm{Mg\,ha^{-1}}$ *for FineKarstAGB. ESA CCI showed a comparatively stronger interval response of 25.1* $\mathrm{Mg\,ha^{-1}}$*, but with wide interquartile ranges across the intervals (85.5–101.5* $\mathrm{Mg\,ha^{-1}}$*).*

Comment 8

Fig. 6: the low performance of other AGB datasets could be caused by temporal mismatch with the NFI plots.

Response: Thank you for this comment. We agree that temporal mismatch can affect the absolute agreement between the external AGB products and the NFI observations. As described in our response to Comment 6, we therefore evaluated the annual estimates and the 2015–2018 per-plot median estimates for Yan, ESA CCI, and Cai.

Within the fixed product-specific supports used for the sensitivity analysis, the 2015–2018 median-based RMSE values were lower than the corresponding product-year values by approximately 5.05 $\mathrm{Mg\,ha^{-1}}$ for Yan, 6.55 $\mathrm{Mg\,ha^{-1}}$ for ESA CCI, and 7.03 $\mathrm{Mg\,ha^{-1}}$ for Cai. Nevertheless, RMSE and MAE still ranged from 45.90 to 66.66 $\mathrm{Mg\,ha^{-1}}$ and from 33.24 to 52.02 $\mathrm{Mg\,ha^{-1}}$, respectively, when estimates representing the NFI period were used. The annual RMSE and MAE results also showed that the broad differences among the products persisted throughout 2015–2021. These results confirm that temporal mismatch contributes to the absolute errors, but its influence on the lower agreement shown in Fig. 6 is limited. The corresponding manuscript changes are reported under our response to Comment 6.

Comment 9

Line 263: how the uncertainty was predicted should be explained. I could not find the uncertainty map from the given link.

Response: Thank you for this important comment. We agree that, although the original manuscript introduced the heteroscedastic uncertainty loss, it did not explain sufficiently how the model-estimated uncertainty was converted into the spatial uncertainty layer. In addition, the uncertainty layer was not included in the repository version available during the review.

We have therefore further clarified how the uncertainty layer is calculated from the model output. For each pixel, the regression branch outputs the AGB estimate, whereas the uncertainty branch outputs the corresponding input-dependent log variance. The corresponding standard deviation is obtained from the predicted log variance, and the lower and upper bounds of an approximately 95% prediction interval are then defined under an approximately Gaussian error distribution as follows:

$$\widehat{\sigma}_i = \exp\left(\frac{1}{2}\log\widehat{\sigma_i^2}\right),$$
$$q_{0.025,i} = \widehat{y}_i - 1.96\widehat{\sigma}_i, \qquad q_{0.975,i} = \widehat{y}_i + 1.96\widehat{\sigma}_i,$$
$$U_i = \widehat{\sigma}_i = \frac{q_{0.975,i} - q_{0.025,i}}{2 \times 1.96} \approx \frac{q_{0.975,i} - q_{0.025,i}}{4}.$$

In Eq. (7), the exponential factor multiplying the squared residual is the inverse of the predicted variance, whereas the additive log-variance term prevents the variance from increasing without penalty. Together, these terms enable the model to estimate the conditional variance of the prediction error for each input. The square root of this variance, i.e. the predicted standard deviation, is then used to define the 95% prediction interval and is reported as pixel-level predictive uncertainty. The conversion of a 95% interval width to an equivalent standard deviation is consistent with Cai et al. (2025).

We have now added a spatial uncertainty map and the corresponding statistical analysis immediately after the FineKarstAGB map in the Results section (Fig. 11). Across the study area, the mean and median uncertainties were 9.27 and 8.18 $\mathrm{Mg\,ha^{-1}}$, respectively, and 92.36% of pixels were concentrated between 5 and 15 $\mathrm{Mg\,ha^{-1}}$. We clarify that this layer represents model-estimated predictive uncertainty and does not encompass every possible source of uncertainty in the input datasets and reference observations.

Changes in manuscript: *Section 3.1.3, "Loss function" (pages 12–13, lines 305–321) [ADDED]*

For each pixel, the regression branch outputs the AGB estimate, whereas the uncertainty branch outputs the corresponding input-dependent log variance. The corresponding standard deviation and the lower and upper bounds of an approximately 95% prediction interval are calculated under an approximately Gaussian error distribution as follows:

$$\widehat{\sigma}_i = \exp\left(\frac{1}{2}\log\widehat{\sigma_i^2}\right),$$
$$q_{0.025,i} = \widehat{y}_i - 1.96\widehat{\sigma}_i, \qquad q_{0.975,i} = \widehat{y}_i + 1.96\widehat{\sigma}_i,$$
$$U_i = \widehat{\sigma}_i = \frac{q_{0.975,i} - q_{0.025,i}}{2 \times 1.96} \approx \frac{q_{0.975,i} - q_{0.025,i}}{4}.$$

In Eq. (7), the exponential factor multiplying the squared residual is the inverse of the predicted variance, whereas the additive log-variance term prevents unconstrained variance inflation. Together, these terms estimate the conditional prediction-error variance for each input patch. The square root of this variance defines the 95% prediction interval and represents the pixel-level predictive uncertainty, allowing uncertainty to vary with the spectral and structural characteristics of the input. The conversion from the 95%

interval width to an equivalent standard deviation follows Cai et al. (2025).

Section 4.3, “Spatial pattern of AGB in Southwest China and typical scenarios”, and Fig. 11 (page 20, lines 444–450; Fig. 11 on page 21) [ADDED]

The pixel-level predictive uncertainty of FineKarstAGB is presented in Fig. 11. Across the study area, the mean and median uncertainties were 9.27 and 8.18 $\mathrm{Mg\,ha^{-1}}$*, respectively, with values ranging from 0 to 79.49* $\mathrm{Mg\,ha^{-1}}$*. Over 90% were concentrated between 5 and 15* $\mathrm{Mg\,ha^{-1}}$*. Spatially, relatively high uncertainties occurred as localized patches and were more frequent in Yunnan, whereas low-to-moderate uncertainties dominated most of the study area. At the provincial scale, Yunnan had the highest mean and median uncertainties (9.87 and 8.54* $\mathrm{Mg\,ha^{-1}}$*), Guizhou had the lowest (8.53 and 7.50* $\mathrm{Mg\,ha^{-1}}$*), and Guangxi and Hunan showed intermediate values. These patterns indicate modest provincial differences, while most pixels exhibit relatively low-to-moderate predictive uncertainty.*

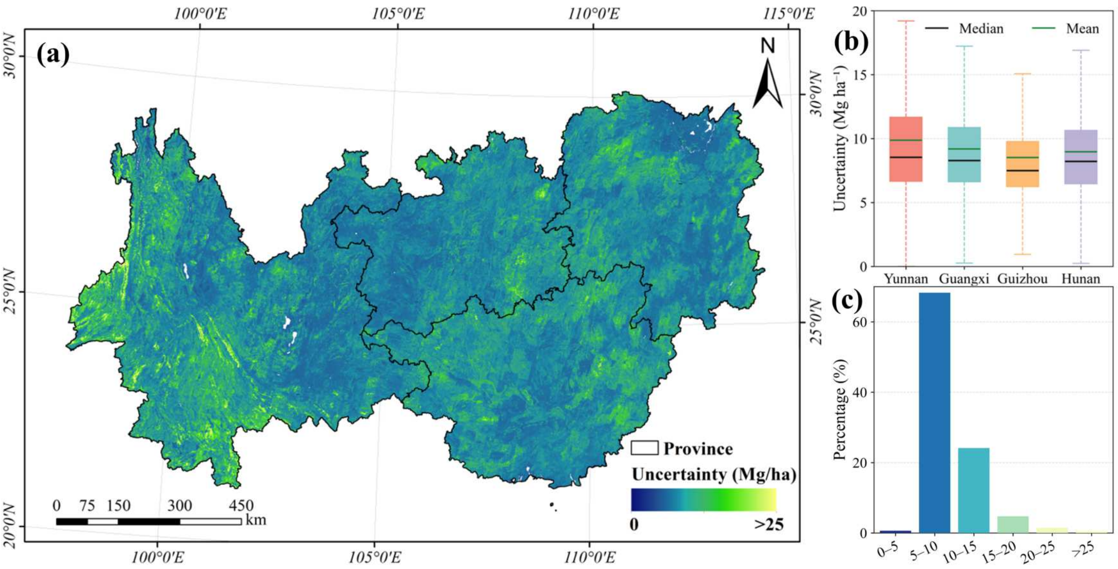


Figure 11. Spatial distribution and statistical characteristics of the predictive uncertainty of FineKarstAGB. (a) Spatial distribution of AGB uncertainty, (b) provincial uncertainty statistics, and (c) frequency distribution of uncertainty.

Authors' Response to Reviewer 2

General Comment

This study generated a dataset of aboveground biomass in southwest China using high-resolution satellite data. This study aims to use direct biophysical parameters, such as height and crown size to model biomass. It should be more reliable than other datasets using environmental variables, since this region has intense human activities which could affect the statistic relationship between AGB and environmental factors. I do agree with authors about this principle.

Response: We sincerely thank the reviewer for recognizing our work, particularly the use of forest structural attributes for AGB estimation in southwest China. This approach is especially relevant because frequent human activities may alter the relationships between AGB and environmental covariates, while coarse-resolution environmental variables may not adequately capture the region's pronounced spatial heterogeneity. We also appreciate the reviewer's valuable and constructive comments, which have helped us improve the scientific rigor and clarity of the manuscript. We respond to each comment below and describe the corresponding revisions made to the manuscript.

Comment 1

study area: the four provinces are not typically defined southwest China. To make it aligned with commonly used definition. The area should include Sichuan, Chongqing, and Tibet.

Response: Thank you for this comment. We agree that the conventional administrative definition of Southwest China, which generally includes Sichuan, Chongqing, Guizhou, Yunnan, and Tibet, is broader than the actual coverage of our dataset. The present dataset was not intended to represent this entire administrative region but focuses specifically on a contiguous four-province study area comprising Yunnan, Guizhou, Guangxi, and Hunan. These provinces contain extensive humid subtropical karst landscapes and encompass pronounced gradients in topography, vegetation, and human disturbance, providing an ecologically coherent setting for fine-scale AGB mapping. To avoid ambiguity, we have explicitly defined the geographic coverage in the Abstract and Study Area section. Accordingly, all analyses and conclusions are intended to refer specifically to this four-province study area rather than the complete administrative region of Southwest China.

Changes in manuscript: *Abstract (page 1, lines 26–27) [REVISED]*

Based on this approach, we generated a fine-grained 30 m AGB dataset (FineKarstAGB) for 2024, covering four provinces in Southwest China (Yunnan, Guizhou, Guangxi, and Hunan).

Section 2.1, "Study area" (page 4, lines 114–115) [REVISED]
The study area is located in Southwest China, encompassing the provinces of Yunnan, Guizhou, Guangxi and Hunan (20°–31°N, 97°–115°E, Fig. 2).

