

Responses to Reviewers' Comments for Manuscript ESSD-2026-546

**ChinaTCC30: An annual 30 m tree canopy cover
dataset for China from the 1970s to 2024**

Addressed Comments for Publication to

Earth System Science Data

by

Jinge Yu, Zhen Yu, Wangya Han, Fangmin Zhang, and Shirong Liu

Response Letter

Dear Editors,

Please find enclosed the revised version of our manuscript entitled “**ChinaTCC30: An annual 30 m tree canopy cover dataset for China from the 1970s to 2024**” with manuscript number **essd-2026-546**. We would like to sincerely thank you and the reviewers for the valuable comments and constructive suggestions, which have helped improve the quality of our manuscript. In this revision, we have carefully addressed all the reviewers’ comments. A summary of the main modifications and a detailed point-by-point response to the comments from the reviewers are provided below.

Sincerely,

Jinge Yu, Zhen Yu, Wangya Han, Fangmin Zhang, and Shirong Liu

Note: To enhance the legibility of this response letter, all the editor’s and reviewers’ comments are typeset in boxes. Point-by-point response sentences are shown in the color text. Rephrased or added sentences are shown in the color boxes. The corresponding parts in the revised manuscript are highlighted to indicate the changes.

Authors' Response to Reviewer 1

General comments

This study develops ChinaTCC30, a long-term tree canopy cover dataset for China spanning 1975-2024 by integrating Landsat observations, visually interpreted samples, machine learning, active learning, and stand-growth-based historical reconstruction. Overall, the manuscript is well structured and addresses an important need for long-term, spatially explicit tree canopy cover information in China. The model shows generally good performance based on the reported validation results, and the comparisons with independent datasets and existing TCC products demonstrate the potential value of ChinaTCC30 for studying forest dynamics and ecological change. However, several concerns remain regarding the validation of the long-term time series, the reliability of the 1975 and 1980 historical reconstructions, the design and independence of the training and testing samples, and the treatment of temporal consistency and uncertainty. My major and specific comments are provided below for clarity.

Response: Thank you very much for your positive assessment of our study and for recognizing the potential value of ChinaTCC30 for investigating long-term forest dynamics and ecological change in China. We also greatly appreciate the reviewer's careful evaluation and the constructive concerns regarding the validation of the long-term time series, the reliability of the 1975 and 1980 historical reconstructions, the independence of the training and testing samples, and the treatment of temporal consistency and uncertainty.

In response to these comments, we have carefully revised the manuscript and conducted additional analyses where necessary. In particular, we have strengthened the description and evaluation of the historical TCC reconstruction, clarified the design and spatial independence of the training and testing samples, provided additional evidence regarding the reliability of the long-term temporal patterns, and further clarified the treatment and interpretation of uncertainty in the dataset. We have also revised the corresponding Methods, Results, Discussion, and Supplementary Materials of the study. Detailed responses to each comment and the corresponding revisions are provided below.

Major comment 1

The reliability of the reconstructed 1975 and 1980 TCC products requires further assessment. Because these historical TCC products are reconstructed from stand-growth relationships and pseudo-samples rather than directly observed Landsat TCC, the reported prediction accuracy of the TCC model alone does not fully characterize the accuracy and uncertainty of the final historical products. It would be important to clarify how uncertainties in reconstructed stand age and stem density propagate into the final TCC estimates. In addition, the spatial and biome-level distribution of the pseudo-samples should be examined, particularly the apparent scarcity of high-TCC samples in southern China and the limited samples for some biomes such as subtropical evergreen broad-leaved forest and tropical monsoon forest and rainforest. The potential influence of these sampling imbalances on the accuracy and spatial reliability of the reconstructed 1975 and 1980 TCC products should be evaluated.

Response: Thank you for this important comment. In the revised manuscript, we further evaluated the reliability of the historical reconstructions using independent historical pseudo-samples stratified by

vegetation region (Table 2). This additional assessment allowed us to examine how reconstruction accuracy varied among regions and biomes, particularly in dense and structurally complex forest types where historical pseudo-samples were relatively limited. Because reliable nationwide direct TCC observations or high-accuracy reference products are unavailable for 1975 and 1980, these independent pseudo-samples, generated from a plot-scale backcasting model whose reliability was supported by five-fold cross-validation, provide the most feasible reference for evaluating the historical reconstructions.

To alleviate the scarcity of high-TCC pseudo-samples in southern China, we additionally searched for eligible stable natural even-aged pure stands with particular attention to southern high-TCC forest regions. This procedure generated 227 additional pseudo-samples through the same back-casting workflow, approximately one quarter of which were located in southwestern and southern China. Based on the expanded pseudo-sample dataset, we re-ran the 1975 and 1980 TCC mapping to improve the representation of high-TCC forest conditions in southern China.

We also clarified the uncertainty propagation pathway in the Discussion. Specifically, uncertainties may arise from the empirical back-casting of stand age and stem density, affect the predicted TCC values of the historical pseudo-samples through the plot-scale TCC model, and then be transferred to the MSS-based spatial reconstruction during model training. These uncertainties may ultimately propagate into the wall-to-wall 1975 and 1980 TCC maps.

Where changed:

Manuscript —3.4.1 Independent accuracy evaluation — Page 18, Lines 429-432. [NEWLY ADDED]

Validation based on pseudo-samples showed larger errors overall. Specifically, the biome-level error pattern was generally consistent with that obtained from the photo-interpreted TCC validation, with larger errors occurring in dense and structurally complex forest biomes (Table 2).

Manuscript —3.4.1 Independent accuracy evaluation — Page 19, Line 440. [NEWLY ADDED]

Table 2. Accuracy assessment across ecoregions based on independent pseudo- tree canopy cover

Ecoregions	n	RMSE (%)	Bias (%)
Alpine vegetation on the Qinghai-Tibet Plateau	73	9.21	0.82
Cold temperate coniferous forest	52	18.34	8.15
Subtropical evergreen broad-leaved forest	683	17.16	3.04
Temperate coniferous and deciduous broad-leaved mixed forest	236	19.47	-1.72
Temperate desert	199	10.88	-1.94
Temperate grassland	163	13.95	-3.01
Tropical monsoon forest and rainforest	68	19.04	0.88
Warm temperate deciduous broad-leaved forest	75	17.42	2.26

Manuscript —3.1 Construction of pseudo-sample data based on stand factors for the years 1975 and 1980 — Page 15, Line 380. [MODIFIED]

The number of pseudo samples in the study area has been updated in the Fig. 4c.

Manuscript —3.3 Training and optimization of retrieval models— Page 17, Lines 397-400. [MODIFIED]

KDAL-RF reached a near-stable performance level at about 2,420 samples ($R^2 = 0.71$, RMSE = 17.10%), whereas random-sampling RF required about 3,140 samples to achieve a comparable level, corresponding to a reduction of approximately 22% in training sample demand (Fig. 6a, b).

Manuscript —3.3 Training and optimization of retrieval models— Page 17, Lines 406-408.

[MODIFIED]

Overall, the active learning strategy achieved higher sample efficiency in both experiments, reducing the required training sample size by about 22% in the pseudo-sample-based experiment and by about 48% in the visually interpreted sample experiment (Fig. 6).

Manuscript — Discussion — Page 36, Lines 769–776. [MODIFIED]

In addition, the uncertainty of the early baseline estimates for 1975 and 1980 is closely related to the historical backtracking procedure. Because stable, direct observations at the national scale are unavailable for these two periods, the corresponding estimates had to be indirectly reconstructed by combining historical stand-factor backtracking with the TCC prediction model. As a result, the early modeling samples are essentially historical reconstructions under model constraints. Their accuracy depends not only on the performance of the TCC pseudo-sample prediction model itself, but also on the reliability of the historical stand-structure backtracking. Specifically, uncertainties in backtracked stand age and stem density may first affect the predicted TCC values of the historical pseudo-samples; these uncertainties can then be transferred to the MSS-based reconstruction model during training and ultimately propagate into the wall-to-wall 1975 and 1980 TCC maps.

Major comment 2

The sample allocation and validation procedure require further clarification. The manuscript reports a total of 10000 samples, among which 2692 are reserved as independent hold-out validation samples. The remaining 7308 samples consist of 2560 initial training samples and 4748 samples forming the unlabeled pool for the K-means-based Diversity Active Learning (KDAL). However, it remains unclear how these 7308 samples are processed across iterative active-learning cycles. The manuscript should clarify how all 10000 samples were allocated and used throughout the active-learning and five-fold cross-validation procedures: specifically, how many samples were selected from the 4748 unlabeled pool and added into the training set during successive KDAL iterations, the final composition and total size of the complete training dataset, and the fate of those unselected samples remaining in the 4748 pool. It should also be explicitly stated whether the 2692 independent validation samples were completely excluded from the unlabeled pool and never involved in any step of active-learning sample querying or model training to avoid potential data leakage. The spatial and TCC distributions of the different sample groups should also be reported, and the spatial locations of the final training- and validation-sample sets should be shown on maps, to demonstrate the representativeness and independence of the training and validation datasets.

Response: Thank you for this helpful comment. We agree that the previous manuscript did not sufficiently clarify how the visually interpreted samples were allocated and used throughout the active-learning and validation procedures. We have therefore revised the manuscript and added a supplementary table to explicitly summarize the sample partitioning and usage across model training, active learning, cross-validation, and independent validation.

Although 10,000 interpretation plots were initially established, plots for which reliable visual interpretation was not possible were subsequently excluded, including those affected by cloud contamination in the VHR basemaps, those located along VHR image mosaic seams, and those for which the VHR imagery could not be loaded. After this quality-control screening, 9,843 valid photo-interpreted TCC samples were retained. This sample quality-control process and the resulting number of valid interpretation plots have been clarified in the revised manuscript.

Among the 9,843 valid samples, 2,692 samples were assigned to an independent validation set before the active-learning procedure. To avoid data leakage, these samples were completely excluded from the unlabeled sample pool, active-learning querying, model training, and five-fold cross-validation. Of the remaining samples, 522 samples were extracted as an independent balanced benchmark/control set for independently evaluating the active-learning strategy, while 6,629 samples were used as the active-learning optimization pool. Samples selected through successive KDAL iterations were added to the training set for model training. The five-fold cross-validation was conducted only within the training sample set used for model development, and the independent validation samples were not involved in any validation fold.

To further demonstrate the representativeness and independence of the different sample groups, we added a new supplementary figure (Fig. S9) showing the spatial distribution of the independent test sets (ITS), Independent balanced benchmark set for active-learning evaluation (IBS), and active-learning optimization pools (ALP) for both photo-interpreted samples and historical pseudo-samples. The donut charts in Fig. S9 summarize the sample sizes and TCC distributions across five TCC intervals (0–20%, 20–40%, 40–60%, 60–80%, and 80–100%). These additions clarify the spatial allocation and TCC-gradient coverage of the different sample subsets used in the modeling and validation workflows.

Where Changed:

Manuscript — Materials and methods — Page 5, Lines 157–161. [MODIFIED]

We established 10,000 interpretation plots across the study area, with each plot dimensioned at 30×30 m to perfectly match the spatial resolution and boundaries of the Landsat grid. Plots for which reliable visual interpretation was not possible were subsequently excluded, including those affected by cloud contamination in the VHR, those located along VHR image mosaic seams, and those for which the VHR imagery could not be loaded. After this quality-control screening, 9,843 valid interpretation plots were retained (Fig. 1a).

Manuscript — 2.3.6 TCC mapping approach — Page 11, Lines 301–302. [NEWLY ADDED]

(the division and spatial distribution of all samples are shown in Supplementary Table S6 and Fig. S9).

Supplement — Table S6 — Page 6. [MODIFIED]

Table S6. Summary of sample partitioning and usage for photo-interpreted samples and historical pseudo-samples.

Sample stream	Subset	n	Description
	Total valid sample pool	9,843	Samples retained after removing plots affected by cloud contamination, poor VHR image quality, or other interpretation issues.
Photo-interpreted samples	Independent test set	2,692	Independent accuracy assessment of annual mapping for 1985–2024.
	Independent balanced benchmark set for active-learning evaluation	522	Benchmark set for evaluating active-learning sample selection.
	Active-learning optimization pool	6,629	Candidate pool for KDAL-based sample optimization.

Historical pseudo-samples	Total valid pseudo-sample pool	7,753	Sample pool for historical TCC reconstruction in 1975 and 1980.
	Independent test set derived from pseudo-sample	1,549	Independent accuracy assessment of historical reconstruction in 1975 and 1980.
	Independent balanced benchmark set for active-learning evaluation	452	Benchmark set for evaluating active-learning sample selection using pseudo-samples
	Active-learning optimization pool	5,752	Candidate pool for KDAL-based sample optimization in historical reconstruction

Supplement —Figure S9 — Page 11. [NEWLY ADDED]

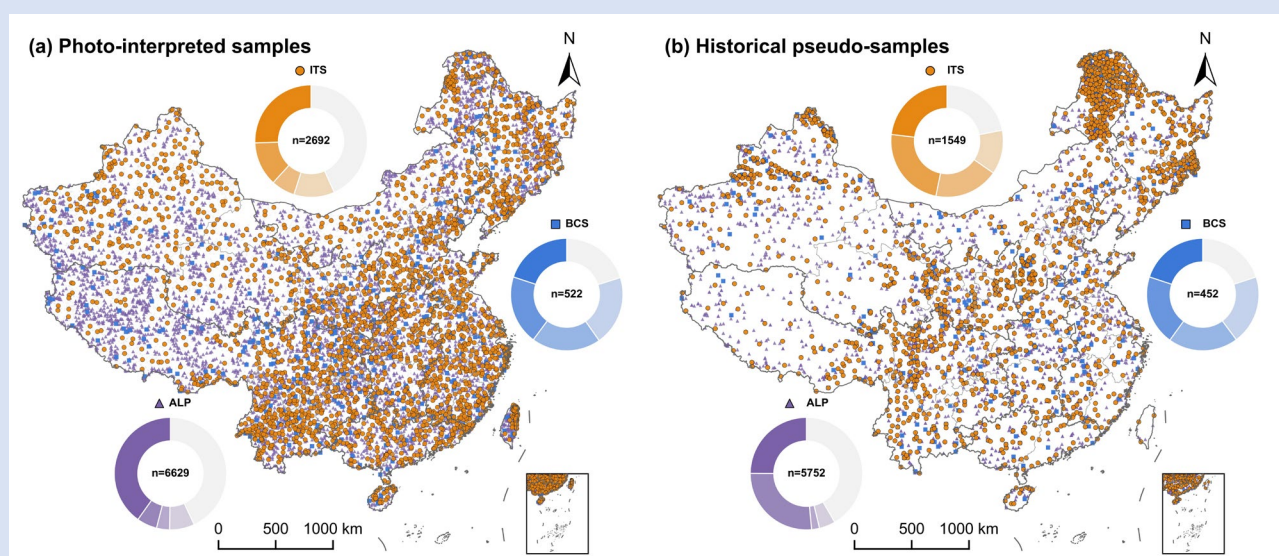


Figure S9. Spatial distribution and TCC-interval composition of sample subsets used for model evaluation and active-learning optimization. Panels show the spatial distributions of independent test sets (ITS), benchmark control sets (BCS), and active-learning optimization pools (ALP) for (a) photo-interpreted samples and (b) historical pseudo-samples. The donut charts summarize the sample size and proportional distribution of each subset across five TCC intervals: 0–20%, 20–40%, 40–60%, 60–80%, and 80–100%. Colors from light to dark indicate increasing TCC intervals.

