



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

Jinge Yu¹, Zhen Yu^{1, 2, 3, 4 *}, Wangya Han^{1, 4}, Fangmin Zhang^{1, 4}, Shirong Liu^{2, 3 *}

¹State Key Laboratory of Climate System Prediction and Risk Management, Nanjing University of Information Science and Technology, Nanjing, 210044, China

²Institute of Forestry and Grassland Carbon Sequestration, Chinese Academy of Forestry, Beijing, 100091, China

³Key Laboratory of Forest Ecology and Environment, China's National Forestry and Grassland Administration, Ecology and Nature Conservation Institute, Chinese Academy of Forestry, Beijing, 100093, China

⁴Key Laboratory of Ecosystem Carbon Source and Sink, China Meteorological Administration (ECSS-CMA), School of Ecology and Applied Meteorology, Nanjing University of Information Science and Technology, Nanjing, 210044, China

*Correspondence to: Zhen Yu (zyu@nuist.edu.cn) & Shirong Liu (liusr@caf.ac.cn)

Abstract. Tree canopy cover (TCC) is a key biophysical indicator of forest structure and function, and long-term fine-resolution TCC data are essential for quantifying carbon cycling, monitoring forest succession, and supporting forest management under ongoing climate and land-use change. However, accurate long-term national-scale TCC mapping remains challenging because strong spatial heterogeneity requires large and representative reference samples that are difficult to obtain consistently across space and time. This challenge is particularly evident in China, where early historical TCC baselines and long-term annual TCC records remain insufficient. To address this gap, we developed an integrated framework that combines active learning, Random Forest, and stand-growth-based historical backtracking to reconstruct China TCC at 30 m (ChinaTCC30) for 1975, 1980, and annually from 1985 to 2024. Specifically, the historical TCC for 1975 and 1980 was reconstructed from MSS imagery using pseudo-samples generated under limited reference conditions, while annual wall-to-wall TCC maps for 1985–2024 were produced from Landsat imagery using visually interpreted TCC samples. Pixel-level uncertainty was further quantified using multi-model predictions. Validation against multiple independent reference datasets demonstrated the robustness and reliability of the resulting dataset, including visually interpreted samples ($R^2 = 0.83$, RMSE = 15.14%) and multi-period National Forest Inventory data (6th–9th NFI; $R^2 = 0.66$ – 0.82 , RMSE = 14.68%–18.81%). Compared with existing products, the resulting ChinaTCC30 dataset shows good national-scale spatial consistency while better capturing fine-scale spatial heterogeneity and long-term temporal continuity. As a spatially explicit TCC dataset spanning nearly five decades, it provides valuable support for forest monitoring, climate change research, and forest conservation and management, but also offers a cost-effective reference for TCC mapping at national and regional scales. The ChinaTCC30 dataset is available via an open-data repository: Part 1 (1975, 1980, and 1985–1988; DOI: <https://doi.org/10.5281/zenodo.21194015>), Part 2 (1989–1994; DOI: <https://doi.org/10.5281/zenodo.21195911>), Part 3 (1995–2000; DOI: <https://doi.org/10.5281/zenodo.21199497>), Part 4 (2001–2006; DOI: <https://doi.org/10.5281/zenodo.21245669>), Part 5 (2007–2012; DOI: <https://doi.org/10.5281/zenodo.21262725>), Part 6 (2013–2018; DOI: <https://doi.org/10.5281/zenodo.21263677>), and Part 7 (2019–2024; DOI: <https://doi.org/10.5281/zenodo.21265832>) (Yu et al., 2026a, b, c, d, e, f, g).



1 Introduction

35 Forest ecosystems represent the most vital terrestrial carbon sink on Earth, playing an indispensable role in sustaining global biodiversity, modulating hydrological cycles, and mitigating anthropogenic climate change (Pan et al., 2013; Pugh et al., 2019). The efficacy of these critical ecological services is intrinsically governed by forest structural configurations (Ziter et al., 2019; Zhang et al., 2023; Bar-On et al., 2018). Tree Canopy Cover (TCC), an important structure parameter in forest inventories representing the continuous fractional area obscured by the vertical projection of tree canopy, directly reflects forest structural characteristics and growth states (Sexton et al., 2013; Yu et al., 2023). By precisely quantifying the density and horizontal distribution of trees, continuous TCC estimations provide a far more sensitive metric for detecting gradual ecological shifts, such as early-stage forest degradation or progressive densification resulting from ecological restorations (Song et al., 2018; Huang et al., 2012). In the context of escalating global land-use and climate changes, tracking the fine-grained spatiotemporal dynamics of TCC is a fundamental prerequisite for accurately evaluating carbon cycling, assessing long-term forest succession, and formulating targeted environmental policies (Brandt et al., 2020a; Bastin et al., 2019).