Comment 2

The influence of plot selection: authors used remote sensing indices (ndvi, crown cover-age, and canopy height) and a RF model to exclude plots which are far away from the model predictions. The remaining plots were finally used for the formal model training and validation. This approach would generate a biased and overestimated accuracy since the sample selection already used the information for model construction. The more scientific approach is to find the real reason of problematic plots and use data source which is independent to the model input to conduct plot selection.

Response: Thank you for raising this important concern. We agree that screening plots using prediction errors obtained from a model fitted to the same plots could introduce circularity and inflate the subsequent accuracy assessment. In our procedure, however, the RF screening was implemented through five-fold cross-validation. The prediction for each plot was therefore generated by an RF model trained on the other four folds, rather than by a model fitted to that plot. Moreover, the RF was used only as a preliminary quality-control tool based on plot-level summaries of mean NDVI, crown coverage, and mean canopy height. Its parameters and predictions were not transferred to the subsequent CSMF-Net. The purpose of this step was to identify pronounced inconsistencies between the historical plot AGB records and the forest conditions represented by the GaoFen-derived features, rather than to select plots according to CSMF-Net prediction performance.

To further examine the reliability of the retained plots using information independent of the CSMF-Net inputs, we conducted an additional temporal-consistency assessment based on repeated NFI8–NFI9 observations, Hansen forest-loss records, and persistent changes in Landsat NDVI, NBR, and NDMI. The annualized logarithmic AGB change between matched NFI8 and NFI9 plots was standardized within forest-origin and age groups using robust z-scores. A plot was classified as stable when its AGB change remained within the robust range of the corresponding forest group ($|z| < 2$), no Hansen forest loss was recorded during 2014–2024, no persistent decline was jointly detected by at least two of the three Landsat indices, and sufficient temporal observations were available. Plots with extreme AGB changes ($|z| > 3.5$) or independent evidence of forest disturbance were classified as anomalous, while the remaining plots were treated as uncertain.

Among the 2,235 retained plots that could be matched between NFI8 and NFI9, 1,852 plots (82.9%)

showed stable temporal trajectories. Although this assessment cannot demonstrate that every retained plot remained unchanged until the acquisition of the GaoFen imagery, it indicates that the retained dataset was predominantly composed of plots with consistent historical forest conditions. Together with the out-of-fold RF procedure, this result supports the use of the screening as a consistency-oriented quality-control step rather than a direct temporal correction of the NFI AGB values.

Accordingly, we have clarified in the revised Methods that the RF predictions were generated through five-fold cross-validation and that the screening was intended to identify substantial inconsistencies rather than to update the historical NFI AGB values to 2024. We have also expanded the Discussion to describe the additional temporal-consistency assessment and clarify the role and limitations of the screening procedure.

Changes in manuscript: *Section 2.3.1, "National Forest Inventory (NFI) data" (page 7, lines 178–195) [REVISED]*
The RF was used only as a preliminary quality-control tool; its parameters and predictions were not transferred to CSMF-Net. Its purpose was to identify pronounced inconsistencies between historical plot AGB and the forest conditions represented by the GaoFen-derived features, rather than to select plots according to CSMF-Net prediction performance or to update the historical NFI AGB values to 2024.
Section 5.3.1, "Input-data and plot-matching uncertainties" (page 27, lines 574–590) [ADDED]
In total, 1,852 matched plots (82.9%) showed stable temporal trajectories. This result indicates that the retained dataset was predominantly composed of plots with consistent historical forest conditions. Nevertheless, it does not demonstrate that every retained plot remained unchanged until the acquisition of the GaoFen imagery, and the screening should be interpreted as consistency-oriented quality control rather than a direct temporal correction of the historical NFI AGB values.

Comment 3

Uncertainty of crown segmentation: authors mentioned the minimal size of detected canopy is one pixel, but how reliable it is? How to solve overlapped crown problem in dense forests.

Response: Thank you for raising this important issue. We have clarified that 0.64 m^2, equivalent to one pixel in the original GaoFen imagery, represents the minimum mapping unit retained in the tree crown dataset. The independent assessment across the study area yielded a precision of 0.832, a recall of 0.853, and an F1 score of 0.843, providing quantitative evidence of the overall reliability of the predicted crown masks. Nevertheless, crowns close to the minimum mapping unit are more sensitive to image resolution, mixed pixels, and boundary uncertainty than larger and more clearly visible crowns.

To separate connected or overlapping crowns in dense forests, the initial segmentation was subjected to the post-processing procedure described by Xu et al. (2026). Potential crown centres were first detected within connected canopy regions, after which the associated pixels were reassigned according to their weighted distances from the identified centres. This procedure enables adjacent crowns to be separated when their centres and visible boundaries can be distinguished.

Nevertheless, crown overlap cannot always be completely resolved in very dense forests, particularly when adjacent crowns have indistinguishable boundaries or overlap extensively. Its influence on the subsequent AGB estimation may be partly reduced by the overall modelling framework. The NFI plots cover diverse forest types, topographic settings, and canopy conditions across the study area, while CSMF-Net integrates the crown segmentation with the spectral and textural information from GaoFen imagery and the vertical structural information provided by canopy height. The AGB estimates therefore do not depend exclusively on the accurate separation of every individual crown. However, these complementary inputs cannot resolve the crown-segmentation problem itself or entirely eliminate the associated uncertainty. We have clarified this limitation in the Discussion. Future work will investigate more advanced crown-deblending methods and the incorporation of three-dimensional canopy information to improve crown separation in dense forests.

Changes in manuscript: *Section 2.2.2, "Tree crown segmentation dataset" (page 6, lines 140–149) [REVISED]*

Tree crowns with a mapped area of at least 0.64 m^2, equivalent to one pixel in the original GaoFen imagery, were retained; this threshold represents the minimum mapping unit of the dataset.

Section 5.3.1, "Input-data and plot-matching uncertainties" (page 27, lines 566–573) [ADDED]

Tree-crown segmentation remains subject to uncertainty in dense forests. Although connected crowns are separated through crown-centre detection and weighted-distance reassignment (Xu et al., 2026), adjacent crowns may still be merged when their centres or visible boundaries cannot be reliably distinguished. The influence of these errors on AGB estimation may be partly reduced because the NFI plots represent diverse canopy conditions and CSMF-Net integrates crown segmentation with spectral and textural information from GaoFen imagery and vertical structural information from canopy height. Nevertheless, these complementary inputs cannot eliminate the underlying crown-segmentation uncertainty. Future work will investigate more advanced crown-deblending methods and three-dimensional canopy information.

Comment 4

Figure 5c: this figure cannot well show the distribution of NFI AGB in each bin, so we do not know the agreement between predicted AGB and NFI AGB. It is better to use violin plot

Response: Thank you for this helpful suggestion. Figure 5c is primarily intended to evaluate how the different AGB products respond to the gradient of observed NFI AGB. The NFI observations are used to define the five AGB classes, and the boxplots compare the distributions, medians, and dispersion of the corresponding product estimates within these classes. Point-by-point agreement with the NFI observations is evaluated more directly in Fig. 5a for FineKarstAGB and in Fig. 6 for the comparison products.

We carefully considered replacing the boxplots with violin plots. However, the NFI observations in the first four classes are restricted to predefined 50 $\mathrm{Mg\,ha^{-1}}$ intervals and are relatively concentrated. Their violin shapes would therefore provide limited additional information beyond the bin definitions. Moreover, the two edge classes contain relatively few observations ($n = 30$), making their kerneldensity shapes sensitive to bandwidth selection, whereas the open-ended $> 200~\mathrm{Mg\,ha^{-1}}$ class has a substantially broader range. Presenting these distributions together would result in strongly unbalanced violin shapes and reduce the clarity of the cross-product comparison. We have therefore retained the original boxplot presentation in Fig. 5c.

Comment 5

The contribution of three inputs should be quantitatively assessed.

Response: Thank you for this important suggestion. We agree that the respective contributions of the three model inputs should be quantitatively evaluated. We therefore conducted leave-one-input-out ablation experiments using the same training, validation, and independent test sets and identical training settings. The full model was compared with three two-input variants, and performance was evaluated using RMSE, MAE, and R^2.

Adding image to the tree crown and canopy height inputs reduced RMSE from 31.61 to 28.51 $\mathrm{Mg\,ha^{-1}}$ and MAE from 20.88 to 18.64 $\mathrm{Mg\,ha^{-1}}$, corresponding to reductions of 9.8% and 10.7%, respectively, while R^2 increased from 0.61 to 0.69. Adding tree crown information to image and canopy height reduced RMSE and MAE by 27.0% and 28.8%, respectively, and increased R^2 from 0.41 to 0.69. Adding canopy height to image and tree crown information produced the largest improvement, reducing RMSE and MAE by 41.2% and 42.7%, respectively, and increasing R^2 from 0.09 to 0.69.

The full model integrating all three inputs achieved the best performance, with an RMSE of 28.51 $\mathrm{Mg\,ha^{-1}}$, an MAE of 18.64 $\mathrm{Mg\,ha^{-1}}$, and an R^2 of 0.69. These results demonstrate that all three inputs provide complementary information. Canopy height provides the largest marginal contribution, tree crown information substantially improves performance when combined with vertical structure, and GaoFen imagery provides additional spectral and spatial information beyond the two structural inputs. We have added the ablation design to Section 3.2 and the quantitative results to new Section 4.1.2, together with new Table 2.

Changes in manuscript: *Section 3.2, "Accuracy assessment and cross-dataset comparison" (page 13, lines 329–331) [ADDED]*

To quantify the contributions of the three model inputs, leave-one-input-out ablation experiments were conducted under identical data partitions and training settings. Each two-input variant was evaluated against the full model using the independent NFI test set.

Section 4.1.2, "Contribution of model inputs" (page 15, lines 370–375), and Table 2 (page 16) [ADDED]

The input-ablation experiments showed that all three inputs contributed to model performance (Table 2). Relative to the corresponding two-input variants, adding GaoFen imagery, tree-crown information, and canopy height reduced RMSE by 9.8%, 27.0%, and 41.2%, respectively, and reduced MAE by 10.7%, 28.8%, and 42.7%, respectively. The full model achieved the highest accuracy (RMSE = 28.51 $\mathrm{Mg\,ha^{-1}}$, MAE = 18.64 $\mathrm{Mg\,ha^{-1}}$, and $R^2 = 0.69$).

Table 2. *Leave-one-input-out ablation results evaluated against the independent NFI test set.*

Model variant	*Image*	*Crown*	*Height*	R^2	*MAE*	*RMSE*
Image + crown	✓	✓	✗	*0.09*	*32.54*	*48.47*
Image + height	✓	✗	✓	*0.41*	*26.19*	*39.07*
Crown + height	✗	✓	✓	*0.61*	*20.88*	*31.61*
Full model	✓	✓	✓	*0.69*	*18.64*	*28.51*

Comment 6

The assessment of topographic effects should be done, since this region has many hills which affect the remote sensing observation. I would like to know how the accuracy changes with the slope.