Specific comment 1

The uncertainty reported in the Abstract is described as being derived from predictions from multiple models, but this should be explained more explicitly and consistently with the Methods.

Response: Thank you for this comment. We have revised the Abstract and Methods to describe the uncertainty estimation more explicitly and consistently. In the revised Abstract, we now clarify that pixel-level uncertainty was quantified as the standard deviation among five cross-validation model predictions.

Manuscript — Abstract — Page 1, Line 23. [MODIFIED]

Pixel-level uncertainty was quantified as the standard deviation among five cross-validation model predictions.

Specific comment 2

Lines 94-97: The authors state that the dataset developed by Cai et al. (2024) may have uncertainties in high-latitude regions because of topographic shadow effects. Since the present study also uses the Landsat data processed by Cai et al. (2024), the authors should clarify whether and how the topographic-shadow problem was addressed in the present study.

Response: We thank the reviewer for this important comment. We agree that topographic shadow effects may also influence ChinaTCC30 because our mapping framework uses the Landsat composites processed by Cai et al. (2024). In the revised manuscript, we have clarified that ChinaTCC30 does not rely solely on spectral predictors. DEM-derived topographic predictors were incorporated into the TCC estimation model to help account for terrain-related variations in illumination and surface reflectance. To evaluate whether these predictors helped reduce topography-associated prediction errors, we conducted an additional ablation experiment comparing models with and without DEM-derived predictors under otherwise identical random forest settings. We have added this clarification and the corresponding ablation results to the revised manuscript and supplementary materials.

Where changed:

Manuscript —4.2 Limitations and future improvement — Page 35, Lines 742–749. [NEWLY ADDED]

Second, complex topographic conditions remain an important source of uncertainty in ChinaTCC30. To mitigate this issue, terrain information derived from the improved NASADEM dataset was incorporated into the TCC estimation model. An ablation analysis showed that these DEM-derived predictors reduced prediction errors in shaded-slope regions (Supplementary Text S6; Fig. S12). Nevertheless, in mountainous and high-latitude regions, terrain shadow, slope-aspect differences, and local variations in illumination geometry may affect surface reflectance (Teillet et al., 1982; Civco, 1989), thereby weakening the relationship between remotely sensed signals and actual canopy structure. Although topographic correction can reduce these effects to some extent, shadowing and radiometric distortion may still introduce local biases in areas with strong topographic relief.

Supplementary —Text S6 — Page 3. [NEWLY ADDED]

An ablation experiment was conducted to evaluate the contribution of NASADEM-derived topographic predictors to TCC estimation under complex terrain conditions. Using identical random forest settings and the same spectral predictors, model performance was compared between configurations with and without the topographic predictors. The inclusion of NASADEM-derived topographic predictors improved TCC estimation accuracy in both sunlit- and shaded-slope regions. In sunlit-slope regions, R^2 increased from 0.73 to 0.75, and RMSE decreased from 19.11% to 18.39%. A larger improvement was observed in shaded-slope regions, where R^2 increased from 0.63 to 0.69 and RMSE decreased from 22.00% to 20.14%. These results suggest that high-quality DEM-derived terrain information may partly reduce terrain-associated prediction errors in Landsat-based TCC mapping, especially under shaded terrain conditions.

Supplementary —Figure S12 — Page 12. [NEWLY ADDED]

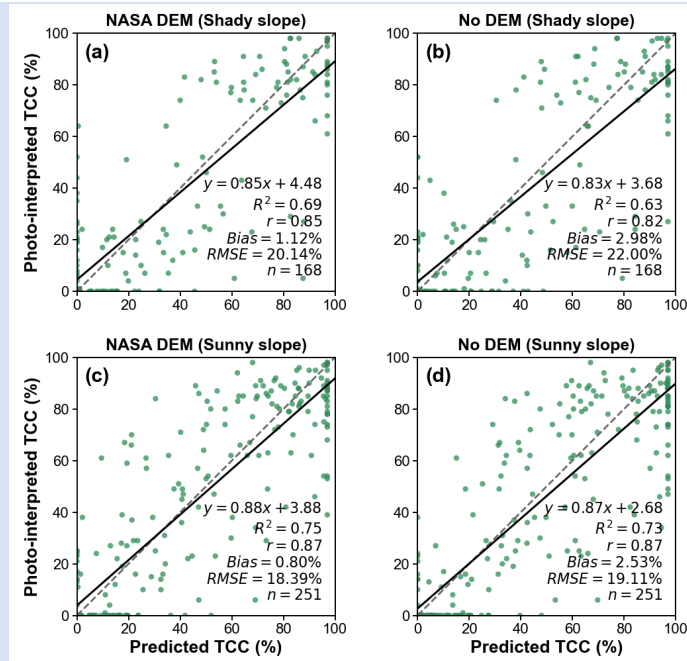


Figure S12. Ablation experiment evaluating the contribution of DEM-derived topographic predictors to TCC estimation in the high-latitude test subset. All models used identical random forest settings and the same spectral predictors (B1, B2, B3, B5, B7, EVI, NDBI, NDVI, MSAVI, ARVI, RVI, NDSI, MNDWI, and NBR), but differed in whether DEM-derived topographic predictors were included. Panels (a, b) show results for shaded-slope regions, and panels (c, d) show results for sunlit-slope regions.

Specific comment 3

Line 105: The term “early historical TCC baselines” is somewhat vague and should be specified more clearly. Since the manuscript subsequently focuses on the 1975 and 1980 historical products, these years should be explicitly identified here.

Response: We thank the reviewer for this helpful suggestion. We agree that the term “early historical TCC baselines” was vague. In the revised manuscript, we have explicitly identified the historical baseline years as 1975 and 1980 to make the statement clearer and more consistent with the subsequent description of the ChinaTCC30 dataset.

Where changed:

Manuscript — Introduction — Page 4, Line 105. [MODIFIED]

In summary, historical baselines for 1975 and 1980 and long-term annual TCC mapping for China remain insufficient.

Specific comment 4

Line 103: The term “stand growth modeling” is inconsistent with the terminology used in the Abstract, where “stand-growth historical backtracking” is used.

Response: We thank the reviewer for pointing out this terminology inconsistency. We have revised “stand growth modeling” to “stand-growth-based historical backtracking” to ensure consistency with the terminology used in the Abstract and throughout the manuscript.

Where changed:

Manuscript — Introduction — Page 4, Lines 105–112. [MODIFIED]

In summary, historical baselines for 1975 and 1980 and long-term annual TCC mapping for China remain insufficient. To address this gap, this study develops an integrated framework that combines Random Forest, active learning-based optimization, and stand-growth-based historical backtracking to enable cost-effective and accurate nationwide TCC mapping across China from 1975 to 2024. Specifically, the objectives are to: (1) evaluate the effectiveness of integrating active learning with machine learning for long-term, low-cost TCC mapping; (2) develop a reliable approach that integrates stand-growth-based historical backtracking with machine learning to accurately back-cast early TCC baselines under severely limited early ground-reference data; and (3) construct a cross-sensor consistent TCC dataset based on multiple generations of Earth observation satellites to support long-term spatiotemporal analysis.

Specific comment 5

Lines 116-118: The authors state that Landsat MSS data were used to generate the 1975 and 1980 products, but it is unclear whether these historical MSS observations underwent the same quality-control procedures as the Landsat data used for the later period.

Response: Thank you for this comment. We agree that the quality-control procedure for the historical MSS imagery should be clarified. Because Landsat MSS data differ from later TM/ETM+/OLI observations in terms of spectral configuration, radiometric quality, geometric accuracy, and the availability of standardized QA bands, the exact same quality-control workflow used for the later Landsat period could not be directly applied. Therefore, we implemented a quality-control procedure tailored to MSS imagery.

Specifically, MSS scenes with severe cloud contamination, missing spectral bands, large geometric displacement, abnormal striping, saturation, or other obvious image artifacts were excluded before compositing. For the retained MSS observations, cloud- and shadow-contaminated pixels were further screened where possible, and valid observations within each baseline window were composited on a pixel-wise basis using a medoid-like approach to reduce the influence of residual clouds, shadows, and radiometric outliers. Remaining spatial gaps were supplemented using the Global Land Survey (GLS) image collection.

Although the compositing and quality-control procedures differed between MSS imagery and later TM/ETM+/OLI imagery because of differences in sensor characteristics and data availability, the two data streams were used in separate modeling workflows. The MSS composites were used only for the 1975 and 1980 historical reconstruction models, whereas the later Landsat composites were used for the annual 1985–2024 mapping models. Therefore, the model trained for the later annual mapping period was not directly applied to the MSS-based historical reconstruction. We have briefly clarified the MSS-specific quality-control, and compositing procedure in the revised Methods, and provided a more detailed description in Supplementary Text S3.

Where changed:

Supplementary —Text S3 — Page 1. [NEWLY ADDED]

Text S3 | Landsat MSS compositing for the 1975 and 1980 historical baselines

For the 1975 and 1980 historical reconstructions, Landsat MSS imagery was processed separately from the later Landsat TM/ETM+/OLI-based annual mapping workflow because of differences in sensor characteristics, spectral configuration, data quality, and the availability of QA information.

MSS observations acquired during 1972-1977 and 1978-1984 were used to construct the nominal 1975 and 1980 baseline composites, respectively. Therefore, these two products should be interpreted as multi-year baseline epochs centered on the nominal years rather than observations from single years. Before compositing, MSS imagery was subjected to pixel-level quality screening. Fill pixels and cloud-contaminated pixels were removed using the QA_PIXEL band, and radiometrically saturated pixels were excluded using the QA_RADSAT band. The retained MSS spectral bands were then used for compositing. For each baseline epoch, Landsat MSS Collection 2 Tier 1 images were used as the primary data source because of their higher geometric quality, and the median value of each MSS band was calculated from all valid observations within the corresponding baseline window. In addition, remaining spatial gaps were supplemented using the Global Land Survey (GLS) image collection. For the 1975 baseline, GLS observations closest to the nominal baseline year were prioritized, whereas GLS observations from the broader 1972-1977 window were used only to fill remaining gaps. The same hierarchical logic was applied to the 1980 baseline reconstruction. In this procedure, all available valid MSS observations were retained, and GLS data were used only as supplementary gap-filling information. The resulting MSS baseline composites were used as spectral inputs for the historical TCC reconstruction models for 1975 and 1980.

Specific comment 6

Lines 120-123: The manuscript states that vegetation-related predictors include multiple spectral-band metrics and vegetation indices, but it is unclear how these temporal variables were summarized. The authors should specify whether the predictors were derived from annual means, growing-season means, medians, or another temporal aggregation method, and should provide the same information for all relevant temporal predictors.

Response: We thank the reviewer for this helpful comment. We agree that the temporal aggregation method used to derive the vegetation-related predictors should be described more clearly. In the revised manuscript, we have clarified that the vegetation-related predictors for the 1985–2024 annual mapping were derived from annual leaf-on Landsat medoid composite images, rather than from annual means, growing-season means, or simple median composites. Specifically, the 30 m Landsat composites used in this study were obtained from the spatiotemporally consistent annual Landsat composite dataset generated by Cai et al. (2025a). For each year, Landsat observations acquired during the leaf-on season were processed using QA-based cloud and cloud-shadow masking, cross-sensor spectral harmonization, medoid compositing, and temporal gap filling. Therefore, all spectral-band metrics and vegetation indices used as predictors for the 1985–2024 period were calculated from the corresponding annual leaf-on medoid composite images.

Where changed:

Manuscript —Materials and methods — Page 4, Lines 116–121. [MODIFIED]

The 30 m imagery was obtained from the spatiotemporally consistent annual leaf-on Landsat composite dataset generated by Cai et al. (2025a) covering the period of 1985–2024. Specifically, the annual dataset was pre-processed in Google Earth Engine during China’s leaf-on season, defined as day of year (DOY) 150–300, QA-based cloud masking (Zhu et al., 2015), cross-sensor harmonization (Roy et al., 2016), and medoid compositing with temporal gap-filling. Therefore, the vegetation-related predictors used for the 1985–2024 annual mapping were calculated from annual growing-season medoid composite images.

Specific comment 7

The temporal coverage and role of the ALCC dataset require clarification. The manuscript indicates that ALCC data cover 1985–2022, but it is unclear which years of ALCC data were actually used in the model. In particular, when predicting TCC for 1975 and 1980, did the authors use the 1985 ALCC layer as a predictor? More importantly, ALCC does not appear to be clearly identified as a predictor in the subsequent methodological description or workflow diagram. The authors should clarify whether ALCC was used as a predictor, for which years, and how it was incorporated into the historical reconstruction.

Response: We thank the reviewer for this helpful comment. We agree that the temporal coverage and role of the land-cover datasets should be clarified more explicitly.

In the revised manuscript, we have clarified that land-cover information was not used in the historical reconstruction for 1975 and 1980, because corresponding annual land-cover data were not available for these two baseline years. Therefore, the ALCC layers was not used as a predictor for reconstructing TCC in 1975 or 1980.