Although remote sensing analyses based on greenness indicators such as NDVI and LAI suggest that China has played an important role in global greening over recent decades (Zhu et al., 2016; Chen et al., 2019; Piao et al., 2020), enhanced greenness does not always correspond to increased woody vegetation cover (Piao et al., 2020; Chen et al., 2019; Sexton et al., 2013). Part of this discrepancy arises because greenness trends may also capture the effects of agricultural intensification, fluctuations in grassland productivity, and seasonal vegetation dynamics, rather than exclusively representing tree canopy expansion (Fensholt et al., 2012). Accordingly, continuous structural measures such as TCC provide a more direct basis for assessing long-term woody expansion than greenness-based indicators alone (Sexton et al., 2013; Hansen et al., 2013).

As the country with the world's fifth-largest forest area and the largest expanse of planted forests, China has experienced remarkable changes in tree cover (FAO, 2020; Cai et al., 2025b). Since the late 1970s, the Chinese government has implemented large-scale ecological restoration initiatives, including the Three-North Shelterbelt Program and the Grain for Green Project, to mitigate environmental degradation (Liu et al., 2008; Lu et al., 2018). These massive afforestation campaigns have reshaped the national landscape, expanding formal forest coverage from 12.98% in the 1980s to 24.02% by 2021 (NFGSDC, 2022). Reconstructing reliable historical baseline productions is a critical prerequisite for evaluating the long-term effectiveness of China's initial ecological restoration policies (Higgs et al., 2014; Bryan et al., 2018; Guo et al., 2026). However, the absence of spatially explicit historical ground-truth records for tree canopy cover fundamentally incapacitates traditional data-driven machine learning models for robust retrospective mapping (Maxwell et al., 2018). Previous studies have demonstrated the feasibility of utilizing established empirical relationships (e.g., age-based models for stand density index, biomass, growth rates, and tree height) to predictively simulate missing forest structural parameters over study time (Pretzsch, 2008; Burkhardt and Tomé, 2012). Most notably, within pure and even-aged stands, the non-linear decay trajectories driven by natural self-thinning mechanisms are well-documented and highly predictable (Burkhardt and Tomé, 2012; Tang et al., 1994). Consequently, in the severe absence of early historical field data, relying on established empirical relationships between contemporary stand



attributes and stand age to back-cast historical stand characteristics (e.g., stem density) through mathematical deduction represents a viable pathway for acquiring early-era plot data (Tang et al., 1994). However, there is currently a conspicuous lack of studies that translate such parameter-derived historical estimates into spatially explicit "pseudo-samples" to drive machine learning models for the large-scale spatial modeling of early-stage TCC.

In recent years, the open availability of high-resolution Earth observation data has greatly facilitated large-scale monitoring of vegetation structure at fine spatial resolution (Carreiras et al., 2006; Woodcock et al., 2008; Hansen et al., 2013; Bastin et al., 2019; Kolluru et al., 2026). In practice, however, accurate wall-to-wall mapping across heterogeneous landscapes requires large and representative training datasets. Acquiring such datasets through extensive field surveys or nationwide manual interpretation is often prohibitively expensive, time-consuming, and logistically demanding. This constrains the ability of some countries and regions to maintain consistently high mapping accuracy, thereby limiting long-term monitoring of environmental dynamics (Qin et al., 2025). To break through the sharp paradox between the scarcity of high-quality samples and exorbitant manual labeling costs, Active Learning (AL) strategies have demonstrated transformative application potential (Qin et al., 2025). AL empowers machine learning models to adaptively evaluate and query the most informative samples for manual annotation during the training process (Settles, 2009). Driven by the dual necessity for model parsimony and high prediction accuracy in large-scale mapping, AL algorithms are increasingly being integrated into quantitative remote sensing frameworks (Tuia et al., 2009; Tuia et al., 2011). By strategically optimizing both the feature and sample spaces, AL not only renders machine learning models significantly more lightweight but also substantially enhances their overall accuracy. Recent studies have comprehensively corroborated the tremendous efficacy of AL in optimizing hybrid models for the retrieval of complex vegetation traits (Verrelst et al., 2016; Verrelst et al., 2020; Berger et al., 2021; Estévez et al., 2022; Qin et al., 2025). However, the effectiveness of AL for long-term TCC mapping remains insufficiently explored, particularly regarding its ability to reduce annotation effort while maintaining high mapping accuracy across large and heterogeneous regions.