Response: Thank you for this important comment. We agree that the complex terrain of the study area may affect remote-sensing observations and that the potential effects of both slope and aspect should be explicitly assessed. We therefore conducted an additional terrain-stratified residual analysis using the independent test plots. The test plots were grouped into seven slope classes and eight aspect classes, and the residual distributions and error metrics were evaluated for each terrain class. Because topographic robustness is an important component of model validation in this mountainous region, we have added the analysis as a new subsection (Section 4.1.3, “Topographic effects on model accuracy”) and included the corresponding figure in the main text (new Fig. 7), rather than placing it only in the Supplement.

The residual distributions in some higher-slope classes were slightly wider than those at low slopes, but neither the residual centers nor their spreads showed a consistent increasing trend with slope. Continuous slope was weakly related to absolute residuals (Pearson $r = 0.046$; Spearman ρ= 0.074). These results indicate that slope had only a limited effect on model errors within the independent test data.

Across the eight aspect classes, the residual distributions were generally centered near zero. Mean residuals ranged from -5.27 to 1.30 $\mathrm{Mg\,ha^{-1}}$ and median residuals ranged from -3.57 to 5.36 $\mathrm{Mg\,ha^{-1}}$. Although the dispersion varied somewhat among aspect classes, no consistent directional pattern of overestimation or underestimation was evident. Overall, the model showed some variation among individual terrain classes but no systematic deterioration in accuracy associated with slope or aspect.

Changes in manuscript: *Section 3.2, “Accuracy assessment and cross-dataset comparison” (page 14, lines 348–352) [ADDED]*

To assess topographic effects on AGB estimation, we examined variations in residual distributions across slope and aspect classes and calculated Pearson and Spearman correlations between continuous slope and absolute residuals to test whether prediction errors increased linearly or monotonically with slope.

Section 4.1.3, “Topographic effects on model accuracy”, and Fig. 7 (page 17, lines 388–394; Fig. 7 on page 17) [ADDED]

The residual distributions in some higher-slope classes were slightly wider than those at low slopes, but neither their centres nor their spreads increased consistently with slope (Fig. 7a). Continuous slope was only weakly related to absolute residuals (Pearson

r = 0.046; Spearman ρ = 0.074), indicating a limited slope-related effect on prediction errors. Residual distributions across the eight aspect classes were also generally centred near zero (Fig. 7b). Mean residuals ranged from -5.27 to 1.30 Mg ha^{-1} *and median residuals from -3.57 to 5.36* Mg ha^{-1}. *Although residual dispersion varied somewhat among aspect classes, no consistent directional pattern of overestimation or underestimation was observed.*

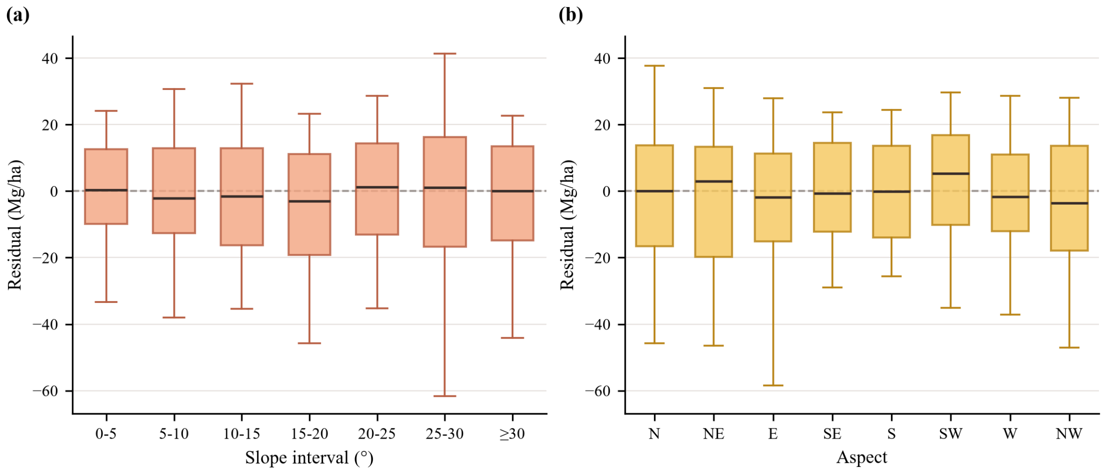


Figure 7. Terrain-stratified residual analysis of CSMF-Net AGB estimates using the independent NFI test set. (a) Residual distributions across slope classes. (b) Residual distributions across aspect classes.

Comment 7

Data problems after I checked the dataset, please see the attached pdf.

Response: Tile- and swath-like discontinuities. Thank you for the careful inspection of the dataset and for providing the specific examples. We agree that residual tile- and swath-like discontinuities are visible in some local areas of FineKarstAGB, including the locations identified in the comment.

Obtaining spatially and temporally consistent sub-meter imagery over such a large region is challenging. Compared with medium- and coarse-resolution sensors that provide broad-swath and frequently repeated coverage, sub-meter GaoFen imagery has a more limited spatial coverage per acquisition, requiring a large number of individual scenes to achieve wall-to-wall regional coverage. The image mosaic used in this study comprises 11,629 GF-2 and GF-7 scenes acquired during 2023–2024. In addition, the humid and mountainous environment of Southwest China is frequently affected by clouds, fog, and haze, which further reduces the availability of usable observations acquired under comparable conditions. Consequently, the final mosaic necessarily combines scenes acquired on different dates and under varying atmospheric, illumination, viewing-angle, and vegetation conditions.

Although geometric and radiometric corrections, image enhancement, cloud detection, mosaicking, and tiling were performed, these inter-scene differences could not always be completely removed. Because the AGB estimates were derived from the GaoFen imagery and its associated structural products, residual inconsistencies in the source mosaic may propagate into local AGB predictions and appear as scene- or orbit-related boundaries. Complete removal of these artifacts would require additional cloud-free overlapping acquisitions or a more temporally and radiometrically homogeneous image composite, which is not consistently available across the entire study area.

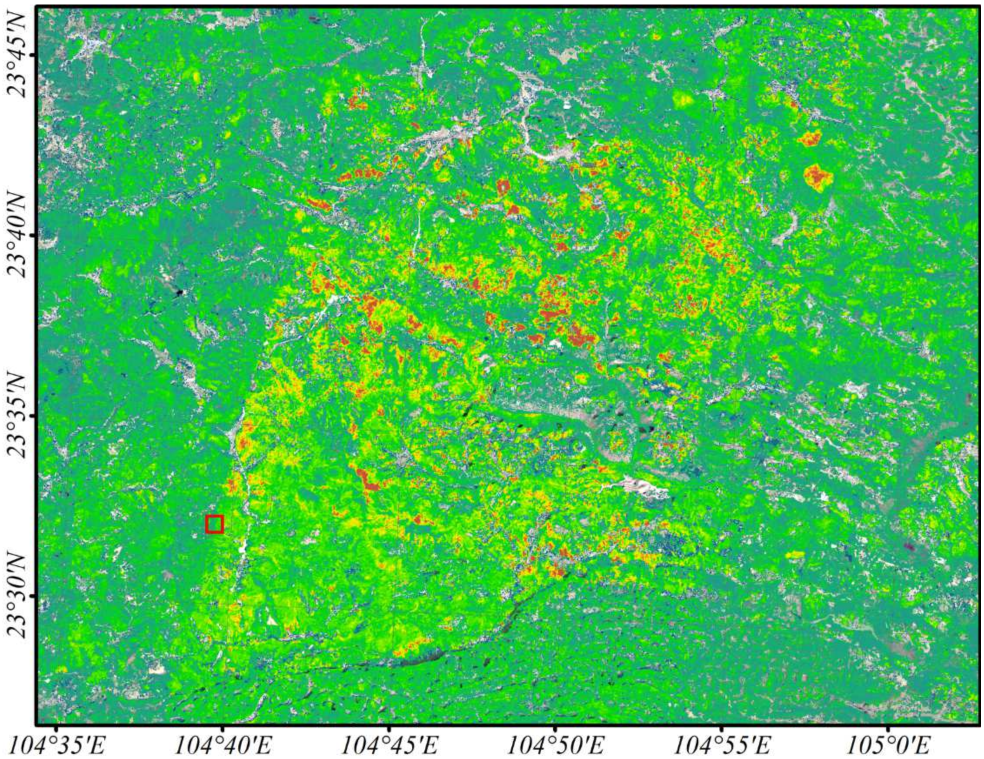
We therefore acknowledge this issue as a limitation of the current dataset. These discontinuities should not be interpreted as abrupt ecological changes, and caution is required when analysing local AGB variations across visible image boundaries. Nevertheless, the independent NFI and UAV LiDAR evaluations assess the overall predictive reliability of the dataset, while the identified artifacts mainly affect the interpretation of fine-scale spatial patterns in specific areas. We have added this limitation to Section 5.3.2 and clarified that wall-to-wall coverage does not imply complete radiometric consistency among all GaoFen scenes. Future versions will explore multi-temporal radiometric harmonization, improved blending of overlapping observations, and seamline optimization to reduce these artifacts.

Changes in manuscript: *Section 5.3.2, "Trade-off between VHR data acquisition and coverage" (pages 27–28, lines 591–613) [ADDED]*

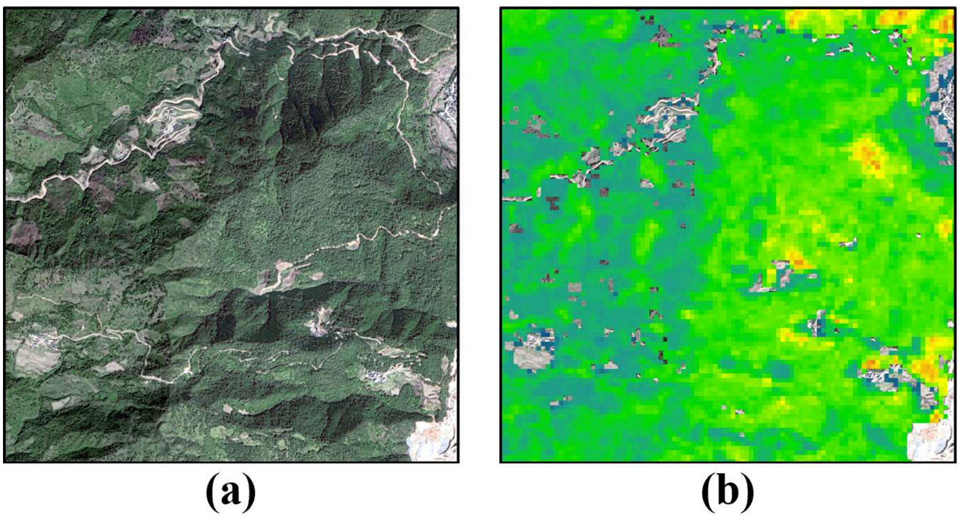
Obtaining spatially and temporally consistent sub-meter imagery over a large region remains challenging. Unlike medium- and coarse-resolution sensors that provide broad-swath and frequently repeated coverage, sub-meter GaoFen imagery requires a large number of individual scenes to achieve wall-to-wall coverage. This difficulty is further increased in the humid and mountainous environment of Southwest China, where frequent clouds, fog, and haze limit the availability of usable observations acquired under comparable conditions. Consequently, the GaoFen mosaic combines scenes acquired on different dates and under varying atmospheric, illumination, viewing-angle, and vegetation conditions. Despite geometric and radiometric corrections, image enhancement, cloud detection, and mosaicking, residual scene- or orbit-related discontinuities remain in some local areas and may propagate into the corresponding AGB estimates. These discontinuities should not be interpreted as abrupt ecological changes, and caution is required when analysing fine-scale AGB variations across visible image boundaries. Future versions of the dataset will explore multi-temporal radiometric harmonization, improved blending of overlapping observations, and seamline optimization to reduce these artifacts.