For the annual TCC mapping period from 1985 to 2024, land-cover information was incorporated as an auxiliary categorical predictor. Specifically, because ALCC covers 1985–2022, it was used as the primary annual land-cover dataset. Since ALCC does not cover 2023 and 2024, CLCD was used as a supplementary land-cover source to provide land-cover information for these two years. The land-cover classes were not subjected to the subsequent feature-selection procedure; instead, they were directly transformed into one-hot-encoded categorical variables and incorporated into the Random Forest model for model construction and prediction.

We have revised the methodological description to explicitly identify the land-cover variables as auxiliary categorical predictors for the 1985–2024 annual mapping period, and to clarify that they were not used for the 1975 and 1980 historical reconstruction.

Where changed:

Manuscript —2.1.2 Auxiliary data — Page 5, Lines 134–141. [MODIFIED]

Instead of using selected land cover categories, such as cropland and water, as binary masks to remove non-forest areas, which can underestimate tree cover in mixed pixels, we introduced land cover information as an auxiliary predictor during model training for 1985–2024. This strategy allows the model to use categorical contextual information, thereby improving TCC estimation in heterogeneous landscapes while preserving sensitivity to sub-pixel tree features. For the 1985–2022 period, the Annual Land Cover of China (ALCC) dataset (Li et al., 2026) was used as the primary land-cover source. Because ALCC does not cover 2023 and 2024, the China Land Cover Dataset

(CLCD) (Yang and Huang, 2021), which uses a land-cover classification scheme consistent with ALCC, was used as a supplementary source to provide land-cover information for these two years.

Specific comment 8

Line 146: “grids” should be corrected to “grid.”

Response: We thank the reviewer for pointing this out. We have corrected “grids” to “grid” in the revised manuscript.

Where changed:

Manuscript — 2.2.1 Photo-interpreted tree cover data — Page 5, Line 161. [MODIFIED]

To precisely quantify the TCC, a 10×10 point grid...

Specific comment 9

Line 149: The binary classification should be described explicitly as “1 for tree canopy and 0 for non-tree canopy”.

Response: We thank the reviewer for this helpful suggestion. We have revised the description of the binary classification to explicitly indicate that 1 represents tree canopy and 0 represents non-tree canopy.

Where changed:

Manuscript — 2.2.1 Photo-interpreted tree cover data — Page 6, Line 163. [MODIFIED]

(1 for tree canopy, 0 for non-tree canopy)

Specific comment 10

Lines 174-175: The role of GEDI tree canopy cover data in the model development is unclear. In the subsequent Methods section (Section 2.3.6), GEDI tree canopy cover does not appear to be used as a training variable or training dataset, whereas it is subsequently used for comparison with other TCC products in the Discussion.

Response: We thank the reviewer for pointing out this ambiguity. We agree that the previous wording may have incorrectly implied that GEDI canopy cover data were used for model development. In fact, GEDI data were not used as training samples, response variables, or predictors in the Random Forest modeling framework. Instead, the GEDI Level 2B Canopy Cover product was used only as an independent reference dataset for product comparison and spatial consistency assessment. We have revised the relevant text to clarify this role and removed wording suggesting that GEDI supported model training.

Where changed:

Manuscript — 2.2.3 Spaceborne Lidar derived canopy cover data — Page 7, Lines 184-186. [MODIFIED]

To further evaluate the spatial consistency and reliability of ChinaTCC30, we utilized the Global Ecosystem Dynamics Investigation (GEDI) Level 2B Canopy Cover product (GEDI02_B) spanning from 2019 to 2024 as an independent reference dataset.

Manuscript — 2.2.3 Spaceborne Lidar derived canopy cover data — Page 7, Lines 191-194. [MODIFIED]

The quality-filtered GEDI canopy cover observations were aggregated to the 30 m ChinaTCC30 grid

using a mean reducer in GEE. Through these procedures, we obtained a nationwide distribution of high-quality GEDI footprints (see Fig. S10 for their spatial distribution), which were used only for independent product comparison and spatial consistency assessment.

Specific comment 11

Line 244: It should be “four standard statistical metrics”.

Response: We thank the reviewer for this helpful suggestion. We have revised the description.

Where changed:

Manuscript —2.3.4 Accuracy assessment — Page 10, Line 278. [MODIFIED]

To quantitatively assess the accuracy of the TCC estimations, four standard statistical metrics were calculated.

Specific comment 12

Line 318: The importance of DVI appears to be relatively low compared with the other predictors in Figure 5b1. Why was DVI nevertheless retained in the final predictor set?

Response: Thank you for pointing this out. As described earlier in the manuscript, the visible bands (Green and Red) and vegetation indices (MSAVI, RVI, and NDVI) exhibited strong correlations with TCC and relatively high feature importance. DVI was therefore not retained in the final predictor set. The original statement in Line 318 was inaccurate, and we have corrected it in the revised manuscript to ensure consistency with the feature selection results.

Where changed:

Manuscript —3.2 Evaluating variable importance and selection — Page 14, Lines 372-375. [MODIFIED]

For the epochal baseline modeling (Fig. 5a1, b1), visible bands (Green, Red) and vegetation indices (MSAVI, RVI, and NDVI) exhibited strong correlations and high feature importance. Consequently, six features were retained for the historical baseline reconstruction, comprising two spectral bands (Green and Red), three spectral indices (MSAVI, RVI, and NDVI), and one topographic factor (Slope).

Specific comment 13

Figure 4b requires clarification regarding the temporal reference of the two TCC variables.

Response: We thank the reviewer for pointing out this ambiguity. In the revised manuscript, we have further clarified the caption of Fig. 4b by specifying the temporal reference of the two TCC variables. Specifically, the 5-fold cross-validation shown in Fig. 4b was conducted for the 2018 reference year. We have also explained that photo-interpreted TCC represents the observed reference TCC, whereas predicted pseudo-TCC represents the model-estimated TCC for the same year.

Where changed:

Manuscript —3.1 Construction of pseudo-sample data based on stand factors for the years 1975 and 1980 — Page 15, Lines 380-383. [MODIFIED]

(b) Performance of the pseudo-plot reconstruction model based on 5-fold cross-validation for the 2018 reference year, where photo-interpreted TCC represents the observed reference TCC and predicted pseudo-TCC represents the model-estimated TCC for the same year.

Specific comment 14

The validation of the 1975 and 1980 TCC products based on pseudo-samples should be further stratified by biome or forest type.

Response: Thank you for this suggestion. We agree that the validation of the historical 1975 and 1980 TCC products should be further examined across different ecological contexts. In the revised manuscript, we have added a stratified validation based on independent historical pseudo-samples across vegetation regions/forest types (Table 2). The results showed that, across the eight ecoregions, RMSE ranged from 9.21% to 19.47% (Table 2). Relatively lower errors were observed in alpine vegetation on the Qinghai-Tibet Plateau and temperate desert regions, whereas higher errors occurred in dense and structurally complex forest regions, such as temperate coniferous and deciduous broad-leaved mixed forest and tropical monsoon forest/rainforest. We have added this pseudo-sample-based ecoregion-stratified evaluation to the revised manuscript to better characterize the biome-level reliability of the 1975 and 1980 historical TCC products.

Where changed:

Manuscript —3.4.1 Independent accuracy evaluation— Page 19, Line 440. [NEWLY ADDED]

Table 2. Accuracy assessment across ecoregions based on independent historical pseudo-TCC samples.

Ecoregions	n	RMSE (%)	Bias (%)
Alpine vegetation on the Qinghai-Tibet Plateau	73	9.21	0.82
Cold temperate coniferous forest	52	18.34	8.15
Subtropical evergreen broad-leaved forest	683	17.16	3.04
Temperate coniferous and deciduous broad-leaved mixed forest	236	19.47	-1.72
Temperate desert	199	10.88	-1.94
Temperate grassland	163	13.95	-3.01
Tropical monsoon forest and rainforest	68	19.04	0.88
Warm temperate deciduous broad-leaved forest	75	17.42	2.26

Manuscript —3.4.1 Independent accuracy evaluation— Page 18, Lines 429-432. [NEWLY ADDED]

Validation based on pseudo-samples showed larger errors overall. Specifically, the biome-level error pattern was generally consistent with that obtained from the photo-interpreted TCC validation, with larger errors occurring in dense and structurally complex forest biomes (Table 2).

Specific comment 15

Line 395: The number of NFI sample plots or observations used to validate the ChinaTCC30 product is not reported.

Response: Thank you for pointing this out. We have revised the manuscript to state that a total of 11,618 NFI sample plots were used to validate the ChinaTCC30 product across the four national forest inventory periods (6th–9th). This information has also been added to the caption of Figure 8.

Where changed:

Manuscript —3.4.2 External validation using NFI data — Page 20, Lines 455-456. [MODIFIED]

Figure 8. Validation of the ChinaTCC30 product against NFI data across four national inventory periods (6th–9th; n = 11,618 NFI sample plots).

Specific comment 16

Panels (e-h) in Figure 14 appear to duplicate the validation results already presented in Figure 8.

Response: Thank you for pointing this out. We agree that panels (e-h) in the previous Figure 14 duplicated the NFI-based validation results already presented in Figure 8. To avoid redundancy and improve the clarity of the manuscript, we have removed these duplicated panels from Figure 14. The revised Figure 14 now focuses only on the validation of CATCD against NFI observations, while Figure 8 retains the detailed multi-period NFI validation of ChinaTCC30.

Where changed:

Manuscript —3.4.2 External validation using NFI data — Page 20, Lines 455-456. [MODIFIED]

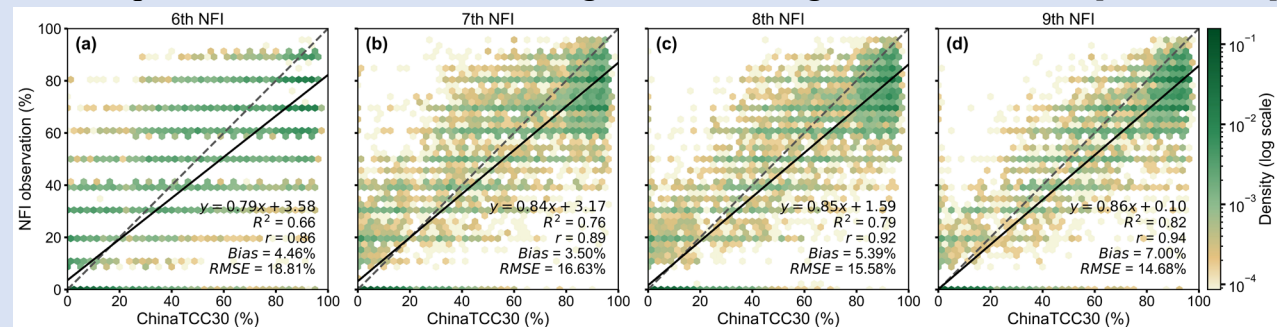


Figure 8. Validation of the ChinaTCC30 product against NFI data across four national inventory periods (6th–9th; n = 11,618 NFI sample plots).

Manuscript —4.1.1 TCC mapping comparison based on multi-source reference data—Page 28, Lines 597-600. [MODIFIED]

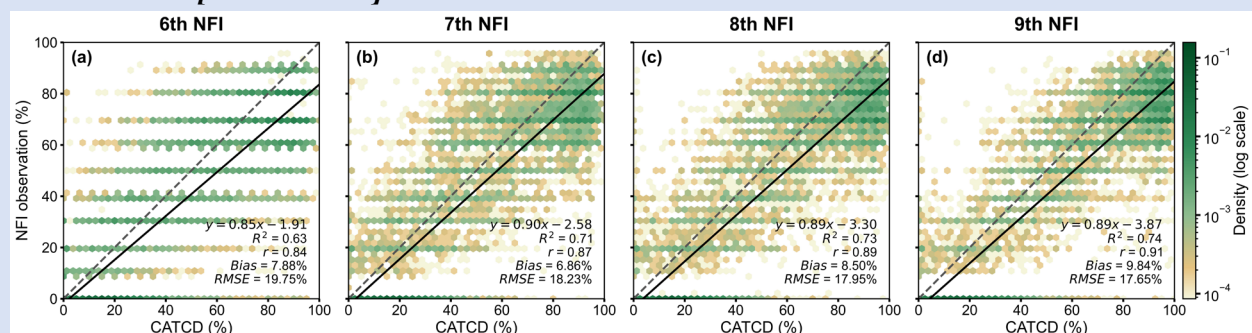


Figure 14. Scatter density plots comparing NFI ground observations with CATCD tree canopy cover estimates. The columns represent four consecutive inventory periods: the 6th (a), 7th (b), 8th (c), and 9th (d) NFI. (n = 11,618 NFI sample plots). ChinaTCC30 validation results are shown in Fig. 8.

Specific comment 17

The year represented by the difference maps in Figure 15 should be explicitly indicated in the figure caption. If different panels correspond to different years, the year should also be clearly labeled within the figure itself to facilitate interpretation.

Response: We thank the reviewer for this helpful suggestion. We have revised the caption of Fig. 15 to explicitly indicate the year represented by each comparison. Specifically, panels (a, d) correspond to the comparison between ChinaTCC30 and GFC in 2000, panels (b, e) correspond to ChinaTCC30

and CATCD in 2010, and panels (c, f) correspond to ChinaTCC30 and GFCC in 2015.

Where changed:

Manuscript —4.1.2 Static and spatial comparison — Page 29, Lines 634-636. [MODIFIED]

Figure 15. Spatial distribution maps of the differences between ChinaTCC30 and three other products and the corresponding frequency distribution histograms. Panels (a, d) show the comparison between ChinaTCC30 and GFC for 2000; panels (b, e) show the comparison between ChinaTCC30 and CATCD for 2010; and panels (c, f) show the comparison between ChinaTCC30 and GFCC for 2015.

Authors' Response to Reviewer 2

General comments

This manuscript presents an annual 30 m tree canopy cover dataset for China from the 1970s to 2024. The methodology is systematic and the dataset is of high value. However, please address the following issues before publication.

Response: We sincerely thank the reviewer for the assessment of our manuscript and for recognizing the value of the ChinaTCC30 dataset. We also appreciate the constructive comments and suggestions. In the revised manuscript, we have carefully addressed all the issues raised by the reviewer, and the detailed point-by-point responses are provided below.

Specific comment 1

The interpretation samples are mainly derived from VHR imagery of 2015–2024, whereas the product provides annual mapping for 1985–2024. Please clarify whether all annual models share the same sample set, and discuss the validity of the underlying assumption that the spectral–TCC relationship is temporally stable and transferable across years, particularly given the radiometric differences between early TM and recent OLI imagery.