Locally extreme AGB values. Thank you for identifying the locally extreme AGB values and providing their geographic location. We carefully inspected the reported area and the corresponding GaoFen imagery. The inspection indicated that some of the extreme values were associated with residual radiometric differences among the mosaicked scenes. These differences increased the contrast between adjacent image regions and subsequently propagated into the local AGB predictions.

We therefore applied more detailed radiometric correction and colour balancing to the affected imagery and regenerated the corresponding local AGB estimates. The revised result reduces the artificial contrast and extreme values associated with the radiometric inconsistencies (Response Fig. 1). Response Fig. 2 provides an enlarged comparison of the area marked by the red box in Response Fig. 1. Comparison with the GaoFen imagery indicates that some local AGB variations remaining after radiometric harmonization correspond to differences in forest growth and stand condition visible in the imagery. These remaining variations may also be related to differences in image acquisition dates and should therefore not all be regarded as anomalous values. The corrected local results have been incorporated into the revised FineKarstAGB product.



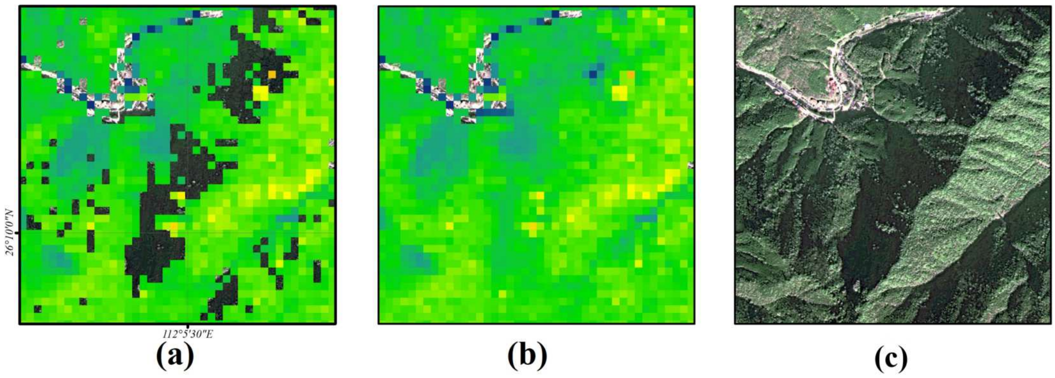
Response Figure 1. Revised FineKarstAGB estimates around the location identified in the comment after refined radiometric correction and colour balancing. The red box indicates the area enlarged in Response Fig. 2.



Response Figure 2. Enlarged comparison of the area marked by the red box in Response Fig. 1. The left panel shows the GaoFen imagery, and the right panel shows the revised FineKarstAGB estimates. The remaining local AGB variations generally correspond to differences in forest growth and stand condition visible in the imagery.

Terrain-shadow omissions. Thank you for identifying the missing values associated with terrain shadows. We carefully inspected the reported area and found that some tree-covered pixels affected by mountain shadows had been incorrectly excluded from the current product. We therefore combined the sub-meter GaoFen imagery with DEM information to identify these omitted areas and restored their corresponding AGB values using the original AGB prediction output. As shown in Response Fig. 3, the shadow-related gaps have been effectively filled while retaining the surrounding spatial pattern of

AGB. The corrected results have been incorporated into the revised FineKarstAGB product.



Response Figure 3. Correction of AGB values omitted because of terrain shadows. (a) Original FineKarstAGB result, (b) corrected result, and (c) corresponding GaoFen imagery. The omitted tree-covered pixels were identified using the GaoFen imagery and DEM information and restored using the original AGB prediction output.

Authors’ Response to Reviewer 3

General Comment

This study used the CSMF-Net to combine NFI, GF, tree crown segmentation, and canopy height to map 30m AGB in Southwest China. Then this model was validated and results were compared with the existing AGB datasets.

Response: Thank you for this concise and accurate summary and for your recognition of our work. We sincerely appreciate the reviewer’s constructive comments and suggestions. We have carefully considered all the comments and made corresponding revisions to improve the clarity and reliability of the manuscript. Detailed point-by-point responses and the associated revisions are provided below.

Comment 1

The overall accuracy of the existing tree canopy segmentation dataset is 0.82. However, when applying the corresponding method to other regions for use, the accuracy situation still needs to be reflected. The article does not mention this.

Response: Thank you for this comment. We agree that the original manuscript did not sufficiently describe the accuracy of the tree crown segmentation dataset across the full study area. We have therefore supplemented Section 2.2.2 with the existing region-wide accuracy assessment. The assessment was based on 919 independent image-interpretation plots covering the major forest types and topographic conditions across the study area. Within these plots, 9,061 crowns were manually delineated, and pixel-wise comparison of the binary crown masks yielded a precision of 0.832, a recall of 0.853, and an F1 score of 0.843. These results provide a more comprehensive evaluation of the tree crown segmentation dataset across its present application area.

Changes in manuscript: *Section 2.2.2, “Tree crown segmentation dataset” (page 6, lines 139–150) [REVISED]*

Tree-crown segmentation accuracy was assessed using 919 independent image-interpretation plots, separate from the NFI plots and covering the major forest types and topographic conditions across the study area. Within these plots, 9,061 crowns were manually delineated, and pixel-wise comparison of the binary crown masks yielded a precision of 0.832, a recall of 0.853, and an F1 score of 0.843. The $R^2 = 0.82$ *reported by Xu et al. (2026) refers to the plot-level agreement between labelled and predicted tree counts in Guangxi.*

Comment 2

When describing the use of GF data in the text, it is necessary to clearly specify which series of data is being referred to.

Response: We thank the reviewer for this helpful suggestion. The GF data used in this study refer specifically to two types of sub-meter GaoFen satellite imagery: GaoFen-2 (GF-2) and GaoFen-7 (GF-7). This information was originally provided in Section 2.2.1, where we described the number of GF-2 and GF-7 scenes, their spatial resolution, acquisition period, and preprocessing workflow. To make this point clearer and avoid ambiguity when "GF" or "GaoFen imagery" is mentioned elsewhere in the manuscript, we have refined the wording in Section 2.2.1 and revised related descriptions to explicitly refer to "GF-2/GF-7 GaoFen optical imagery" where appropriate.

Changes in manuscript: *Section 2.2.1, "Sub-meter GaoFen imagery" (page 5, lines 126–137) [REVISED]*
The sub-meter optical imagery used in this study was acquired over the study area by GaoFen-2 (GF-2) and GaoFen-7 (GF-7) of the GaoFen satellite constellation. The image collection includes 10,330 GF-2 scenes and 1,299 GF-7 scenes acquired during 2023–2024, with summer 2024 imagery used as the primary source and 2023 imagery used to fill remaining spatial gaps. The panchromatic and multispectral (PMS) sensors onboard GF-2 provide a 0.8 m panchromatic band and four 3.2 m multispectral bands, while GF-7, as a stereo-mapping satellite, offers a similar spectral configuration with a 0.8 m panchromatic band and four 3.2 m multispectral bands (Zhang et al., 2023b; Pan, 2015).

Comment 3

Section 2.2.3, "the errors are substantially lower than those reported for previously published canopy height products ($r \leq 0.22$)". What exactly is the canopy height product? It seems there are more than one product. Please clearly state the accuracy values for comparing them.

Response: Thank you for this helpful comment. We agree that the original statement did not clearly identify the canopy height products included in the comparison or provide their individual error metrics.

The statement referred to the resolution-matched comparison reported by Tong et al. (2025), in which the GaoFen-derived canopy height dataset used in this study was compared with the China forest canopy height product of Liu et al. (2022), the global canopy height product of Lang et al. (2023), and the global forest canopy height product of Potapov et al. (2021). In that comparison, the GaoFen-derived dataset achieved an MAE of 1.94 m and an RMSE of 2.76 m. The corresponding MAE and RMSE were 7.36 and 8.92 m for Liu et al. (2022), 5.47 and 6.98 m for Lang et al. (2023), and 6.27 and 8.63 m for Potapov et al. (2021), respectively.

To further address the reviewer's concern, we conducted an additional independent evaluation using UAV LiDAR-derived CHM data from LIDARNET. The UAV LiDAR CHM was aggregated to the native grid of each canopy height product to account for differences in spatial resolution. In this independent evaluation, the GaoFen-derived dataset achieved an MAE of 1.50 m and an RMSE of 2.82 m. The corresponding MAE and RMSE were 6.62 and 8.44 m for Liu et al. (2022), 8.37 and 9.30 m for Lang et al. (2023), and 4.13 and 5.88 m for Potapov et al. (2021), respectively. These additional results independently confirm that the GaoFen-derived canopy height dataset has lower absolute height errors than the three comparison products.

We have revised Section 2.2.3 to identify the comparison products and report the relevant MAE and RMSE values, and we have added the new independent validation results to the revised Fig. S2.

Changes in manuscript: *Section 2.2.3, "Canopy height dataset" (page 6, lines 151–161) [REVISED]*

These errors were substantially lower than those reported by Tong et al. (2025) for the canopy-height products of Liu et al. (2022b), Lang et al. (2023), and Potapov et al. (2021) (MAE ≥ 5.47 m; RMSE ≥ 6.98 m). Further accuracy assessment using independent UAV LiDAR data (described in Section 2.3.2) yielded an MAE of 1.50 m and an RMSE of 2.82 m for the GaoFen-derived canopy height dataset, while the other three products showed MAE ≥ 4.13 m and RMSE ≥ 5.88 m. The corresponding validation results are shown in Fig. S2.

Supplementary Fig. S2 [REVISED]

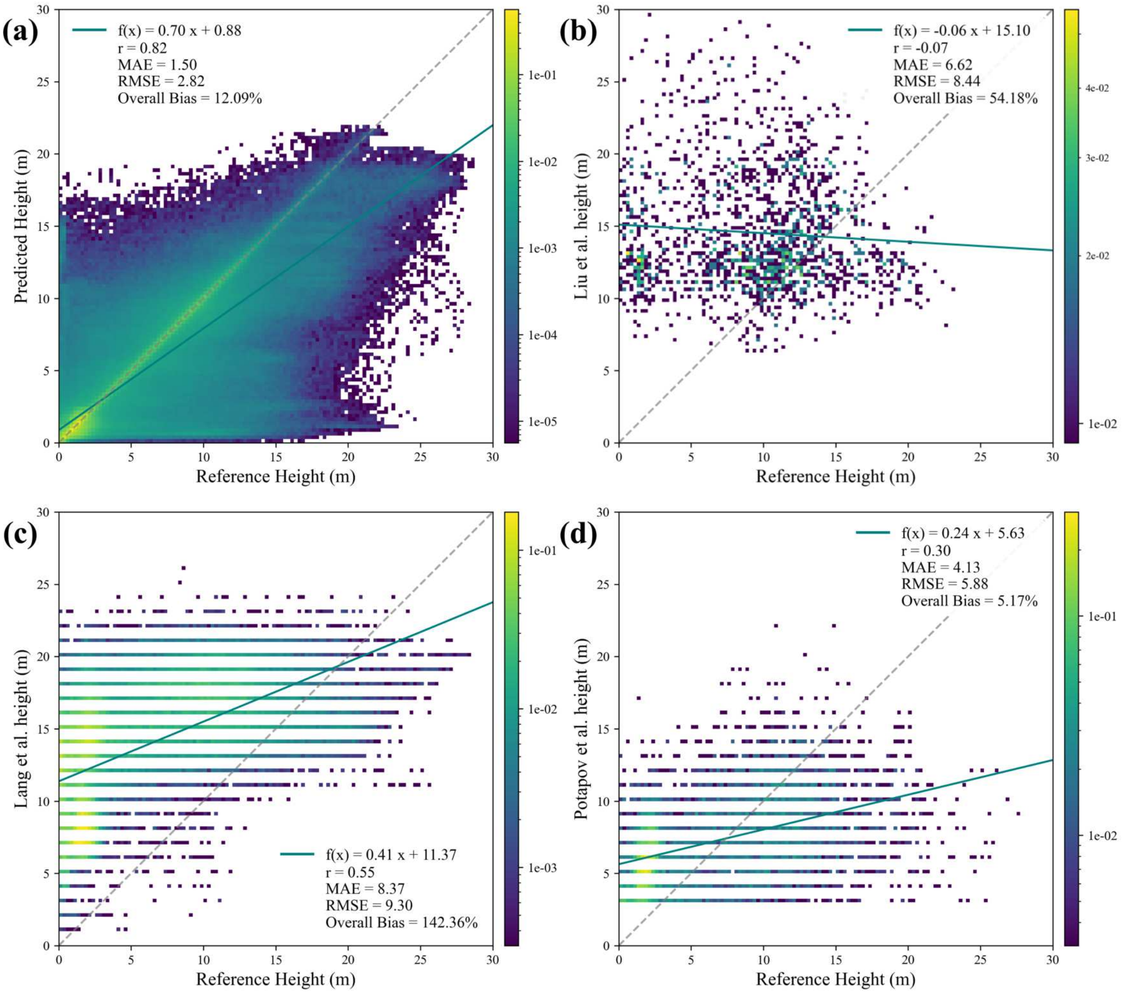


Figure S2. Comparison of canopy-height products with independent UAV LiDAR-derived CHMs from LIDARNET (http://lidar.pku.edu.cn). (a) GaoFen-derived canopy-height dataset used in this study. (b) National canopy-height product of Liu et al. (2022). (c) Global canopy-height product of Lang et al. (2023). (d) Global forest-canopy-height product of Potapov et al. (2021). The UAV LiDAR CHMs were aggregated to the native resolution of each product before evaluation.

Comment 4

Lines 179-181: How these spectral and structural variables (mean NDVI, crown coverage, ad canopy height) can be combined with AGB to reduce errors?

Response: Thank you for this comment. We agree that the original manuscript did not clearly explain how mean NDVI, crown coverage, mean canopy height, and NFI AGB were used in the sample-screening procedure. These three variables were not directly combined with AGB to correct the NFI measurements, nor were these summary statistics used as additional inputs to CSMF-Net. Instead, they were used only in a preliminary quality-control step to identify plots showing strong disagreement between the historical NFI observations and the forest conditions represented by the more recent remote-sensing data.

For each candidate NFI plot, mean NDVI was derived from the GaoFen imagery, crown coverage from the tree crown segmentation layer, and mean canopy height from the canopy height layer. These variables were used as predictors in a parsimonious Random Forest model, with the recorded NFI AGB as the response variable. To avoid screening plots using in-sample fitted values, the procedure was implemented using five-fold cross-validation. In each iteration, four folds were used to train the model and the remaining fold was used for prediction, ensuring that the screening estimate for each plot was generated by a model that had not used that plot for training.

Plots for which the out-of-fold prediction differed from the recorded NFI AGB by more than 50% were excluded from subsequent model development. The Random Forest predictions were used only for sample screening and did not replace or adjust the original NFI AGB values. Thus, this procedure reduced the influence of plots showing pronounced temporal or data inconsistencies. We have revised the corresponding text to clarify this procedure and also acknowledge that such screening cannot completely eliminate the uncertainty associated with the temporal mismatch.

Changes in manuscript: *Section 2.3.1, "National Forest Inventory (NFI) data" (page 7, lines 178–195) [REVISED]*

To identify gross inconsistencies between historical plot AGB and the spectral and structural conditions represented by the recent imagery, we applied an automated cross-source consistency screening workflow. First, an initial data-cleaning step removed plots with incomplete records, missing AGB values, zero or negative AGB, or zero or negative mean tree height. Second, three simple features strongly related to AGB were derived from the high-resolution remote-sensing data for the remaining plots: mean NDVI, crown coverage, and mean canopy height. These features represent the spectral condition, horizontal canopy occupancy, and vertical canopy structure of each plot, respectively.

Third, a parsimonious Random Forest model was constructed using these three variables as predictors and the recorded NFI AGB as the response. To avoid screening plots using in-sample fitted values, five-fold cross-validation was applied. In each iteration, the model was trained using four folds and applied to the remaining fold, ensuring that every plot received an out-of-fold prediction from a model that had not used that plot for training. Plots for which the out-of-fold prediction differed from the recorded NFI AGB by more than 50% were excluded from subsequent model development. The RF was used only as a preliminary quality-control tool; its parameters and predictions were not transferred to CSMF-Net. Its purpose was to identify pronounced inconsistencies between historical plot AGB *and the forest conditions represented by the GaoFen-derived features, rather than to select plots according to CSMF-Net prediction performance or to update the historical NFI AGB values to 2024. This procedure was used as sample-compatibility quality control rather than as a direct temporal correction; it removed plots exhibiting severe structural–label incompatibilities while maintaining the representativeness and diversity of the AGB distribution. However, it could not completely eliminate the uncertainty associated with the temporal mismatch.*

Comment 5

Fig. 3: The first part of the data, the canopy segmentation data, was copied from another algorithm. It is necessary to distinguish which parts are original and which ones are already available algorithms, and it is necessary to clearly define this.

Response: Thank you for this important comment. We agree that the origins of the tree crown segmentation and canopy height datasets should be more clearly distinguished from the work conducted in this study. Both datasets were produced by our co-author Xuan Yang independently of the present AGB study, using the same methodological frameworks as those described by Xu et al. (2026) and Tong et al. (2025), respectively. They are therefore introduced as input datasets in the Data section rather than as methods developed in this study.

The canopy height dataset has been described in the preprint of Tong et al. (2025), which is directly cited in the manuscript. The tree crown segmentation dataset has not been published separately; therefore, Xu et al. (2026) is cited to identify the methodological framework used to generate it. We have revised Fig. 3 and the corresponding descriptions to clarify the provenances and role of these two datasets.

The present study focuses on integrating GaoFen imagery, the two structural datasets, and NFI observations for regional AGB estimation. The work conducted in this study includes constructing the multi-feature plot-level samples; developing CSMF-Net, including its parallel feature-extraction branches, Multi-feature Cross-Attention module, regression head, and uncertainty estimation; applying the trained model across the study area to generate the 30 m FineKarstAGB dataset; and evaluating the resulting dataset. Therefore, the originality of this study lies in the multi-feature fusion and AGB estimation framework, its regional application, and the resulting FineKarstAGB dataset, rather than in the production of the two structural input datasets.

To make this distinction explicit, we have revised Fig. 3 by labeling the tree crown segmentation and canopy height datasets as input datasets produced independently of the present study, while identifying CSMF-Net and FineKarstAGB as the components developed and generated in this study, respectively. The figure caption has been revised accordingly.

Changes in manuscript: *Sections 2.2.2–2.2.3, "Tree crown segmentation dataset" and "Canopy height dataset" (page 6, lines 139–161) [REVISED]*

To characterize the highly heterogeneous patterns of tree distribution across the study area, this study used a tree-crown segmentation dataset derived from sub-meter GaoFen imagery and a U-Net model. The dataset originates from previous tree-mapping efforts (Xu et al., 2026). Using the same methodological framework, tree-crown segmentation was extended across the four-province study area, resulting in more than 21 billion mapped trees.

This study used a wall-to-wall canopy-height dataset with a spatial resolution of 0.8 m covering the entire study area. As described by Tong et al. (2025), the dataset was generated using a U-Net model with a ResNet-50 backbone, with sub-meter GaoFen imagery as input and airborne LiDAR-derived CHMs as reference data.

Fig. 3 caption (page 9) [REVISED]

Workflow of plot-level aboveground biomass estimation. The tree crown segmentation and canopy height layers are input datasets produced independently of the present AGB study using the methodological frameworks of Xu et al. (2026) and Tong et al. (2025), respectively; CSMF-Net and FineKarstAGB were developed and generated in the present study.

Comment 6

Section 3.1: The last sentence has clearly demonstrated the advantages of the CSMF-Net model in AGB prediction. Therefore, there is no need to conduct any verification regarding the purpose of this study. Overall, the research approach is clear and there is a certain degree of innovation. However, many parts of the text are not clearly explained, which affects the reliability of the results.

Response: Thank you for recognizing the advantages of CSMF-Net for AGB prediction, as well as the clarity and innovation of the overall research approach. We agree that several descriptions in the original manuscript were not sufficiently clear, which could make it difficult for readers to assess the reliability and reproducibility of the study. To address this concern in Section 3.1, we revised the opening paragraph to explain more directly the roles of the three inputs and the overall information flow through CSMF-Net. The revised paragraph clarifies that GaoFen imagery provides spectral and spatial context, tree crown segmentation represents horizontal canopy structure, and canopy height represents vertical forest structure. It also explains that these complementary features are extracted through separate branches, integrated by the Multi-feature Cross-Attention module, and subsequently used to estimate plot-level AGB and its associated uncertainty.

Considering that the reviewer's concern about unclear descriptions may also refer to the manuscript as a whole, we made additional targeted revisions in response to the specific comments provided below. In particular, we: (1) explicitly identified GF-2 and GF-7 as the GaoFen imagery used in this study and clarified their acquisition period and spatial coverage; (2) clarified the sources and roles of the tree crown segmentation and canopy height datasets and distinguished these input datasets from the methodological contribution of the present study in Fig. 3; (3) explained how mean NDVI, crown coverage, and mean canopy height were used in the NFI sample-screening procedure; (4) identified the existing canopy height products used for comparison and reported their corresponding accuracy values; (5) supplemented the amount, acquisition time, spatial coverage, and spatial distribution of the NFI and UAV LiDAR data; (6) explicitly stated that FineKarstAGB represents AGB conditions in 2024. We also revised the Discussion to organize the cross-dataset comparisons around more clearly defined and comparable aspects. Detailed modifications are described in our responses to the corresponding comments below. Together, these revisions improve the transparency of the data, methods, validation procedures, and interpretation of the results.