Response: We thank the reviewer for this important comment. We agree that the temporal representativeness of the interpretation samples and the transferability of the spectral–TCC relationship should be clarified. In the revised manuscript, we have clarified the sample allocation strategy for different mapping periods. Specifically, the annual 30 m TCC mapping from 1985 to 2024 used the same visually interpreted TCC sample set, which was mainly derived from VHR imagery acquired during 2015–2024. In contrast, the historical 60 m TCC reconstruction for 1975 and 1980 did not use these interpretation samples directly, but relied on a separately constructed pseudo-sample dataset developed within the stand-growth-based historical backtracking framework. To improve transparency, we further summarized the allocation of all samples used in the mapping framework in Supplementary Table S6 and visualized their spatial distribution in Supplementary Fig. S9.

We acknowledge that this strategy relies on the assumption that the relationship between TCC and the predictor space is sufficiently stable and transferable across years after consistent preprocessing. To reduce potential temporal and sensor-related inconsistencies, the annual Landsat composites used in this study were generated using a unified preprocessing framework, including QA-based cloud and cloud-shadow masking, cross-sensor spectral harmonization, medoid compositing, and temporal gap filling. These procedures helped improve temporal consistency and reduce radiometric differences among Landsat sensors, particularly between early TM observations and recent OLI imagery.

Nevertheless, we agree that the spectral–TCC relationship may not be perfectly invariant over time. Residual uncertainties may remain because of differences in sensor radiometry, atmospheric correction quality, image availability, phenological conditions, and background land-cover composition. These uncertainties may be more pronounced in early TM-based years, for which contemporaneous VHR reference samples are unavailable. We have therefore added a discussion of this assumption and its potential limitation in the revised manuscript.

Where changed:***Manuscript — 2.3.6 TCC mapping approach — Page 11, Lines 296-302. [MODIFIED]***

Compared with previous annual TCC mapping efforts that mainly focused on the Landsat TM era and later periods (Hansen et al., 2013; Sexton et al., 2013; Cai et al., 2024), this study further integrates active-learning-based sample optimization, stand-growth-based historical backtracking, MSS-based early-period reconstruction to generate annual and historical estimates of TCC across China from 1970s to 2024, respectively (Fig. 3). Historical 60 m TCC for 1975 and 1980 was reconstructed from Landsat MSS imagery and pseudo-samples developed within a historical backtracking framework, whereas annual 30 m TCC maps for 1985–2024 were produced from Landsat imagery and visually interpreted TCC samples (the division and spatial distribution of all samples are shown in Supplementary Table S6 and Fig. S9).

Manuscript — 4.2 Limitations and future improvement — Page 35, Lines 757-768. [NEWLY ADDED]

Fourth, the temporal representativeness of the visually interpreted samples may also introduce uncertainty into annual TCC mapping. The interpretation samples were mainly derived from VHR imagery acquired during 2015–2024, whereas ChinaTCC30 provides annual TCC estimates for 1985–2024. In the annual mapping framework, the visually interpreted TCC sample pool was used as the reference sample set, while predictor variables were extracted from the corresponding annual Landsat composites for each target year. This strategy assumes that the relationship between TCC and the predictor space is sufficiently stable and transferable across years after consistent preprocessing. The use of QA-based cloud and shadow masking, cross-sensor spectral harmonization, medoid compositing, and temporal gap filling helped reduce sensor-related inconsistencies and improve temporal consistency (Flood, 2013; Roy et al., 2016; Wulder et al., 2019; Cai et al., 2025a). Nevertheless, residual uncertainties may remain because of differences in sensor radiometry, atmospheric correction quality, image availability, phenological conditions, and background land-cover composition. These uncertainties may be more pronounced in early TM-based years, where contemporaneous VHR reference samples are unavailable.

Specific comment 2

The GEDI footprint is 25 m, while the product is 30 m. Please explain the spatial matching/aggregation method between the two, and whether the GEDI geolocation error was considered.

Response: We thank the reviewer for this helpful comment. We agree that the spatial matching between the 25 m GEDI footprints and the 30 m ChinaTCC30 product should be clarified.

In the revised manuscript, we have clarified that GEDI observations are discrete footprint measurements rather than wall-to-wall continuous raster observations. The GEDI02_B product provides canopy cover information at the nominal 25 m footprint scale, and the GEDI02_B monthly product used in GEE is a rasterized representation of these discrete footprint observations. Therefore, valid GEDI values occur only at GEDI footprint locations, while non-footprint locations remain masked.

Before comparison with ChinaTCC30, the quality-filtered GEDI canopy cover observations were aggregated to the 30 m ChinaTCC30 grid using ‘reduceResolution’ with a mean reducer in GEE.

Specifically, valid GEDI canopy cover values contributing to each 30 m ChinaTCC30 grid cell were averaged to generate a 30 m GEDI-based reference layer, while grid cells without valid GEDI observations remained masked. The corresponding ChinaTCC30 values were then extracted from the same 30 m grid cells for comparison. This procedure allowed the GEDI–ChinaTCC30 comparison to be conducted at a consistent 30 m grid scale.

We also considered the potential influence of GEDI geolocation uncertainty by applying strict quality-control filters before aggregation, including retaining observations with ‘l2b_quality_flag = 1’ and ‘degrade_flag = 0’, and selecting nighttime observations with ‘solar_elevation < 0’. The use of ‘degrade_flag = 0’ helped exclude observations potentially affected by degraded pointing or geolocation quality.

Where changed:

Manuscript —2.2.3 Spaceborne Lidar derived canopy cover data — Page 7, Lines 184-194. [MODIFIED]

To further evaluate the spatial consistency and reliability of ChinaTCC30, we utilized the Global Ecosystem Dynamics Investigation (GEDI) Level 2B Canopy Cover product (GEDI02_B) spanning from 2019 to 2024 as an independent reference dataset. Specifically, we extracted the directional gap probability-derived canopy cover metric at the nominal 25 m footprint resolution. Given the inherent noise in spaceborne LiDAR waveforms, a stringent multi-step quality control procedure was implemented within the Google Earth Engine (GEE) platform to screen the GEDI footprints. Initially, we retained only high-quality footprints by applying standard quality masks, selecting data points with l2b_quality_flag equal to 1 and degrade_flag equal to 0. Additionally, to avoid the interference caused by solar background illumination, we exclusively selected nighttime observations with a solar elevation angle below zero (solar_elevation < 0). The quality-filtered GEDI canopy cover observations were aggregated to the 30 m ChinaTCC30 grid using a mean reducer in GEE. Through these procedures, we obtained a nationwide distribution of high-quality GEDI footprints (see Fig. S10 for their spatial distribution), which were used only for independent product comparison and spatial consistency assessment.

Specific comment 3

The pseudo-samples for the 1975/1980 baselines are based only on natural even-aged pure stands, whereas many of China's forests are dominated by plantations and mixed stands. Please discuss in the limitations the potential systematic bias this may introduce in major afforestation regions (e.g., the Three-North and Grain-for-Green areas).

Response: We thank the reviewer for this important comment. We agree that pseudo-samples derived from natural even-aged pure stands may introduce systematic uncertainty when applied to forests with different structural characteristics, such as plantations, shelterbelts, and mixed stands. In this study, stable natural even-aged pure stands were used because the stand-growth-based historical backcasting procedure requires relatively homogeneous stand structure and a clear stand age-density relationship. Such stands provide a more ecologically interpretable basis for estimating historical canopy conditions under model constraints.

We have revised the limitation discussion to clarify the potential bias associated with this sample design. Specifically, growth relationships derived mainly from natural even-aged pure stands may not fully represent plantation-dominated shelterbelts or structurally mixed forests, which could

increase the uncertainty of the 1975 and 1980 baseline estimates in regions where such forest types were already present. At the same time, we clarified that the Grain-for-Green Program began after the 1975 and 1980 baseline periods, with pilot implementation starting in 1999 and large-scale implementation occurring later. Therefore, Grain-for-Green itself would not directly introduce bias into the 1975/1980 baseline reconstruction, although post-1999 afforestation should be considered when interpreting subsequent temporal changes. We now discuss these limitations more explicitly in the revised manuscript and note that early baseline estimates in plantation-dominated shelterbelt or mixed-stand regions should be interpreted with caution.

Where changed:

Manuscript —4.2 Limitations and future improvement — Page 36, Lines 777-782. [MODIFIED]

Another source of uncertainty is related to the ecological scope of the pseudo-samples. The historical backtracking procedure was based on the selected NFI sample plots, because these stands provide relatively homogeneous structure and clearer stand age-density relationships. However, growth relationships derived from such stands may not fully represent plantation-dominated shelterbelts, structurally mixed stands, or forests strongly affected by management. Therefore, the 1975 and 1980 baseline estimates may be more uncertain in regions where these forest types were already present, such as parts of the Three-North Shelterbelt region.

Specific comment 4

Section 2.3 does not describe the trend analysis method, yet Section 3.5 and Fig. 10 use Sen's slope and significance testing ($p < 0.05$). Please add the trend detection method used in the Methods.

Response: We thank the reviewer for pointing out this omission. In the revised manuscript, we have added the pixel-wise trend detection method to Section 2.3.7. Specifically, the Theil-Sen median slope estimator was used to quantify the magnitude and direction of temporal changes in TCC, while the Mann-Kendall test was used to assess the statistical significance of monotonic trends. Pixels were classified according to the sign of the Theil-Sen slope and the Mann-Kendall p value, with $p < 0.05$ used as the threshold for significant trends. The detailed formulas and variable definitions for the Theil-Sen estimator and Mann-Kendall test were provided in Supplementary Text S5.

Where changed:

Manuscript —2.3.7 TCC change trend analysis — Pages 11-12, Lines 325-332. [MODIFIED]

A pixel-wise trend analysis was conducted to evaluate changes in TCC from 1985–2024. The Theil–Sen median slope estimator (Sen, 1968) was used to quantify the magnitude and direction of TCC trends, while the Mann–Kendall test (Mann, 1945) was applied to evaluate their statistical significance. The Theil–Sen estimator provides a robust non-parametric estimate of the temporal slope by calculating the median of all pairwise slopes in the time series, and the Mann–Kendall test identifies whether the monotonic trend is statistically significant (see Text S5 for the details). Pixels were then grouped according to trend direction and significance level into five categories: significant gain ($p < 0.05$ and trend > 0), significant loss ($p < 0.05$ and trend < 0), non-significant gain ($p \geq 0.05$ and trend > 0), non-significant loss ($p \geq 0.05$ and trend < 0), and unchanged ($p \geq 0.05$ and trend = 0).

Supplement —Text S5 — Pages 2-3. [NEWLY ADDED]

Text S5 | Theil-Sen trend analysis and Mann-Kendall test

A pixel-wise trend analysis was conducted to quantify long-term changes in annual tree canopy cover (TCC). The Theil–Sen estimator was used to estimate the magnitude and direction of the temporal trend as the median of all pairwise slopes in each TCC time series (Eq. S1). This non-parametric estimator provides a robust estimate of temporal change and has been widely used for environmental time-series analysis (Sen, 1968). The Mann–Kendall test was used to assess the statistical significance of the monotonic trend (Eqs. S2–S5). This rank-based non-parametric test was proposed by Mann (1945) and further developed by Kendall (1975). Based on the sign of the Theil–Sen slope and the Mann–Kendall p value, pixels were classified into five categories: significant gain ($p < 0.05$ and slope > 0), significant loss ($p < 0.05$ and slope < 0), non-significant gain ($p \geq 0.05$ and slope > 0), non-significant loss ($p \geq 0.05$ and slope < 0), and unchanged ($p \geq 0.05$ and slope $= 0$).

$$\rho = \text{median}\left(\frac{X_j - X_i}{j - i}\right) \quad (S1)$$

where ρ denotes the Theil–Sen slope estimate; X_i and X_j represent the annual TCC values of a given pixel at time steps i and j , respectively; i and j are temporal indices, with $1 \leq i < j \leq n$; and n is the number of annual observations. The median is calculated from the slopes of all possible pairs of observations. A positive ρ indicates an increasing TCC trend, whereas a negative ρ indicates a decreasing trend.

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i) \quad (S2)$$

where S is the Mann–Kendall test statistic, calculated by summing the signs of pairwise differences across the TCC time series. The variables x_i , x_j , i , j , and n are defined as in Eq. (S1), and sgn denotes the sign function defined in Eq. (S3).

$$\text{sgn}(x_j - x_i) = \begin{cases} 1, & x_j - x_i > 0 \\ 0, & x_j - x_i = 0 \\ -1, & x_j - x_i < 0 \end{cases} \quad (S3)$$

where $\text{sgn}(x_j - x_i)$ equals 1 when the later observation is greater than the earlier observation, 0 when the two observations are equal, and -1 when the later observation is smaller than the earlier observation.

$$\text{VAR}(S) = \frac{(n(n-1)(2n+5) - \sum_{i=1}^m t_i(t_i-1)(2t_i+5))}{18} \quad (S4)$$

where $\text{VAR}(S)$ is the variance of the Mann–Kendall statistic; m is the number of groups of tied TCC values; and t_i is the number of observations in the i -th tied group. The tie-correction term accounts for repeated values within the TCC time series.

$$Z_S = \begin{cases} \frac{S-1}{\sqrt{\text{VAR}(S)}}, & S > 0 \\ 0, & S = 0 \\ \frac{S+1}{\sqrt{\text{VAR}(S)}}, & S < 0 \end{cases} \quad (S5)$$

where Z_S is the standardized Mann–Kendall test statistic, S is the Mann–Kendall statistic, and $\text{VAR}(S)$ is its variance. Under the null hypothesis of no monotonic trend, Z_S approximately follows

a standard normal distribution. The corresponding two-sided p value was derived from Z_S and used to determine trend significance.

Specific comment 5

The compositing windows are 1972–1977 for the 1975 baseline and 1978–1984 for the 1980 baseline, which are relatively wide. Please discuss how these windows affect the temporal representativeness of the baseline years.

Response: We thank the reviewer for this important comment. We agree that the relatively wide MSS compositing windows may affect the temporal representativeness of the 1975 and 1980 baseline estimates, and this point should be clarified.