> ***Changes in manuscript:*** *Section 3.1, "Canopy Structure-driven Multi-feature Fusion Network" (page 10, lines 244–262) [REVISED]*
> *As illustrated in Fig. 4, this study proposes a Canopy Structure-driven Multi-feature Fusion Network (CSMF-Net) for plot-level AGB regression. The model integrates three spatially aligned inputs with complementary information: sub-meter GaoFen imagery provides the spectral and spatial context of the plot, tree crown segmentation characterizes the horizontal distribution of the canopy, and canopy height represents its vertical structure. These inputs are processed through separate feature-extraction branches and subsequently integrated by the Multi-feature Cross-Attention module, which enables the imagery features to interact with the horizontal and vertical structural features. The fused representation is then used by the regression head to predict a single plot-level AGB value and its associated uncertainty.*

> **Comment 7**
>
> How much NFI survey data and LIDAR data were used? There needs to be a space to display or mark the data. Some schematic diagrams are shown in Figure 3. But were only these covered location data actually used? The text needs to be clear about this and provide spatial display.

Response: Thank you for this suggestion. We agree that the amounts and actual spatial distributions of the NFI and UAV LiDAR data should be distinguished more clearly from the schematic workflow shown in Fig. 3. We would like to clarify that Fig. 3 presents only the methodological workflow; the graphical elements in this figure do not represent the actual sampling locations or spatial coverage of the reference data.

After data screening, a total of 3,637 NFI plots were used in this study, including 2,545 plots for model training, 364 for validation, and 728 for independent testing. Their spatial distribution is shown in Fig. S3. As detailed in our responses to Comments 3 and 4, the UAV LiDAR data comprise 18 datasets acquired in 2020, covering approximately 1,208 ha in Guangxi. All 18 datasets were used both to independently validate the canopy height dataset and to evaluate the AGB estimates after aggregation to the AGB grid.

To provide a clear spatial display, we have added Fig. S4, showing the distribution of all 18 UAV LiDAR datasets across Guangxi Province, together with GaoFen imagery and the corresponding UAV

LiDAR-derived CHM for a representative site. We have also revised Section 2.3.2 to clarify the sample numbers and the distinction between the schematic workflow and the actual data distributions.

Changes in manuscript: *Section 2.3.1, "National Forest Inventory (NFI) data" (page 7, lines 178–195) [REVISED]*

The final set of selected plots (n = 3,637) was used for model training (n = 2,545), validation (n = 364), and testing (n = 728), and their distribution is shown in Fig. S3.

Section 2.3.2, "UAV LiDAR data" (pages 7–8, lines 197–206) [REVISED]

The evaluation CHM data used in this study were obtained from LIDARNET (http://lidar.pku.edu.cn) and covered typical forest scenes in Guangxi Province. The data comprise 18 UAV LiDAR datasets acquired in 2020, covering approximately 1,208 ha across Guilin ($n = 10$), Fangchenggang ($n = 6$), and Wuzhou ($n = 2$). The LiDAR point clouds were acquired using Riegl VUX-1, with an average point-cloud density exceeding 10 $\mathrm{pts\,m^{-2}}$. The resulting CHMs have a spatial resolution of 0.05 m. Their spatial distribution and a representative example are shown in Fig. S4.

Supplementary Fig. S4 [ADDED]

Comment 8

What is the specific situation of the GF data? The article mentions some numbers of the data, but is it true that the distribution is across all regions? How are the areas without data distribution handled?

Response: Thank you for this important comment. We agree that the original manuscript did not explain the spatial coverage of the GaoFen imagery clearly enough. The GF-2/GF-7 scenes were not used as isolated image footprints; instead, they were processed with cloud detection, mosaicked, and tiled to generate a continuous GaoFen image mosaic covering the entire four-province study area. Specifically, summer 2024 imagery was used as the primary source, and 2023 imagery was used to supplement remaining spatial gaps. After preprocessing and mosaicking, the final processed GaoFen imagery consisted of 13,648 image tiles distributed across Yunnan, Guizhou, Guangxi, and Hunan. This wall-to-wall coverage is shown in Fig. S1, and the provincial tile statistics are reported in Table S1.

Changes in manuscript: *Section 2.2.1, "Sub-meter GaoFen imagery" (page 5, lines 126–137) [REVISED]*

The image collection includes 10,330 GF-2 scenes and 1,299 GF-7 scenes acquired during 2023–2024, with summer 2024 imagery used as the primary source and 2023 imagery used to fill remaining spatial gaps. ... All raw images were processed using a comprehensive workflow, including dehazing and enhancement, geometric and radiometric correction, pan-sharpening, cloud detection, mosaicking, and tiling, to generate a continuous GaoFen image mosaic covering the entire four-province study area. The processed imagery was organized into tiles based on a hierarchical grid system analogous to the Google Maps zoom-level framework. The final processed imagery consists of 13,648 image tiles distributed across Yunnan, Guizhou, Guangxi, and Hunan, with the spatial mosaic shown in Fig. S1 and provincial tile statistics summarized in Table S1.

Supplementary Fig. S1 caption [REVISED]

Figure S1. Spatial mosaic of the processed GF-2/GF-7 GaoFen imagery covering the four-province study area. The mosaic contains 13,648 image tiles and is displayed using the near-infrared, red, and green bands.

Comment 9

The text mentions that the AGB data from this study needs to be compared with many existing biomass data. It is suggested to pay attention to whether the definitions of forests in different data sets are consistent. If the definitions differ greatly, then comparing the AGB values with them would be of little significance, as the differences in definitions and spatial distribution would lead to significant discrepancies in the results. It is necessary to check whether the definitions of each data for the forest are consistent.

Response: Thank you for this important comment. We agree that differences in forest definitions and spatial masks can affect comparisons among AGB products. This issue was taken into account when designing the comparative evaluation and interpreting the resulting differences. We also re-examined the original publications and product documentation to verify the spatial-domain definitions of the datasets. The definitions are not fully consistent: Yang Map applies the forest class of GlobeLand30 2020; Yan Map follows the FAO Global Forest Resources Assessment definition; ESA CCI is not explicitly masked to forest areas; and Cai Map defines forest using a tree-cover threshold of 20%. FineKarstAGB is not restricted by a conventional forest mask and additionally represents biomass in tree-covered areas outside contiguous forests, including agroforestry systems, urban trees, and scattered farmland trees.

These differences were explicitly considered in the comparative analyses. First, the evaluation against NFI data was conducted exclusively using forest inventory plots. For each AGB product, only NFI plots with valid corresponding product values were retained; locations outside the valid coverage of a product were excluded rather than treated as zero AGB or prediction errors. Second, the pixel-wise differences shown in Fig. 13 were calculated only over areas where both FineKarstAGB and the corresponding comparison product contained valid AGB estimates. Third, the comparisons in Figs. 10, 12–14 were designed to examine differences in spatial coverage, spatial patterns, and numerical distributions. No-data gaps caused by product-specific forest masks were therefore interpreted as differences in product coverage rather than AGB estimation errors.

Therefore, although the products do not use identical forest definitions, the NFI-based evaluation and the pixel-wise comparisons over overlapping valid areas remain meaningful. Nevertheless, forest definitions and spatial masks are important sources of differences in product coverage and spatial distribution and should be considered when interpreting the comparison results. We have clarified these comparison principles in the revised Fig. 13 caption and Discussion.

Changes in manuscript: *Section 5.1, "Comparison with other AGB datasets" (pages 23–24, lines 511–520 and 531–534) [REVISED]*

Beyond these product-specific factors, discrepancies between FineKarstAGB and the comparison datasets may also arise from differences in acquisition year, spatial resolution, input data, and modeling strategy. FineKarstAGB represents 2024 conditions, whereas Yang Map, Yan Map, ESA CCI, and Cai Map correspond to 2019, 2021, 2022, and 2023, respectively. The comparison products therefore do not represent exactly the same forest conditions, including changes caused by growth, harvesting, or recovery. This temporal difference is particularly important in regions where human disturbance can drive rapid changes in forest landscapes. FineKarstAGB derives its 30 m estimates from 0.8 m imagery and structural layers, whereas the comparison products use different combinations of medium- or coarse-resolution imagery, spaceborne lidar observations, and interpolated structural variables. These differences in input scale and information content can lead to contrasting representations of heterogeneous landscapes. Fine-resolution inputs are particularly relevant for representing fragmented scenes such as agroforestry systems, forest–shrub mosaics, urban green spaces, and scattered trees within 30 m pixels.

In addition, the forest mask applied by Yan et al. (2023) produces visible no-data gaps. These omissions suggest that forest masks that are temporally mismatched with the AGB reference year or contain classification errors may exclude areas that currently contain tree biomass from existing AGB maps.

Fig. 13 caption (page 25) [REVISED]

Spatial distribution and statistics of AGB differences between FineKarstAGB and the comparison datasets: (a) Yang et al. (2023), (b) Yan et al. (2023), (c) ESA CCI, and (d) Cai et al. (2024). Pairwise differences were calculated after spatial alignment using only pixels with valid estimates in both products. FineKarstAGB was aggregated to the native 100 m grid for comparison with ESA CCI. μ and σ denote the mean and standard deviation of the AGB differences, respectively.

Comment 10

Clearly state which year's AGB data were obtained in this study.

Response: Thank you for this comment. We apologize that the reference year of the AGB dataset was not stated sufficiently clearly in the original manuscript. FineKarstAGB represents AGB for 2024. It was generated primarily using GF-2 and GF-7 imagery acquired during the summer of 2024. We have now explicitly stated the 2024 reference year in the Abstract and the caption of Fig. 10.

Changes in manuscript: *Abstract (page 1, lines 26–27) [REVISED]*

Based on this approach, we generated a fine-grained 30 m AGB dataset (FineKarstAGB) for 2024, covering four provinces in Southwest China (Yunnan, Guizhou, Guangxi, and Hunan).

Fig. 10 caption (page 20) [REVISED]

Spatial pattern and statistical characteristics of FineKarstAGB for 2024. (a) Spatial distribution of AGB, (b) boxplots of provincial AGB values for the four provinces, and (c) pixel-based frequency distribution of AGB across the study area.

Comment 11

The discussion section is not focused enough. The first and second parts are all comparing with other AGB data, but no clear discussion points have been identified. For instance, the algorithms of these products are different, and the definitions of forests in different data are also different. It is necessary to identify the comparable points.

Response: Thank you for this important comment. We agree that the focus and internal division of the first two Discussion subsections were not sufficiently clear. Our original intention was for Section 5.1 to explain inter-product differences through a sequence of difference statistics, spatial difference patterns, product-specific causes, common causes, and evidence from representative regions, while Section 5.2 was intended to discuss the significance of fine-scale canopy vertical structure for characterizing AGB heterogeneity. However, the final paragraph of Section 5.2 compared the canopy-height inputs and production methods of Yang Map, Yan Map, and Cai Map. Although this information was intended to support the role of vertical structure, its placement introduced product-comparison material into Section 5.2 and weakened the distinction and logical flow between the two subsections.

To clarify the comparable points, we now state that the products were spatially aligned and that pixel-wise differences were calculated only over mutually valid pixels. Specifically, FineKarstAGB was aggregated to 100 m for comparison with ESA CCI. Differences in forest masks affect the spatial coverage of the products, but do not invalidate the pixel-wise comparison because masked or no-data pixels were excluded. The comparison therefore focuses on differences in AGB magnitude and spatial patterns, as well as the factors contributing to these differences.

We have accordingly reorganized the two subsections while retaining their original scientific arguments. Section 5.1 now follows the intended sequence more explicitly, and the discussion of canopy-height and other product inputs has been moved from Section 5.2 and integrated into the corresponding product-specific explanations in Section 5.1. Section 5.2 now focuses on the ecological and remote-sensing significance of fine-scale vertical structure and the supporting evidence from Figs. 12 and 14. In addition, after aggregating FineKarstAGB to 100 m for comparison with ESA CCI, the mean and standard deviation of the differences changed from 6.40 and 57.17 $\mathrm{Mg\,ha^{-1}}$ to 8.09 and 54.52 $\mathrm{Mg\,ha^{-1}}$, respectively; the overall spatial pattern and interpretation remain unchanged.

Changes in manuscript: *Section 5.1, "Comparison with other AGB datasets", with Figs. 13–14 (pages 22–24, lines 475–537; Figs. 13–14 on pages 25–26) [REVISED AND REORGANIZED]*

To systematically characterize the differences among the AGB products and identify their potential causes, comparisons were conducted at both the overall statistical and spatial-pattern levels. At the overall statistical level, pixel-wise comparisons indicate that our estimates are closer to those of Yang et al. (2023) and Yan et al. (2023), with mean differences of -14.66 $\mathrm{Mg\,ha^{-1}}$ *and -15.36* $\mathrm{Mg\,ha^{-1}}$*, respectively (Fig. 13). In contrast, although the comparison with ESA CCI yields the smallest mean difference (*μ *=*

8.09 $\mathrm{Mg\,ha^{-1}}$*), the substantially larger dispersion of differences (*σ *= 54.52* $\mathrm{Mg\,ha^{-1}}$*) indicates pronounced inconsistencies at local scales. Comparisons with the dataset of Cai et al. (2024) reveal a much larger overall bias, with a mean difference of -40.98* $\mathrm{Mg\,ha^{-1}}$*. It is important to note that these discrepancies do not manifest as spatially homogeneous or systematic biases across the study region, but instead exhibit pronounced spatial heterogeneity.*

The spatial patterns of the differences further highlight the influence of modeling strategies on the representation of AGB spatial structure. In the comparison with Yang et al. (2023) (Fig. 13a), the differences display a clear regional boundary, particularly evident in southwestern Yunnan, where FineKarstAGB tends to be lower than the estimates of Yang et al. to the west of the boundary and higher east of the boundary. This abrupt spatial transition is closely associated with the vegetation-zone-based modeling strategy adopted by Yang et al. A similar boundary pattern is also observed in the comparison with Yan et al. (2023) (Fig. 13b), notably in western Guizhou and northern Guangxi. These results suggest that while geographically or vegetation-stratified modeling approaches can improve internal coherence within individual zones, they may introduce discontinuities at zone boundaries. Such boundary effects are especially pronounced in the study area, where complex topography, steep ecological gradients, and highly heterogeneous vegetation amplify inconsistencies.

In addition to the effects of vegetation zonation, differences in structural inputs may also contribute to the patterns observed for the datasets of Yang et al. (2023) and Yan et al. (2023). Both products incorporated the national forest canopy height product of Liu et al. (2022), which was derived through interpolation of GEDI and ICESat-2 observations. Under complex terrain and canopy conditions, limitations in these observations may reduce the local representativeness of the resulting canopy-height information (Fu et al., 2025). The combined effects of zonation-based modeling and uncertainty in structural inputs therefore provide plausible explanations for the regional boundaries and local discrepancies described above.

The comparison between FineKarstAGB and ESA CCI shows pronounced differences (Fig. 13c). These differences may partly arise from the global modeling framework and coarser spatial resolution of ESA CCI. In mountainous and fragmented forest landscapes, spatial aggregation effects tend to smooth fine-scale AGB variability, limiting the ability of such products to capture ecologically meaningful AGB heterogeneity. Although no clear zonal boundaries are evident in the comparison with the dataset of Cai et al. (2024) (Fig. 13d), the overall magnitude of the bias is substantially large, with very high values in western and southern Yunnan. For the dataset of Cai et al. (2024), the large differences may partly reflect uncertainties in the GEDI reference data under complex terrain and the lack of fine-grained canopy-height input in its spatial prediction. GEDI measurements are susceptible to slope effects and footprint geolocation uncertainties under complex terrain, which can increase the uncertainty of AGB reference values (Tang et al., 2023). Its spatial prediction instead relied more strongly on remote-sensing texture, spectral, and tree-cover information. This difference in input information may contribute to the observed discrepancies, although it is worth noting that tree-cover information used in such approaches retains distinct advantages for long-term, large-scale monitoring of forest densification and degradation.

Beyond these product-specific factors, discrepancies between FineKarstAGB and the comparison datasets may also arise from differences in acquisition year, spatial resolution, input data, and modeling strategy. FineKarstAGB represents 2024 conditions, whereas Yang Map, Yan Map, ESA CCI, and Cai Map correspond to 2019, 2021, 2022, and 2023, respectively. The comparison products therefore do not represent exactly the same forest conditions, including changes caused by growth, harvesting, or recovery. This temporal difference is particularly important in regions where human disturbance can drive rapid changes in forest landscapes. FineKarstAGB derives its 30 m estimates from 0.8 m imagery and structural layers, whereas the comparison products use different combinations of medium- or coarse-resolution imagery, spaceborne lidar observations, and interpolated structural variables. These differences in input scale and information content can lead to contrasting representations of heterogeneous landscapes. Fine-resolution inputs are particularly relevant for representing fragmented scenes such as agroforestry systems, forest–shrub mosaics, urban green spaces, and scattered trees within 30 m pixels.

The comparison of representative regions with significant differences (Fig. 14) illustrates the previously discussed causes of discrepancies among AGB datasets. In both the representative areas and their zoomed-in views, the AGB estimates from this study exhibit spatial patterns that are consistent with canopy height and GaoFen imagery. In contrast, the results of Yang et al. (2023) and Yan et al. (2023) clearly reflect the influence of geographic and vegetation zonation. The Pu'er region in Yunnan is located within tropical monsoon forest–rainforest zones (Fig. 14d, e, Rows 1–2) and therefore shows generally higher AGB estimates, whereas the Tongren region in Guizhou lies within subtropical evergreen broadleaf forests (Fig. 14d, e, Rows 3–4) and is characterized by predominantly *moderate to high AGB values; differences among vegetation zones are thus amplified in these datasets. In addition, the forest mask applied by Yan et al. (2023) produces visible no-data gaps. These omissions suggest that forest masks that are temporally mismatched with the AGB reference year or contain classification errors may exclude areas that currently contain tree biomass from existing AGB maps. The discrepancies observed in the dataset of Cai et al. (2024) (Fig. 14g) are mainly reflected in areas with complex terrain and dense forests, consistent with the product-specific differences in reference data and input variables discussed above. Together, these representative regions provide local evidence for the causes inferred from the regional difference patterns.*

Section 5.2, "Significance of fine-scale vertical structure in characterizing AGB heterogeneity" (pages 24 and 26, lines 538–554) [REVISED AND REORGANIZED]

Taken together, the ecological relationship between canopy height and biomass and the spatial evidence in Figs. 12 and 14 indicate that fine-scale vertical structure complements spectral and crown features by providing direct information on vegetation stature. This information is particularly important for distinguishing local biomass contrasts associated with differences in canopy height within dense forests, fragmented agroforestry systems, forest–shrub mosaics, and scattered tree cover. Thus, canopy-height information supports not only AGB estimation but also the representation and interpretation of biomass heterogeneity across structurally complex landscapes.

Fig. 13 caption (page 25) [REVISED]

Spatial distribution and statistics of AGB differences between FineKarstAGB and the comparison datasets: (a) Yang et al. (2023), (b) Yan et al. (2023), (c) ESA CCI, and (d) Cai et al. (2024). Pairwise differences were calculated after spatial alignment using only pixels with valid estimates in both products. FineKarstAGB was aggregated to the native 100 m grid for comparison with ESA CCI. μ and σ denote the mean and standard deviation of the AGB differences, respectively.

Authors' Response to Community Comment 1

General Comment

It is an Innovative work. This study adopted 0.8 m sub-meter Gaofen-2/7 satellite imagery to characterize the highly fragmented and spatially heterogeneous karst landscapes of Southwest China, effectively resolving the mixed-pixel limitation of conventional medium-resolution remote sensing data. This work produced fine-scale individual tree crown segmentation and high-precision canopy height products using sub-meter GaoFen data. By integrating spectral information, horizontal crown structure, and vertical canopy height, the proposed CSMF-Net deep learning model substantially reduced biomass underestimation and saturation in dense forests. The FineKarstAGB dataset accurately quantified the biomass of scattered trees in agroforestry mosaics, rural landscapes and urban areas, which are ignored by existing coarse-resolution products.

But the most obvious limitation of this study lies in the severe temporal mismatch between reference data and remote sensing observations. The field NFI plots were collected during 2014–2018, while the GaoFen imagery and canopy height data were acquired around 2023–2024, creating a time gap of nearly 10 years. Southwest China's karst area is dominated by young and fast-growing forests under long-term ecological restoration. Forest stand structure and AGB have changed rapidly over the decade. As a result, the historical plot biomass values cannot truly reflect the forest actual conditions in 2024, which inevitably brings systematic bias and underestimation risks to AGB modelling. Although the authors attempted to filter inconsistent samples using high-resolution image features, this strategy can only reduce rather than fundamentally eliminate the temporal error. The study did not further quantify how much biomass was underestimated over the decade, making it difficult to fully evaluate the magnitude of uncertainty caused by time inconsistency.

Response: Thank you for recognizing the innovation and value of our work and for raising the important concern regarding the temporal mismatch between the NFI measurements and the GaoFen observations.

Strict plot-by-plot temporal matching is rarely attainable in regional wall-to-wall biomass mapping because forest inventories and satellite observations are generally acquired according to different multiyear schedules. For example, Yang et al. (2023) used field plots collected between 2011 and 2019 to produce an AGB map for 2019. Nevertheless, we agree that the lack of plot-specific timestamps means that the temporal mismatch in our study cannot be completely eliminated.