The use of multi-year compositing windows was mainly constrained by the limited availability and uneven spatial coverage of early Landsat MSS observations. For the 1975 and 1980 historical baselines, single-year MSS observations were insufficient to generate spatially continuous and reliable national-scale composites for China. Therefore, images from 1972–1977 and 1978–1984 were composited to represent the 1975 and 1980 historical baseline periods, respectively. This strategy increased the number of valid observations, reduced the influence of clouds, shadows, and data gaps, and improved the spatial completeness of the MSS-based inputs.

We also note that the pseudo-samples used for historical backcasting were selected as temporally stable samples. These stable samples were intended to reduce the influence of short-term land-cover changes, disturbances, or abrupt canopy changes within the compositing windows. Therefore, the resulting baselines primarily represent relatively stable historical canopy conditions around the target years, rather than transient changes occurring within individual years.

Nevertheless, we acknowledge that the wide compositing windows may smooth or obscure rapid changes in TCC that occurred within the corresponding periods, especially in areas affected by disturbance, harvesting, afforestation, or land-use conversion. Therefore, the 1975 and 1980 products should be interpreted as historical baseline estimates representing conditions around the target periods, rather than as exact single-year observations. We have added this clarification and limitation to the revised manuscript.

Where changed:

Manuscript —4.2 Limitations and future improvement — Page 36, Lines 782-786. [MODIFIED]

In addition, because the historical MSS composites were constructed from relatively broad time windows (1972-1977 for the nominal 1975 baseline and 1978-1984 for the nominal 1980 baseline), these products should be interpreted as multi-year baseline estimates centered on the nominal years rather than strict single-year observations. This design improves spatial completeness under sparse early Landsat imagery coverage, but it may reduce temporal specificity in areas where rapid land-cover changes occurred within the compositing windows.

Specific comment 6

Fig. 10 is titled 1985–2024, while Section 3.5 discusses 1975–2024. Please clarify whether the pixel-level trend analysis includes the 1975/1980 baselines, and keep this consistent throughout.

Response: We thank the reviewer for pointing out this inconsistency. We agree that the temporal periods used for different temporal analyses should be stated consistently. In this study, the pixel-level trend analysis was conducted only for the continuous annual TCC time series from 1985 to

2024. The 1975 and 1980 products are historical baseline estimates represented by two individual raster layers, rather than continuous annual observations. Therefore, they were not included in the Theil–Sen slope estimation or the Mann–Kendall significance test, which require a continuous annual time series. In contrast, the center-of-gravity analysis was used to characterize the broader spatiotemporal shift of national TCC from 1975 to 2024. For this analysis, we included the 1975 and 1980 historical baseline layers and the subsequent annual maps summarized at a five-year interval. Thus, the pixel-level trend analysis covers 1985–2024, whereas the center-of-gravity analysis covers 1975–2024 at a five-year temporal step.

In the revised manuscript, we have clarified this distinction in the Methods and Results sections and checked the related text and figure captions to ensure consistent use of the corresponding analysis periods.

Where Changed:

Manuscript —3.5 TCC temporal dynamics — Page 20, Line 458. [MODIFIED]

The long-term TCC maps show clear spatiotemporal dynamics of TCC across China over the past four decades (Fig. 9).

Manuscript —3.5 TCC temporal dynamics — Page 20, Lines 473-474. [MODIFIED]

To further characterize the long-term greening trend, a pixel-wise trend analysis of TCC was conducted from 1985 to 2024 (Fig. 10).

Specific comment 7

Section 3.5 states a "westward shift of the center-of-gravity trajectory," while the Conclusion states a "southwestward shift." Please make these consistent.

Response: We thank the reviewer for pointing out this inconsistency. We have carefully checked the manuscript and revised the relevant descriptions in Section 3.5 and the Conclusion to consistently describe the center-of-gravity trajectory as a southwestward shift.

Where changed:

Manuscript —3.5 TCC temporal dynamics — Page 20, Lines 466-469. [MODIFIED]

The center-of-gravity trajectory also reveals a distinct southwestward shift during 1975–2024 (Fig. 9b), indicating that TCC expansion was spatially unbalanced across China. In particular, the relative increase in western and inland regions exceeded that in the eastern part of the country, resulting in a gradual southwestward displacement of the TCC centroid.

Specific comment 8

The title of panel e is missing in Figure 12.

Response: We thank the reviewer for pointing this out. We have added the missing title for panel (e) in Figure 12.

Where changed:

Manuscript —3.5 TCC temporal dynamics — Page 26, Lines 557-561. [MODIFIED]

Figure 12. Validation of four tree canopy cover products against photo-interpreted reference data. Scatter density plots comparing the photo-interpreted tree cover with (a) MODIS VCF, (b) DWTP, (c) CATCD, and (d) ChinaTCC30 (Ours). (e) Distributions of product-derived TCC within different photo-interpreted TCC intervals, shown using grouped boxplots for the four products.

Specific comment 9

Figure numbers are missing in Section 4.2 (Line 670-671); "Fig. a9–10," etc. should read "Fig. 16 a9–10."?

Response: We thank the reviewer for pointing this out. We have corrected the missing figure numbers in Section 4.2.

Where changed:

Manuscript —4.2 Limitations and future improvement — Page 35, Line 738. [MODIFIED]

This strategy effectively reduced false positives of TCC over bare surfaces, as well as the misclassification of non-tree areas as TCC in urban impervious surfaces and agroforestry mosaics (Fig. 16a9–10, b9–10).

Specific comment 10

The caption of Fig. 8, "across nation four inventory periods," should be revised to "across four national inventory periods."

Response: We thank the reviewer for pointing this out. We have revised the caption of Fig. 8 by changing “across nation four inventory periods” to “across four national inventory periods.”

Where changed:

Manuscript —3.4.2 External validation using NFI data — Page 20, Line 456. [MODIFIED]

Figure 8. Validation of the ChinaTCC30 product against NFI data across four national inventory periods (6th–9th; n = 11,618 NFI sample plots).

Specific comment 11

Typo: "Radom Forest" in Section 3.3 (Line 340) should be "Random Forest".

Response: Response: Thank you for pointing out this typo. We have corrected “Radom Forest” to “Random Forest” in Section 3.3.

Where changed:

Manuscript —3.3 Training and optimization of retrieval models — Page 17, Line 396. [MODIFIED]

K-means-based diversity active learning-Random Forest (KDAL-RF)

Authors' Response to Reviewer 3

General comments

The dataset has potential value for long-term tree canopy cover monitoring in China. However, several methodological and data-consistency issues should be clarified before the reliability and originality of ChinaTCC30 can be fully assessed. My main concerns relate to the distinction from the existing CATCD framework, the historical reconstruction, temporal transferability of the model, auxiliary-data consistency, and temporal breakpoint correction.

Response: We sincerely thank the reviewer for the assessment of our manuscript and for recognizing the value of the ChinaTCC30 dataset. We also appreciate the constructive comments and suggestions. In the revised manuscript, we have carefully addressed all the issues raised by the reviewer, and the detailed point-by-point responses are provided below.

Major comment 1

The methodological novelty relative to CATCD should be clarified more explicitly.

The 1985–2024 mapping framework shares several key components with Cai et al. (2024), including nationwide 30 m VHR interpretation plots, Random Forest with 500 trees, five-fold ensemble prediction, averaging of the five prediction layers, and the use of their standard deviation as an uncertainty layer. CATCD had already implemented this RF ensemble framework. ChinaTCC30 introduces additional components, including the 10×10 interpretation grid, KDAL, historical backcasting, ALCC predictors, and temporal breakpoint correction.

I suggest that the authors explicitly distinguish which methodological components are inherited or adapted from CATCD and which are newly developed in this study. A concise comparison with Cai et al. (2024) would help readers assess the methodological contribution of ChinaTCC30.

Relevant manuscript locations: Introduction, Lines ~95–105; Sect. 2.3.1–2.3.2; Sect. 2.3.6.

Response: We thank the reviewer for this helpful suggestion. We agree that the methodological relationship between ChinaTCC30 and CATCD should be clarified more explicitly. In the revised manuscript, we have added a concise statement at the beginning of the mapping-framework section to distinguish the components adapted from previous annual TCC mapping efforts from the methodological extensions introduced in this study.

Specifically, we clarified that the 1985–2024 annual mapping framework follows the RF ensemble strategy used in CATCD, including Random Forest modeling, five-fold ensemble prediction, pixel-wise averaging of the five prediction layers, and the use of their standard deviation as a pixel-level uncertainty estimate. We also clarified that the main methodological extensions of ChinaTCC30 include the 10×10 point-grid VHR interpretation protocol for deriving continuous TCC reference values, K-means-based Diversity Active Learning (KDAL) for training-sample optimization, the use of ALCC land-cover information as an auxiliary predictor, stand-growth-based historical backtracking, MSS-based reconstruction of the 1975 and 1980 baselines, and stable-reference-based temporal breakpoint correction.

To avoid repeatedly restating this comparison in each subsection, we summarized these differences in the opening paragraph of the mapping framework section. This revision helps readers assess which parts of the framework were adapted from CATCD and which components represent new developments in ChinaTCC30.

Where changed:

Manuscript —2.3.6 TCC mapping approach— Page 11, Lines 296-302. [MODIFIED]

Compared with previous annual TCC mapping efforts that mainly focused on the Landsat TM era and later periods (Hansen et al., 2013; Sexton et al., 2013; Cai et al., 2024), this study further integrates active-learning-based sample optimization, stand-growth-based historical backtracking, MSS-based early-period reconstruction to generate annual and historical estimates of TCC across China from 1970s to 2024, respectively (Fig. 3). Historical 60 m TCC for 1975 and 1980 was reconstructed from Landsat MSS imagery and pseudo-samples developed within a historical backtracking framework, whereas annual 30 m TCC maps for 1985–2024 were produced from Landsat imagery and visually interpreted TCC samples (the division and spatial distribution of all samples are shown in Supplementary Table S6 and Fig. S9).

Manuscript —2.3.6 TCC mapping approach— Page 11, Lines 318-323. [MODIFIED]

After annual TCC mapping, a $0.5^{\circ} \times 0.5^{\circ}$ grid-based temporal jump check was conducted across the periods surrounding Landsat sensor transitions to identify potential sensor-transition discontinuities. For tiles showing anomalous jumps in the transition year, a local correction based on linear regression of the temporal trajectories of stable reference pixels was applied only to pixels with unchanged land-cover classes and no detected forest disturbance within the correction window (further details on the identification and correction of affected tiles are provided in Text S4).

Major comment 2

The reconstruction of the 1975 and 1980 historical baselines requires additional methodological detail.

The manuscript states that historical TCC pseudo-samples were generated by back-calculating stand age and stem density and combining these variables with climate and topography. However, several important details remain unclear, including how “stable natural even-aged pure stands” were identified, the source and reference year of the stand-age data.

In addition, 1,700 MSS scenes acquired during 1972–1977 and 1978–1984 were composited to represent the nominal 1975 and 1980 baselines, respectively. Please clarify the compositing procedure. For example, were observations from all years weighted equally, how were multiple observations combined, and what fraction of the final mosaics was contributed by each acquisition year? This information is important because the “1975” and “1980” maps represent multi-year epochs rather than observations from single years.

Relevant manuscript locations: Sect. 2.3.3, Lines ~220–240; Sect. 2.3.6, Lines ~263–273.

Response: We thank the reviewer for this important and constructive comment. In the revised manuscript, we have clarified and updated the selection procedure for historical backtracking sample plots. To increase the number of eligible samples while maintaining ecological consistency, we

adjusted the previous strict definition to a sequential screening procedure based on plot-level NFI attributes. The revised screening incorporated land-cover type, species composition, stand age, degree of naturalness, forest health status, stand structural uniformity, and site-condition class. The detailed selection criteria are provided in Supplementary Text S1.

We have also clarified the source and reference year of the stand-age data. Stand-age information was obtained from the NFI records, and the corresponding inventory year was used as the reference year for back-calculation. Historical stand ages for 1975 and 1980 were then reconstructed according to the time interval between the inventory reference year and the target historical year. This information has been clarified in the NFI data description in the Materials and Methods section.

Regarding the MSS compositing procedure, we have revised the Landsat data section to clarify that the 1975 and 1980 MSS mosaics were generated as multi-year baseline composites rather than strict single-year images. For the nominal 1975 baseline, quality-screened MSS observations acquired during 1972-1977 were used; for the nominal 1980 baseline, quality-screened MSS observations acquired during 1978-1984 were used. Observations from different years within each compositing window were not assigned explicit temporal weights. Instead, after pixel-level QA screening, all valid observations within each baseline window were combined using a pixel-wise band-wise median compositing approach. Remaining gaps were supplemented using the Global Land Survey (GLS) image collection. The MSS-specific quality-control, compositing, and gap-filling procedure has been briefly described in the Landsat data section, with further details provided in Supplementary Text S3. In addition, the GLS images used for gap filling are distributed as pre-composited global mosaics and do not provide pixel-level source-year or scene-ID information in the version used in this study. Therefore, we could not reliably quantify the exact contribution of each acquisition year to the GLS-filled portions of the final mosaics. We therefore did not report year-specific contribution fractions for the GLS component, but clarified that the 1975 and 1980 products should be interpreted as multi-year baseline composites rather than single-year observations, and further discussed in the revised Discussion section how this compositing strategy may affect the temporal representativeness and uncertainty of the early TCC baselines.

Where changed:

Supplement —Text S1 — Page 1. [NEWLY ADDED]

Text S1 | Selection of sample plots for historical backtracking

Historical backtracking sample plots were selected through a sequential screening procedure using plot-level attributes available from the NFI. The screening incorporated land-cover type, species composition, stand age, degree of naturalness, and forest health status. A plot was retained only when it satisfied all of the following criteria: (1) Temporal land-cover stability. The plot was consistently recorded as the same forest land-cover type throughout all NFI survey periods available for the study period. Plots with land-cover conversion, inconsistent land-cover records, or missing observations were excluded. (2) Pure species composition. The dominant tree species accounted for more than 90% of the recorded stand composition. Plots with mixed species composition or a change in the dominant species group among inventory observations were excluded. (3) Relatively uniform stand structure. Plots with a relatively small coefficient of variation in DBH were considered to have a more uniform size structure and were retained as candidate backtracking plots. In contrast, plots with large DBH variability were excluded because they may indicate heterogeneous size structure. (4) Low disturbance and healthy stand condition. Forest health records were used to exclude plots

affected by substantial disturbance, degradation, or damage. Only plots with a low recorded disturbance level and without evidence of major canopy-altering events were retained. Finally, to reduce confounding effects arising from environmental heterogeneity, candidate plots were further stratified according to the site-condition class. For each species or genus, the largest homogeneous subset of plots sharing the same site-condition category was retained. This additional constraint was intended to ensure that variation in stand structure was primarily associated with stand development rather than differences in site conditions.