The random-forest procedure was used to screen gross inconsistencies between historical plot AGB and the NDVI, crown coverage, and canopy height observed in the recent imagery, rather than to update the NFI labels to 2024. We independently examined the temporal consistency of the retained plots using repeated NFI8–NFI9 observations. Stability was determined from the annualized logarithmic AGB change relative to plots of comparable stand origin and age, together with Hansen forest-loss records and persistent declines in Landsat NDVI, NBR, and NDMI during 2014–2024. Of the 2,235 matched retained plots, 1,852 plots (82.9%) were classified as stable, 243 plots (10.9%) as anomalous, and 140 plots as uncertain. The dominance of independently stable plots suggests that the consistency screening potentially favored samples with greater temporal stability, although it cannot establish that every retained plot remained unchanged until 2024.

We further quantified the possible influence of the remaining time gap using a growth-slowdown model that was independent of CSMF-Net and the GaoFen-derived predictors. Annual AGB change between NFI8 and NFI9 was modeled as a nonlinear function of initial AGB and stand age, with forest type, forest origin, province, elevation, and slope included as covariates. Cubic B-spline terms represented the nonlinear effects of initial AGB and age, ridge regularization controlled model complexity, and performance was assessed by five-fold spatial-block cross-validation. AGB and stand age were then updated recursively at annual steps, allowing the predicted increment to slow as biomass accumulated and stands aged. The estimated median potential AGB increases were 13.45, 17.00, and 20.23 $\mathrm{Mg\,ha^{-1}}$ for assumed intervals of 6, 8, and 10 years, respectively. These values quantify a plausible range of temporal uncertainty rather than measured plot-level changes, and were not used to update the NFI labels or the FineKarstAGB map. We have clarified the role of the screening procedure in Section 2.3.1, added the quantitative method and Fig. S8 to the Supplement, and revised the limitation discussion in Section 5.3.1.

Changes in manuscript: *Section 2.3.1, "National Forest Inventory (NFI) data" (page 7, lines 178–195) [REVISED]*

To identify gross inconsistencies between historical plot AGB and the spectral and structural conditions represented by the recent imagery, we applied an automated cross-source consistency screening workflow.

This procedure was used as sample-compatibility quality control rather than as a direct temporal correction; it removed plots exhibiting severe structural–label incompatibilities while maintaining the representativeness and diversity of the AGB distribution.

Section 5.3.1, "Input-data and plot-matching uncertainties" (page 27, lines 574–590) [ADDED]

To examine the historical stability of the retained plots using information independent of the CSMF-Net inputs, we additionally evaluated the 2,235 retained plots that could be matched between NFI8 and NFI9. The annualized logarithmic AGB change between repeated observations was standardized within forest-origin and age groups using robust z-scores. A plot was classified as stable when its AGB change remained within the robust range of its forest group ($|z| < 2$), no Hansen forest loss was recorded during 2014–2024, no persistent decline was jointly detected by at least two of Landsat NDVI, NBR, and NDMI, and sufficient temporal observations were available. Plots with extreme AGB changes ($|z| > 3.5$) or independent evidence of disturbance were classified as anomalous, while the remaining plots were treated as uncertain. In total, 1,852 matched plots (82.9%) showed stable temporal trajectories. This result indicates that the retained dataset was predominantly composed of plots with consistent historical forest conditions. Nevertheless, it does not demonstrate that every retained plot remained unchanged until the acquisition of the GaoFen imagery, and the screening should be interpreted as consistency-oriented quality control rather than a direct temporal correction of the historical NFI AGB values.

To further quantify the remaining temporal uncertainty, an independent nonlinear growth-slowdown analysis based on repeated NFI8–NFI9 observations estimated median potential AGB increases of 13.45–20.23 $\mathrm{Mg\,ha^{-1}}$ *over assumed intervals of 6–10 years (Fig. S8). These scenario-based estimates were not used to update the NFI labels or FineKarstAGB map, but indicate the possible magnitude of the remaining temporal uncertainty.*

Supplementary Text S1, "Growth-slowdown sensitivity analysis" (page 2), and Fig. S8 (page 10) [ADDED]

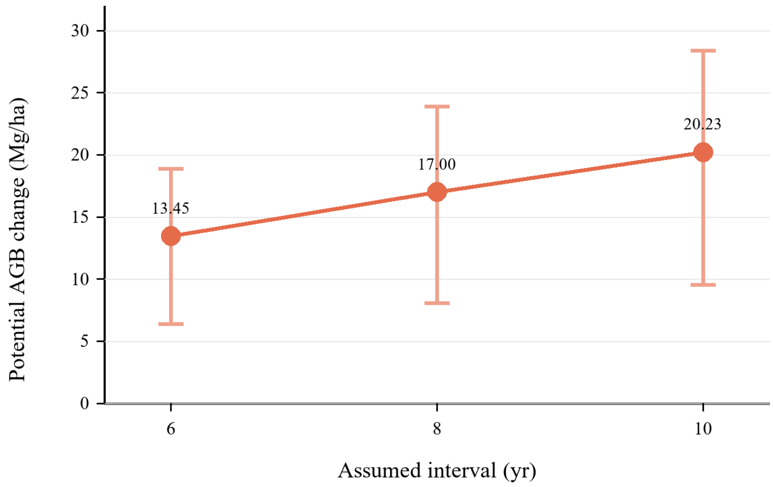


Figure S8. Potential median AGB changes estimated under assumed intervals of 6, 8, and 10 years; error bars indicate the interquartile range. These values are sensitivity scenarios rather than direct updates of the NFI labels.

Authors' Response to Community Comment 2

General Comment

Overall, this study develops a novel canopy structure-driven multi-feature fusion network and produces a 30 m resolution aboveground biomass dataset for the karst region of Southwest China. The research topic is highly valuable for regional carbon cycle assessment, ecological restoration evaluation and carbon neutrality targets. The data foundation is solid, the methodological framework is innovative, the accuracy evaluation system is comprehensive, and the results are reliable with clear scientific and application significance. The manuscript is well-structured and merits acceptance after moderate revision.

There are significant spatial distribution differences between our results and those of other products (Figure 3). What are the reasons for this?

Figure 7(a), few AGB values below 50. Is this reasonable?

Response: Thank you for your positive assessment of the scientific value, methodological innovation, data foundation, and potential applications of this study. We also appreciate your constructive comments. The two specific issues raised are addressed below.

Regarding the spatial differences shown in the revised Fig. 13, the causes of the differences between FineKarstAGB and the comparison products have been systematically discussed in Section 5.1. These differences arise from several factors, including the input data, reference data, modelling strategies, spatial resolutions, acquisition years, and forest masks. More specifically, geographically or vegetation-stratified modelling may introduce discontinuities at zone boundaries, whereas the coarser resolution and global modelling framework of ESA CCI tend to smooth local AGB variability in mountainous and fragmented landscapes. Differences involving the Cai product may also be associated with its reliance on GEDI observations, which can be affected by terrain and footprint geolocation uncertainty. In contrast, FineKarstAGB integrates sub-meter GaoFen imagery with tree crown and canopy height information, allowing it to retain more detailed variations in horizontal and vertical forest structure. We have further clarified these common and product-specific causes in Section 5.1 and interpreted the spatial differences together with the NFI and UAV LiDAR evaluations and representative landscape comparisons.

Regarding the relatively small number of AGB values below 50 $\mathrm{Mg\,ha^{-1}}$ in the revised Fig. 8a, this panel is based specifically on the forest subset of the independent UAV LiDAR validation data. Its purpose is to evaluate the structural relationship between forest AGB and canopy height. Of the 3,591 forest validation samples, 130 samples (3.62%) had AGB values below 50 $\mathrm{Mg\,ha^{-1}}$, with a minimum value of approximately 1.97 $\mathrm{Mg\,ha^{-1}}$. The relatively small number of low-AGB observations mainly reflects the composition of this forest validation subset. Samples outside contiguous forests were evaluated separately in Fig. 9, while the complete pixel-level AGB frequency distribution across the study area is presented in Fig. 10c. We have clarified the sampling domain and purpose of Fig. 8a in the Results and its caption.

Changes in manuscript: *Section 5.1, "Comparison with other AGB datasets" (pages 22–24, lines 475–537) [REVISED]*

Beyond these product-specific factors, discrepancies between FineKarstAGB and the comparison datasets may also arise from differences in acquisition year, spatial resolution, input data, and modeling strategy. FineKarstAGB represents 2024 conditions, whereas Yang Map, Yan Map, ESA CCI, and Cai Map correspond to 2019, 2021, 2022, and 2023, respectively. The comparison products therefore do not represent exactly the same forest conditions, including changes caused by growth, harvesting, or recovery. This temporal difference is particularly important in regions where human disturbance can drive rapid changes in forest landscapes. FineKarstAGB derives its 30 m estimates from 0.8 m imagery and structural layers, whereas the comparison products use different combinations of medium- or coarse-resolution imagery, spaceborne lidar observations, and interpolated structural variables. These differences in input scale and information content can lead to contrasting representations of heterogeneous landscapes. Fine-resolution inputs are particularly relevant for representing fragmented scenes such as agroforestry systems, forest–shrub mosaics, urban green spaces, and scattered trees within 30 m pixels.

The comparison of representative regions with significant differences (Fig. 14) illustrates the previously discussed causes of discrepancies among AGB datasets. In both the representative areas and their zoomed-in views, the AGB estimates from this study exhibit spatial patterns that are consistent with canopy height and GaoFen imagery. In contrast, the results of Yang et al. (2023) and Yan et al. (2023) clearly reflect the influence of geographic and vegetation zonation. The Pu'er region in Yunnan is located within tropical monsoon forest–rainforest zones (Fig. 14d, e, Rows 1–2) and therefore shows generally higher AGB estimates, whereas the Tongren region in Guizhou lies within subtropical evergreen broadleaf forests (Fig. 14d, e, Rows 3–4) and is characterized by predominantly moderate to high AGB values; differences among vegetation zones are thus amplified in these datasets. In addition, the forest mask applied by Yan et al. (2023) produces visible no-data gaps. These omissions suggest that forest masks that are temporally mismatched with the AGB reference year or contain classification errors may exclude areas that currently contain tree biomass from existing AGB maps. The discrepancies observed in the dataset of Cai et al. (2024) (Fig. 14g) are mainly reflected in areas with complex terrain and dense forests, consistent with the product-specific differences in reference data and input variables discussed above. Together, these representative regions provide local evidence for the causes inferred from the regional difference patterns.

Section 4.2, "Evaluation with independent UAV LiDAR CHM data" (page 17, lines 396–400) [ADDED]

The structural-consistency assessment in Fig. 8 uses the forest subset of the independent UAV LiDAR validation data ($n = 3{,}591$) and is intended to evaluate whether mapped AGB follows expected canopy-height variation rather than to represent the regional AGB frequency distribution.

Fig. 8 caption (page 18) [REVISED]

Structural-consistency assessment of the AGB datasets against mean UAV LiDAR-derived canopy height using forest validation samples. Results are shown for (a) FineKarstAGB ($n = 3{,}591$), (b) Yang et al. (2023), (c) Yan et al. (2023), (d) ESA CCI, and (e) Cai et al. (2024). Sample sizes in panels (b)–(e) vary because only locations with valid values in the corresponding products were included.