Supplement —Text S3 — Pages 1-2. [NEWLY ADDED]

Text S3 | Landsat MSS compositing for the 1975 and 1980 historical baselines

For the 1975 and 1980 historical reconstructions, Landsat MSS imagery was processed separately from the later Landsat TM/ETM+/OLI-based annual mapping workflow because of differences in sensor characteristics, spectral configuration, data quality, and the availability of QA information. MSS observations acquired during 1972-1977 and 1978-1984 were used to construct the nominal 1975 and 1980 baseline composites, respectively. Therefore, these two products should be interpreted as multi-year baseline epochs centered on the nominal years rather than observations from single years. Before compositing, MSS imagery was subjected to pixel-level quality screening. Fill pixels and cloud-contaminated pixels were removed using the QA_PIXEL band, and radiometrically saturated pixels were excluded using the QA_RADSAT band. The retained MSS spectral bands were then used for compositing. For each baseline epoch, Landsat MSS Collection 2 Tier 1 images were used as the primary data source because of their higher geometric quality, and the median value of each MSS band was calculated from all valid observations within the corresponding baseline window. In addition, remaining spatial gaps were supplemented using the Global Land Survey (GLS) image collection. For the 1975 baseline, GLS observations closest to the nominal baseline year were prioritized, whereas GLS observations from the broader 1972-1977 window were used only to fill remaining gaps. The same hierarchical logic was applied to the 1980 baseline reconstruction. In this procedure, all available valid MSS observations were retained, and GLS data were used only as supplementary gap-filling information. The resulting MSS baseline composites were used as spectral inputs for the historical TCC reconstruction models for 1975 and 1980.

Major comment 3

More information is needed on the fitting of the stand age–stem density relationships.

Supplementary Tables S4–S5 provide empirical age–stem-density relationships for nine species or genera, but the sample size used to fit each relationship, the optimization method, and the parameter ranges or boundary constraints are not reported.

*Several fitted coefficients are exactly 100000 or 15000. For example, several power models have a coefficient of 100000, while the exponential models for *Pinus densata* and *Betula platyphylla* have coefficients of exactly 15000. These repeated exact values raise the possibility that some fitted parameters reached imposed optimization bounds.*

Please report the sample sizes, fitting algorithm, initial values where relevant, and parameter

bounds, and confirm whether any fitted parameters reached the prescribed limits. This is particularly important because these relationships are subsequently used to reconstruct historical stand conditions.

Relevant location: Supplementary Tables S4–S5, pp. 3–4.

Response: We thank the reviewer for this helpful comment. We re-fitted and re-validated the stand age–stem density relationships for each species or genus. We have added more details on the fitting procedure, including the candidate model forms, fitting algorithm, sample size, initial values, parameter bounds, and independent validation results in the Supplementary. Specifically, three empirical models, including exponential decay, power-law decay, and logarithmic decline models, were fitted using the nonlinear least-squares regression. The best model for each species or genus was selected based on AIC and further evaluated using an independent test set. We also checked whether fitted parameters reached the prescribed optimization bounds and clarified this information in the revised supplementary tables.

Where changed:

Manuscript —3.1 Construction of pseudo-sample data based on stand factors for the years 1975 and 1980 — Page 14, Lines 352-354. [MODIFIED]

Model robustness was further assessed using an independent 20% hold-out test set, yielding R^2 values of 0.52–0.68 and r of 0.74–0.85 (Fig. S1; Table S5).

Manuscript —3.1 Construction of pseudo-sample data based on stand factors for the years 1975 and 1980 — Page 14, Lines 357-358. [MODIFIED]

A 5-fold cross-validation confirmed the model's accuracy, yielding an R^2 of 0.91, an RMSE of 10.84%, a MAE of 6.33%, and a bias of -0.15% (Fig. 4b).

Manuscript —3.1 Construction of pseudo-sample data based on stand factors for the years 1975 and 1980 — Page 15, Lines 380-384. [MODIFIED]

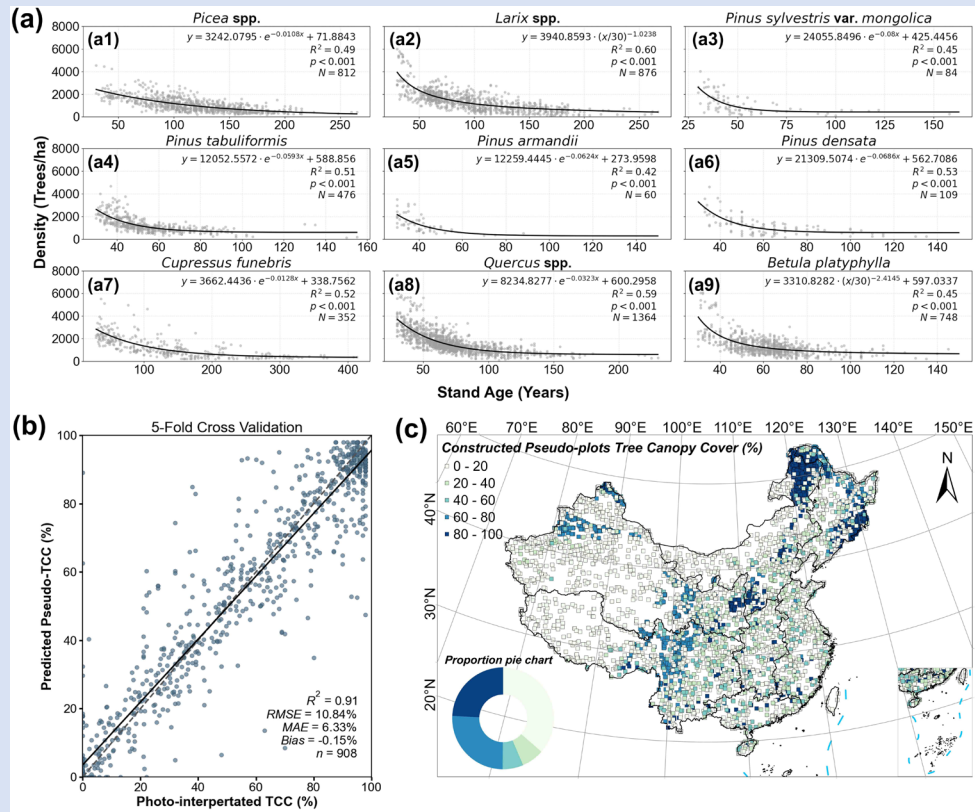


Figure 4. Generation and deduction model validation for historical TCC pseudo-plots. (a) Stand age-density self-thinning curves for nine major tree species/genera (a1-a9), serving as the parameter basis for the deduction algorithm. (b) Performance of the pseudo-plot reconstruction model based on 5-fold cross-validation for the 2018 reference year, where photo-interpreted TCC represents the observed reference TCC and predicted pseudo-TCC represents the model-estimated TCC for the same year. (c) Spatial distribution of the ultimately generated historical pseudo-plots ($n = 7753$).

Manuscript —3.3 Training and optimization of retrieval models— Page 17, Lines 397-400. [MODIFIED]

KDAL-RF reached a near-stable performance level at about 2,420 samples ($R^2 = 0.71$, $RMSE = 17.10\%$), whereas random-sampling RF required about 3,140 samples to achieve a comparable level, corresponding to a reduction of approximately 22% in training sample demand (Fig. 6a, b).

Manuscript —3.3 Training and optimization of retrieval models— Page 17, Lines 406-408. [MODIFIED]

Overall, the active learning strategy achieved higher sample efficiency in both experiments, reducing the required training sample size by about 22% in the pseudo-sample-based experiment and by about 48% in the visually interpreted sample experiment (Fig. 6).

Supplement —Text S2 — Page 1. [NEWLY ADDED]

Text S2 | Fitting of stand age–stem density relationships for historical backtracking

Stand age–stem density relationships were fitted separately for each dominant tree species or species group using bounded nonlinear least-squares regression. For each species, three candidate empirical models were evaluated: an exponential decay model [$y = a \cdot e^{-bx} + c$], a power-law model [$y = a \cdot x^{-b} + c$], and a logarithmic model [$y = a - b \cdot \ln(x)$], where x represents stand age, y represents stem density, and a , b , and c are fitted parameters. For each species, the final stand age–density equation was selected

from the three candidate models according to the minimum Akaike information criterion (AIC). Model performance was evaluated using R^2 , adjusted R^2 , Pearson's correlation coefficient (r).

Major comment 4

The temporal coverage of the auxiliary datasets is inconsistent with their stated use.

Table S1 reports ALCC for 1985–2022 and TerraClimate for 1985–2024. However, the manuscript states that historical climate variables were used for 1975 and 1980, and ALCC is also used to generate zero-TCC pseudo-samples for the historical reconstruction. ALCC is additionally described as an auxiliary predictor in the annual mapping framework.

Please clarify:

- the actual climate dataset and years used for the 1975 and 1980 reconstruction;*
- how ALCC information beginning in 1985 was used to identify persistent bare-land/water areas for 1975 and 1980;*
- how the land-cover predictor was handled for 2023 and 2024, when ALCC is unavailable.*

In particular, the use of post-1985 land-cover information for constructing historical zero-TCC pseudo-samples should be discussed because it may introduce information from later periods into the historical reconstruction.

Relevant manuscript locations: Sect. 2.1.2, Lines ~125–135; Sect. 3.1, around Line 305; Supplement Table S1.

Response: We thank the reviewer for pointing out this inconsistency. We have revised the manuscript and Supplementary Table S1 to clarify the temporal coverage and use of the auxiliary datasets.

First, we clarified that the historical climate variables used for the 1975 and 1980 reconstructions were not from TerraClimate. Instead, they were obtained from the 1901–2025 monthly climate dataset for China at 1000 m resolution (Peng et al., 2019), from which climate variables corresponding to the 1975 and 1980 baseline periods were extracted.

Second, we clarified that ALCC was not used as a direct land-cover map for 1975 or 1980. It was only used to identify candidate locations with a high likelihood of persistent non-canopy cover for generating zero-TCC pseudo-samples. Specifically, pixels consistently classified as water bodies or bare land throughout the 1985–2022 ALCC time series were first selected, and only interior pixels away from class boundaries were retained. These candidates were then further verified using the corresponding Landsat MSS baseline composites: water candidates were retained only when $NDWI > 0$, and bare-land candidates were retained only when $NDVI < 0.10$. Therefore, zero-TCC pseudo-samples were generated only for pixels satisfying both the long-term ALCC stability criterion and the MSS-based spectral verification criterion.

Third, we clarified how the land-cover predictor was handled for 2023 and 2024. Because ALCC was unavailable after 2022, the land-cover predictor for 2023 and 2024 was supplemented using CLCD

data with harmonized land-cover classes. This has been clarified in the section 2.3.3.

Where changed:

Manuscript —2.1.2 Auxiliary data — Page 5, Lines 143-147. [MODIFIED]

Climatic variables, including temperature and precipitation, were obtained from China-specific monthly mean temperature and precipitation datasets with a spatial resolution of approximately 1 km developed by Peng et al (2019). Annual temperature and precipitation predictors were derived from the monthly climate layers and used for plot-scale pseudo-sample model construction and prediction (see Table S1 for further details).

Manuscript —2.1.2 Auxiliary data— Page 5, Lines 137-141. [MODIFIED]

For the 1985–2022 period, the Annual Land Cover of China (ALCC) dataset (Li et al., 2026) was used as the primary land-cover source. Because ALCC does not cover 2023 and 2024, the China Land Cover Dataset (CLCD) (Yang and Huang, 2021), which uses a land-cover classification scheme consistent with ALCC, was used as a supplementary source to provide land-cover information for these two years.

Manuscript —2.3.3 Construction of Pseudo-plot Samples for Historical Baseline Reconstruction— Pages 9-10, Lines 263-276. [MODIFIED]

Furthermore, to ensure that the pseudo-sample dataset represented the full landscape continuum from non-canopy surfaces to high-TCC forests, reliable zero-value pseudo-plots (0% TCC) were generated for the 1975 and 1980 baselines. Because direct TCC reference observations were unavailable for these historical periods, these zero-TCC pseudo-plots were generated at locations with a high likelihood of persistent non-canopy cover, namely persistent water bodies and bare land. These locations were identified using a conservative two-step screening procedure. First, the 1985–2022 ALCC time series (Li et al., 2026) was used to identify pixels that were consistently classified as water bodies or bare land during all years covered by the land-cover dataset. To reduce mixed-pixel effects along class boundaries, only interior pixels that were not adjacent to other land-cover types were retained as candidate persistent non-canopy locations. Second, these candidates were further verified using spectral features from the Landsat MSS composites for the corresponding historical baseline periods. Candidate persistent water-body pixels were retained only when their NDWI was greater than 0, whereas candidate persistent bare-land pixels were retained only when their NDVI was lower than 0.10 in the corresponding MSS composite. This spectral verification step was based on the established use of NDWI for identifying open water and NDVI for identifying bare surfaces (McFeeters, 1996; Tucker, 1979; Mzid et al., 2021). Only pixels satisfying both the long-term ALCC stability criterion and the MSS-based spectral verification criterion were used to generate zero-TCC pseudo-plots for the 1975 and 1980 reconstructions.

Major comment 5

Temporal transfer of the model from recent reference labels to 1985–2014 has not been sufficiently evaluated.

The 10,000 visually interpreted plots were derived predominantly from VHR imagery acquired during 2015–2024, and Landsat predictors were extracted from the corresponding label years. The resulting model is nevertheless used to generate annual TCC maps back to 1985.

This implies substantial temporal transfer from predominantly recent training labels to earlier Landsat periods, particularly 1985–2014. Although sensor harmonization may reduce spectral differences, it does not by itself demonstrate that the TCC–predictor relationship remains transferable across several decades.

Please provide a more direct evaluation of model performance outside the primary reference-label period, preferably using temporally independent observations from earlier years. At minimum, the manuscript should discuss this temporal extrapolation more explicitly and quantify performance for earlier versus later periods where reference data are available. I also suggest conducting a Landsat spectral-consistency assessment across the major sensor periods, using temporally stable reference targets or other suitable invariant samples. Such an analysis could help demonstrate whether the harmonized spectral feature distributions remain sufficiently consistent for transferring a model trained predominantly on recent observations to earlier decades.

Relevant manuscript locations: Sect. 2.2.1, Lines ~145–150; Sect. 2.3.6, Lines ~263–280.

Response: We thank the reviewer for this important comment. We agree that the temporal transferability of a model trained mainly with recent visually interpreted labels should be more explicitly evaluated and discussed. We have revised Sect. 2.2.1, Sect. 2.3.6, and the related validation and discussion sections to clarify the temporal distribution of the reference labels and to provide additional evidence for model performance outside the primary VHR-label period.

First, we have added a more explicit description of the temporal-transfer issue. The visually interpreted TCC reference plots were mainly derived from VHR imagery acquired during 2015–2024, and the corresponding Landsat predictors were extracted for the label years. Therefore, applying the trained model to earlier Landsat periods involves temporal transfer of the learned TCC–predictor relationship. We have clarified this point in the revised manuscript.

Second, to provide a more direct evaluation outside the primary VHR-label period, we quantified product performance using temporally independent NFI observations from different inventory periods. The comparison showed that ChinaTCC30 maintained reasonable agreement with NFI-based canopy-cover observations in earlier periods, although the agreement for the 6th NFI period was weaker than that for later inventory periods. Specifically, the validation results were $R^2 = 0.66$ and $RMSE = 18.81\%$ for the 6th NFI period (1999–2003), $R^2 = 0.76$ and $RMSE = 16.63\%$ for the 7th NFI period (2004–2008), $R^2 = 0.79$ and $RMSE = 15.58\%$ for the 8th NFI period (2009–2013), and $R^2 = 0.82$ and $RMSE = 14.68\%$ for the 9th NFI period (2014–2018). These results indicate that the product retained meaningful predictive skill before the primary VHR-label period.

Third, following the reviewer’s suggestion, we have clarified the spectral-consistency basis of the Landsat predictors used in this study. The annual Landsat features were derived from the China Earth Observation Data Cube, a seamless annual leaf-on Landsat composite dataset for China from 1985 to 2023. This dataset was specifically designed to harmonize observations across multiple Landsat sensors and to address cloud and shadow contamination, reflectance consistency, and data gaps. Cai et al. (2025) reported high consistency between gap-filled proxy values and actual Landsat references for both stable and dynamic pixel sequences, as well as strong agreement between the annual composites and representative Landsat reference images. In the revised version, we have cited this dataset paper and clarified that the use of these harmonized annual Landsat composites provides a more consistent spectral basis for applying the TCC model across major Landsat sensor periods.

Where changed:

Manuscript —4.2 Limitations and future improvement— Page 35, Lines 757-768. [NEWLY ADDED]

Fourth, the temporal representativeness of the visually interpreted samples may also introduce uncertainty into annual TCC mapping. The interpretation samples were mainly derived from VHR imagery acquired during 2015–2024, whereas ChinaTCC30 provides annual TCC estimates for 1985–2024. In the annual mapping framework, the visually interpreted TCC sample pool was used as the reference sample set, while predictor variables were extracted from the corresponding annual Landsat composites for each target year. This strategy assumes that the relationship between TCC and the predictor space is sufficiently stable and transferable across years after consistent preprocessing. The use of QA-based cloud and shadow masking, cross-sensor spectral harmonization, medoid compositing, and temporal gap filling helped reduce sensor-related inconsistencies and improve temporal consistency (Flood, 2013; Roy et al., 2016; Wulder et al., 2019; Cai et al., 2025a). Nevertheless, residual uncertainties may remain because of differences in sensor radiometry, atmospheric correction quality, image availability, phenological conditions, and background land-cover composition. These uncertainties may be more pronounced in early TM-based years, where contemporaneous VHR reference samples are unavailable.

Major comment 6

The temporal breakpoint correction is not sufficiently described to be reproduced.

The manuscript states that stable ALCC pixels and local temporal trajectories were used to identify and correct non-physical breakpoints. However, the breakpoint-detection algorithm, temporal window, significance threshold, definition of stable pixels, and correction equation are not provided.

This distinction is important because genuine forest disturbances, such as harvesting, fire, typhoons, or land conversion, can also produce abrupt changes in TCC.

Please provide the full correction procedure in the Methods and report, at least, the number or proportion of pixels corrected, the years most affected, and the typical correction magnitude. It would also be useful to demonstrate that genuine abrupt canopy disturbances are retained by the correction procedure.

Relevant manuscript location: Discussion, Lines ~630–655. This procedure would be more appropriate as part of the Methods.

Response: We thank the reviewer for this important suggestion. We agree that this post-processing step should be described in the Methods rather than only in the Discussion. In the revised manuscript, we have added a concise description of the temporal breakpoint correction procedure to the Methods, and provided the full procedure in Supplementary Text S4.

Specifically, the procedure now includes three steps: (1) detection of potentially affected tiles using $0.5^\circ \times 0.5^\circ$ tiles and two consecutive three-year windows before and after Landsat sensor transitions; tiles were flagged when the inter-window mean TCC difference exceeded the mean difference across all tiles by more than three standard deviations; (2) selection of stable reference pixels within flagged

tiles, defined as pixels with unchanged ALCC land-cover classes and no detected forest disturbance during the correction and transition-assessment periods; and (3) local level-shift correction based on linear regression of the stable-pixel TCC trajectory before and after the transition. The estimated tile-specific additive offset was applied only to eligible stable pixels in the pre-transition period, and corrected values were constrained to 0–100%.

Using this procedure, 137 tiles, accounting for 3.18% of all tiles, were identified as exhibiting anomalous positive TCC jumps, with the detected transition concentrated around 2000. We have visualized these anomalous tiles and their assigned transition year in Supplementary Fig. S11. Importantly, pixels affected by land-cover change or forest disturbance were excluded from the correction and retained unchanged.

Where changed:

Manuscript —2.3.6 TCC mapping approach — Page 11, Lines 318-323. [MODIFIED]

After annual TCC mapping, a $0.5^\circ \times 0.5^\circ$ grid-based temporal jump check was conducted across the periods surrounding Landsat sensor transitions to identify potential sensor-transition discontinuities. For tiles showing anomalous jumps in the transition year, a local correction based on linear regression of the temporal trajectories of stable reference pixels was applied only to pixels with unchanged land-cover classes and no detected forest disturbance within the correction window (further details on the identification and correction of affected tiles are provided in Text S4).

Supplement —Text S4 — Page 2. [NEWLY ADDED]

Text S4 | Post-processing of sensor-transition discontinuities in TCC mapping

A targeted stable-reference-based correction was conducted after annual TCC mapping to reduce localized discontinuities potentially associated with Landsat sensor transitions. The procedure consisted of three steps: (1) Identification of potentially affected tiles. The study area was divided into $0.5^\circ \times 0.5^\circ$ tiles. For each tile, the difference between mean TCC values calculated over two consecutive three-year windows before and after the sensor transition was derived. Tiles were flagged as anomalous when their inter-window TCC difference exceeded the mean difference across all tiles by more than three standard deviations. (2) Selection of stable pixels within flagged tiles. Within each flagged tile, pixels were retained as stable reference pixels only when their ALCC land-cover class remained unchanged throughout the correction and transition-assessment periods and no forest disturbance was detected. Pixels affected by land-cover change or forest disturbance were excluded from subsequent correction (see Table S7 for the details of landcover and forest disturbance data). (3) Local level-shift correction. The annual median TCC of the retained stable pixels was used to construct a local reference trajectory. Linear trends were fitted before and after the transition, and the post-transition trajectory was back-extrapolated to estimate the local level difference at the breakpoint. The resulting tile-specific additive offset was applied only to eligible stable pixels during the pre-transition period, with corrected values constrained to the physical range of 0–100%. Applying this screening procedure, 137 of the tiles (3.18%) were flagged as exhibiting anomalous positive TCC jumps during the sensor-transition period (Fig. S11a, b).

Supplement —Table S7 — Page 6. [MODIFIED]

Table S7. Summary of datasets used for TCC transitions post-processing.

Dataset	Regions	Temporal scope	Spatial resolution	Sources
ALCC	China	1985-2022	30 m	(Li et al., 2026)

CLCD	China	1985, 1990-2024	30 m	(Yang and Huang, 2021)
China Forest Disturbance Observer (1986 - 2020)	China	1986-2020	30 m	https://liuzh822.users.earthengine.app/view/chinaforestdisturbance

Specific comment 1

Title and spatial/temporal resolution

The title describes ChinaTCC30 as an “annual 30 m” dataset from the 1970s to 2024. However, the Methods describe 60 m historical TCC for 1975 and 1980 and annual 30 m TCC only from 1985 onward. The Abstract also describes ChinaTCC30 as a 30 m product for 1975, 1980, and annually from 1985–2024.

Please make the title, abstract, Methods, and released-product description consistent and explicitly state whether the released 1975/1980 rasters were resampled from their native 60 m MSS resolution.

Response: We thank the reviewer for pointing out this inconsistency. We have revised the Abstract, Methods, and released-product description to explicitly distinguish the spatial resolution and temporal coverage of different components of ChinaTCC30.

In the revised manuscript, we clarify that the 1975 and 1980 historical TCC baselines were reconstructed from Landsat MSS imagery at the native 60 m resolution. To ensure grid consistency and pixel alignment with the annual 30 m TCC maps from 1985 to 2024, the 1975 and 1980 baseline layers were resampled to the 30 m ChinaTCC30 grid before release. Therefore, the released ChinaTCC30 dataset provides TCC layers on a consistent 30 m grid for 1975, 1980, and annually from 1985 to 2024, but the 1975 and 1980 layers should be interpreted as resampled products derived from native 60 m MSS-based reconstructions.

We retained the dataset name/title because ChinaTCC30 primarily refers to the annual 30 m mapping component from 1985 to 2024, which constitutes the majority of the dataset, while the 1975 and 1980 layers are historical baseline extensions. We have updated the relevant text in the Abstract, Methods, and product description to avoid implying that.

Where changed:

Manuscript —Abstract — Page 1, Lines 17-20. [MODIFIED]

To address this gap, we developed an integrated framework that combines active learning, Random Forest, and stand-growth-based historical backtracking to generate ChinaTCC30, which includes historical TCC baselines for 1975 and 1980 reconstructed from Landsat MSS imagery at 60 m resolution and annual 30 m TCC maps from 1985 to 2024.

Manuscript —2.3.6 TCC mapping approach — Page 11, Lines 296-302. [MODIFIED]

Compared with previous annual TCC mapping efforts that mainly focused on the Landsat TM era and later periods (Hansen et al., 2013; Sexton et al., 2013; Cai et al., 2024), this study further integrates active-learning-based sample optimization, stand-growth-based historical backtracking, MSS-based early-period reconstruction to generate annual and historical estimates of TCC across China from 1970s to 2024, respectively (Fig. 3). Historical 60 m TCC for 1975 and 1980 was

reconstructed from Landsat MSS imagery and pseudo-samples developed within a historical backtracking framework, whereas annual 30 m TCC maps for 1985–2024 were produced from Landsat imagery and visually interpreted TCC samples (the division and spatial distribution of all samples are shown in Supplementary Table S6 and Fig. S9).

Specific comment 2

The trend and center-of-gravity analyses need methodological descriptions.

The Results report pixel-wise significance at $p < 0.05$ and use mean Sen's slopes, but I could not find a corresponding statistical-method description in the Methods section. Please specify the significance test used, and indicate whether temporal autocorrelation and the large number of simultaneous pixel-level tests were considered.

The center-of-gravity analysis is also presented without an equation. Please specify whether the centroid was weighted by TCC, forest area, or area-weighted TCC, and state the coordinate system used for the calculation.

There is also an internal inconsistency: the Results describe a westward shift, whereas the Conclusion reports a southwestward shift. Please verify the calculation and make the wording consistent.

Relevant manuscript locations: Results, Lines ~405–435; Fig. 9–10; Conclusion, Lines ~715–720.

Response: We thank the reviewer for this helpful comment. We have revised the Methods section to clarify the statistical procedures used for the pixel-wise trend analysis and the center-of-gravity analysis.

For the annual TCC time series from 1985 to 2024, the trend magnitude was estimated using the Theil–Sen median slope estimator, and trend significance was evaluated using the Mann–Kendall test. Pixels were classified according to the sign of Sen's slope and the nominal significance level of the Mann–Kendall test, with $p < 0.05$ used as the significance threshold. Because the Theil–Sen estimator and Mann–Kendall test are standard non-parametric methods, we kept the Methods section concise and provided the detailed equations, variable definitions, significance criterion, and trend-classification rules in Supplementary Text S5. We also clarified that temporal autocorrelation was not explicitly considered in the current analysis, and that the reported significance values represent nominal pixel-wise results.

We have also added the methodological description of the center-of-gravity analysis. The centroid was calculated as a TCC-weighted centroid, using the TCC value of each valid pixel as the weight. The coordinate system used for the calculation has been specified in the revised manuscript. In addition, we rechecked the center-of-gravity results and revised the wording in the Results and Conclusion to ensure that the reported shift direction is internally consistent.

Manuscript —2.3.7 TCC change trend analysis — Pages 11-12, Lines 325-332. [NEWLY ADDED]

A pixel-wise trend analysis was conducted to evaluate changes in TCC from 1985–2024. The Theil–Sen median slope analysis (Sen, 1968) was used to quantify the magnitude and direction of TCC trends, while the Mann–Kendall test (Mann, 1945) was applied to evaluate their statistical

significance. The Theil–Sen method provides a robust non-parametric estimate of the temporal slope by calculating the median of all pairwise slopes in the time series, and the Mann–Kendall test identifies whether the monotonic trend is statistically significant (see Text S5 for further details). Pixels were then grouped according to trend direction and significance level into five categories: significant gain ($p < 0.05$ and trend > 0), significant loss ($p < 0.05$ and trend < 0), non-significant gain ($p \geq 0.05$ and trend > 0), non-significant loss ($p \geq 0.05$ and trend < 0), and unchanged ($p \geq 0.05$ and trend $= 0$).

To examine the spatiotemporal movement of TCC from 1975–2024, we calculated the annual center of gravity of TCC within the study area using the following equations:

$$G_x = \frac{\sum_{i=1}^n \sum_{j=1}^m (TCC_{ij} \times \text{Lon}_{ij})}{\sum_{i=1}^n \sum_{j=1}^m (TCC_{ij})} \quad (6)$$

$$G_y = \frac{\sum_{i=1}^n \sum_{j=1}^m (TCC_{ij} \times \text{Lat}_{ij})}{\sum_{i=1}^n \sum_{j=1}^m (TCC_{ij})} \quad (7)$$

where G_x and G_y are the longitude and latitude of the TCC gravity center, respectively; n and m denote the numbers of raster rows and columns; Lon_{ij} and Lat_{ij} represent the geographic coordinates of the center of the raster cell at row i and column j ; and TCC_{ij} is the corresponding TCC value.

Supplement —Text S5 — Pages 2-3. [NEWLY ADDED]

Text S5 | Theil-Sen trend analysis and Mann-Kendall test

A pixel-wise trend analysis was conducted to quantify long-term changes in annual tree canopy cover (TCC). The Theil–Sen estimator was used to estimate the magnitude and direction of the temporal trend as the median of all pairwise slopes in each TCC time series (Eq. S1). This non-parametric estimator provides a robust estimate of temporal change and has been widely used for environmental time-series analysis (Sen, 1968). The Mann–Kendall test was used to assess the statistical significance of the monotonic trend (Eqs. S2–S5). This rank-based non-parametric test was proposed by Mann (1945) and further developed by Kendall (1975). Based on the sign of the Theil–Sen slope and the Mann–Kendall p value, pixels were classified into five categories: significant gain ($p < 0.05$ and slope > 0), significant loss ($p < 0.05$ and slope < 0), non-significant gain ($p \geq 0.05$ and slope > 0), non-significant loss ($p \geq 0.05$ and slope < 0), and unchanged ($p \geq 0.05$ and slope $= 0$).

$$\rho = \text{median}\left(\frac{X_j - X_i}{j - i}\right) \quad (S1)$$

where ρ denotes the Theil–Sen slope estimate; X_i and X_j represent the annual TCC values of a given pixel at time steps i and j , respectively; i and j are temporal indices, with $1 \leq i < j \leq n$; and n is the number of annual observations. The median is calculated from the slopes of all possible pairs of observations. A positive ρ indicates an increasing TCC trend, whereas a negative ρ indicates a decreasing trend.

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i) \quad (S2)$$

where S is the Mann–Kendall test statistic, calculated by summing the signs of pairwise differences across the TCC time series. The variables x_i , x_j , i , j , and n are defined as in Eq. (S1), and sgn denotes the sign function defined in Eq. (S3).

$$\text{sgn}(x_j - x_i) = \begin{cases} 1, x_j - x_i > 0 \\ 0, x_j - x_i = 0 \\ -1, x_j - x_i < 0 \end{cases} \quad (S3)$$

where $\text{sgn}(x_j - x_i)$ equals 1 when the later observation is greater than the earlier observation, 0 when the two observations are equal, and -1 when the later observation is smaller than the earlier observation.

$$\text{VAR}(S) = \frac{(n(n-1)(2n+5) - \sum_{i=1}^m t_i(t_i-1)(2t_i+5))}{18} \quad (S4)$$

where $\text{VAR}(S)$ is the variance of the Mann–Kendall statistic; m is the number of groups of tied TCC values; and t_i is the number of observations in the i -th tied group. The tie-correction term accounts for repeated values within the TCC time series.

$$Z_S = \begin{cases} \frac{S-1}{\sqrt{\text{VAR}(S)}}, S > 0 \\ 0, S = 0 \\ \frac{S+1}{\sqrt{\text{VAR}(S)}}, S < 0 \end{cases} \quad (S5)$$

where Z_S is the standardized Mann–Kendall test statistic, S is the Mann–Kendall statistic, and $\text{VAR}(S)$ is its variance. Under the null hypothesis of no monotonic trend, Z_S approximately follows a standard normal distribution. The corresponding two-sided p value was derived from Z_S and used to determine trend significance.

Manuscript —3.5 TCC temporal dynamics — Page 20, Lines 466-469. [MODIFIED]

The center-of-gravity trajectory also reveals a distinct southwestward shift during 1975–2024 (Fig. 9b), indicating that TCC expansion was spatially unbalanced across China. In particular, the relative increase in western and inland regions exceeded that in the eastern part of the country, resulting in a gradual southwestward displacement of the TCC centroid.

Specific comment 3

Please clarify the numerical representation of the released uncertainty rasters.

In targeted checks of the released uncertainty rasters for 1975, 1980, 1985, and 2024, the valid uncertainty values were stored as integers, with the inspected range approximately 0–8, and a large number of pixels had an uncertainty value of exactly 0.

The manuscript states that uncertainty is calculated as the standard deviation among five model predictions. Please clarify whether the standard deviations were rounded or quantized before storage, the units and scale factor of the released uncertainty rasters, and whether a value of zero truly indicates zero inter-model variability. A short numerical description of the uncertainty files in the Data Availability section or repository documentation would be useful.

Relevant manuscript locations: Sect. 2.3.6, Lines ~275–280; Data Availability, Lines ~700 onward.

Response: We thank the reviewer for this careful check and helpful comment. We agree that the numerical representation of the released uncertainty rasters should be clarified. The pixel-level uncertainty was first calculated as the standard deviation among the five fold-specific model predictions using the original floating-point prediction values. The resulting uncertainty values were then rounded only at the final export stage and stored as Int8 GeoTIFF files. Thus, the released uncertainty rasters are expressed in TCC percentage points with a scale factor of 1. A stored value of 0 may indicate that the five models produced identical predictions, which commonly occurs in bare-land areas where all models consistently predict TCC as 0. However, a value of 0 may also result from a small floating-point standard deviation being rounded to 0 during final storage. Therefore, zero values should not always be interpreted as exact zero inter-model variability in the original floating-point calculations. This information has also been added to the description on the corresponding Zenodo repository page.

Specific comment 4

The spatial dimensions of the historical and annual GeoTIFF files are not fully consistent.

In the released data, the inspected 1975/1980 rasters have dimensions of:

237482 × 148426

whereas the inspected 1985 and 2024 rasters have dimensions of:

237483 × 148427.

The files have the same inspected grid origin and pixel size, but the historical files contain one fewer column on the eastern edge and one fewer row on the southern edge.

For a dataset intended for long-term pixel-wise analysis, it would be preferable for all rasters to use exactly the same grid extent and array dimensions. Please check and, if possible, harmonize the historical and annual grids. If the difference is intentional, it should be documented clearly in the repository metadata.

Relevant manuscript location: Data Availability, Lines ~700 onward.

Response: We thank the reviewer for carefully checking the released GeoTIFF files and pointing out this grid inconsistency. We have reprocessed the historical TCC rasters for 1975 and 1980 and regenerated the corresponding GeoTIFF files to ensure that they are fully aligned with the annual 1985–2024 rasters. We also verified the row and column dimensions of the rasters for the corresponding years and confirmed that they are consistent in QGIS (The raster size information displayed by QGIS is uniformly as follows: Width: 237483, Height: 148427).

In the revised data release, all TCC layers now share the same grid origin, pixel size, spatial extent, row and column dimensions, and coordinate reference system. This harmonization ensures that the full time series can be used directly for long-term pixel-wise analysis without additional grid adjustment. The updated files have been uploaded to Zenodo (Part 1: 1975, 1980, and 1985–1988;

DOI: <https://doi.org/10.5281/zenodo.22841763>).

Where changed:

Data published page—ChinaTCC30 dataset tif file. [MODIFIED]

ChinaTCC30_1975.tif, ChinaTCC30_1980.tif, Uncertainty_1975.tif, and Uncertainty_1980.tif have been replaced with the consistent rasters. (Part 1: 1975, 1980, and 1985–1988; DOI: <https://doi.org/10.5281/zenodo.22841763>)

Specific comment 5

Typographical error at Line 340

“K-means-based diversity active learning-Radom Forest” should be corrected to “Random Forest”.

Response: We thank the reviewer for pointing this out. We have corrected the typographical error in Line 340.

Where changed:

Manuscript —3.3 Training and optimization of retrieval models — Page 17, Line 396. [MODIFIED]

Under this setting, K-means-based diversity active learning-Random Forest (KDAL-RF) consistently outperformed random-sampling RF across most sampling stages, with higher R^2 and lower RMSE (Fig. 6a, b).

Specific comment 6

Number of accuracy metrics

The manuscript states that “five standard statistical metrics” were calculated, but only r , R^2 , RMSE, and Bias are subsequently listed. Please revise “five” to “four” or add the missing metric.

Relevant manuscript location: Sect. 2.3.4, Lines ~243–247.

Response: We thank the reviewer for pointing this out. We have revised “five” to “four” in Section 2.3.4 Accuracy assessment.

Where changed:

Manuscript —2.3.4 Accuracy assessment— Page 10, Line 278. [MODIFIED]

To quantitatively assess the accuracy of the TCC estimations, four standard statistical metrics were calculated...

Specific comment 7

Please moderate causal interpretations of the observed TCC trends.

The Results state that the positive TCC trajectory reflects the “notable effectiveness of afforestation and ecological restoration efforts.” However, the present analysis maps TCC changes rather than explicitly attributing them to individual conservation or restoration programs.

I suggest using more cautious wording such as “consistent with” or “may be associated with” ecological restoration unless a dedicated attribution analysis is provided.

Relevant manuscript location: Results, around Lines ~420–435.

Response: We thank the reviewer for this helpful suggestion. In the revised Results section, we have moderated the causal wording. We now describe the observed greening pattern as being “consistent with” or “possibly associated with” multiple contributing factors, including large-scale afforestation, ecological restoration, vegetation recovery, climate variability, and land-management changes. We have avoided statements implying direct attribution to any single program or driver.

Manuscript —3.5 TCC temporal dynamics — Page 21, Lines 483–497. [MODIFIED]

Overall, these results suggest a generally positive trajectory of tree canopy dynamics across China. This greening pattern may be associated with multiple factors, including large-scale afforestation, ecological restoration, vegetation recovery, climate variability, and land-management changes. In contrast, Guangxi and Hainan exhibited relatively severe canopy loss, as indicated by the more negative mean Sen’s slopes among pixels with significant TCC decline (Fig. 10c). In Guangxi, these stronger negative trends may be associated with intensive forest management in landscapes dominated by fast-growing plantations. As the province with the largest area of eucalyptus in China (NFGA, 2019), Guangxi is characterized by widespread short-rotation plantation management and recurrent harvesting. Such practices can cause abrupt canopy removal over short periods and may therefore contribute to more negative trajectories in satellite-derived TCC (Cohen et al., 2010; Hansen et al., 2013). In Hainan, the relatively large negative slopes may be associated with the combined effects of recurrent environmental disturbances and land-use transitions. Frequent typhoons, droughts, and occasional cold or frost events can increase canopy vulnerability to abrupt damage (Ahrends et al., 2015); where recovery is incomplete or disturbances recur, these impacts may translate into sustained negative TCC trends. In addition, the expansion of tropical cash-crop plantations, such as rubber, may have further contributed to canopy decline or structural simplification in some areas by replacing evergreen broadleaf forests and secondary vegetation with more homogeneous plantation systems (Ziegler et al., 2009).

Manuscript —4.1.3 Dynamic and trend comparison— Page 32, Lines 673–674. [MODIFIED]

The increase after 1985 points to a continued recovery of Chinese forests over the past several decades and may be associated with forest conservation and management measures, among other factors.

Specific comment 8

Please justify the use and resampling of the climate data.

The manuscript states that temperature and precipitation from TerraClimate were bilinearly resampled to 30 m and then used for plot-scale pseudo-sample construction and prediction. Because the native spatial resolution of TerraClimate is approximately 4.6 km, bilinear resampling to 30 m does not introduce new fine-scale climatic information.

Please explain why TerraClimate was selected instead of a higher-spatial-resolution climate dataset, if suitable alternatives are available, and clarify the intended role of these resampled climate

variables in the 30 m modeling framework. It would also be useful to discuss whether the coarse native resolution of the climate data may limit the representation of local climatic gradients.

Relevant manuscript location: Sect. 2.1.2, Lines ~130–135.

Response: We thank the reviewer for this important and helpful comment. We agree that bilinearly resampling climate data from its native spatial resolution to 30 m does not introduce new fine-scale climatic information. Therefore, we revised both the climate data source and the preprocessing workflow for the climate variables.

In the revised analysis, the TerraClimate temperature and precipitation variables were replaced by China-specific monthly mean temperature and precipitation datasets with a spatial resolution of approximately 1 km. Compared with TerraClimate, these finer-resolution climate datasets provide more detailed spatial information and are more suitable for representing regional climatic gradients across China.

Following the reviewer’s concern, we removed the previous 30 m bilinear resampling step from the climate-variable preprocessing workflow. Annual temperature and precipitation predictors were first derived from the monthly climate layers at their native 1 km spatial resolution. These annual climatic predictors were then directly used as auxiliary environmental variables in the plot-scale TCC model construction and pseudo-plots prediction.

We have revised Sect. 2.1.2 and Table S1 accordingly. This revision was also consistently propagated through the historical reconstruction workflow, including the regeneration of historical pseudo-samples and the reconstruction of the 1975 and 1980 TCC baselines.

Where changed:

Manuscript —2.1.2 Auxiliary data — Page 5, Lines 143-147. [MODIFIED]

Climatic variables, including temperature and precipitation, were obtained from China-specific monthly mean temperature and precipitation datasets with a spatial resolution of approximately 1 km developed by Peng et al. (2019). Annual temperature and precipitation predictors were derived from the monthly climate layers and used for plot-scale pseudo-sample model construction and prediction (see Table S1 for further details).

Supplement —Table S1 — Page 4. [MODIFIED]

Dataset	Regions	Temporal scope	Spatial resolution	Sources
Landsat	China	1985-2024	30 m	(Cai et al., 2025)
ALCC	China	1985-2022	30 m	(Li et al., 2026)
CLCD	China	1985, 1990-2024	30 m	(Yang and Huang, 2021)
NASA DEM	Global	—	30 m	https://www.earthdata.nasa.gov/data/catalog/lpcloud-nasadem-hgt-001
1901-2025				
Monthly climate dataset for China	China	1901-2025	1000 m	(Peng et al., 2019)