Substantial efforts have been made to characterize land-surface dynamics, resulting in an increasing availability of TCC-related products. For instance, the Hansen Global Forest Change dataset provides a 30 m baseline tree-cover map for 2000 and monitors subsequent forest change using annual loss and gain information (Hansen et al., 2013). However, it lacks continuous TCC information before 2000 and cannot directly represent interannual canopy-cover dynamics. The continuous canopy cover product developed by Sexton et al. (2013) is better suited than binary forest maps for representing canopy-cover gradients, but it lacks the annual continuity needed for nationwide, multi-decadal analyses. More recently, the China annual tree cover dataset (CATCD) has greatly improved understanding of long-term TCC changes since 1985. The forest-area changes estimated from this dataset are generally more consistent with those reported by the National Forest Inventory (NFI) than those derived from other datasets, although uncertainties may increase in high-latitude regions due to terrain shadow effects and changes in the local solar zenith angle (Cai et al., 2024). Overall, existing products remain limited in terms of reference-sample construction, temporal extent, annual continuity, regional applicability, and cross-dataset consistency. These limitations highlight the need to further extend the temporal coverage of annual TCC datasets and to develop a mapping framework that simultaneously achieves low cost, long-term consistency, and high accuracy.



In summary, early historical TCC baselines 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 modeling 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 modeling 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.

2 Materials and methods

2.1 Remote sensing data

2.1.1 Landsat data

This study utilized both the 30 m and 60 m Landsat images for tree cover dataset reconstruction. The 30 m images were a spatiotemporally consistent Landsat composite dataset generated by Cai et al. (2025a) covering the period of 1985-2024. The dataset was pre-processed in Google Earth Engine using growing-season phenological filtering, QA-based cloud masking (Zhu et al., 2015), cross-sensor harmonization (Roy et al., 2016b), and medoid compositing with temporal gap-filling. The 60 m Landsat MSS imagery was additionally acquired from Google Earth Engine to support reconstruction of the early TCC baseline. Because valid MSS observations were sparse during the early period, multi-year MSS imagery was composited into single target-year mosaics. Using this approach, we generated two MSS-based composite images representing approximately 1975 and 1980 for China. Based on these remote sensing inputs, we constructed a multidimensional feature space to distinguish tree cover from spectrally complex backgrounds, incorporating spectral bands and derived indices where sensor availability permitted. The candidate features included the visible, near-infrared, and shortwave-infrared bands, together with NDVI, EVI, DVI, RVI, ARVI, MSAVI, NBR, NDBI, MNDWI, and NDSI. Because the MSS sensor differs in spectral band configuration from later Landsat sensors, the set of derivable features was correspondingly adjusted, including Green, Red, NIR, RVI, DVI, MSAVI, and NDVI.

2.1.2 Auxiliary data

To reduce misclassification in complex ecotones, such as water-land interfaces and agroforestry landscapes, we incorporated the Annual Land Cover of China (ALCC) dataset (Li et al., 2026) into the modeling framework. Instead of using selected ALCC 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. This strategy allows the model to use categorical contextual information, thereby improving tree cover estimation in heterogeneous landscapes while preserving sensitivity to sub-pixel tree features. In addition, topographic variables, including elevation, slope, and aspect,



derived from the 30 m NASA DEM digital elevation model and accessed via Google Earth Engine, were included as environmental predictors to help characterize the spatial distribution of tree growth across topographic gradients. Climatic variables, including temperature and precipitation, were obtained from the TerraClimate dataset, also accessed via Google Earth Engine. The climatic data were bilinearly resampled to a spatial resolution of 30 m and used for plot-scale pseudo-sample model construction and prediction (see Table S1 for further details on the auxiliary data).

2.2 Multi-source reference data

A multi-tiered validation strategy was employed to rigorously assess the dataset's performance at both regional and national levels. Specifically, this multi-source validation suite incorporates: (1) multi-temporal validation samples derived from Google Earth Very-High-Resolution (VHR) imagery; (2) TCC validation samples spanning four phases National Forest Inventory (1999-2003, 2004-2008, 2009-2013, 2014-2018); and (3) spaceborne canopy cover estimates from the Global Ecosystem Dynamics Investigation (GEDI). The integration of these diverse data sources ensures a rigorous and multi-dimensional accuracy assessment.

2.2.1 Photo-interpreted tree cover data

The acquisition of reliable reference data involved a rigorous visual interpretation protocol based on a systematic sampling design. 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. To precisely quantify the TCC, a 10×10 point grids was overlaid onto each plot, yielding 100 evenly spaced sampling locations per pixel. Each point was manually evaluated against VHR satellite imagery (predominantly from 2015 to 2024) via Google Earth Pro, receiving a binary indicator (1 for canopy, 0 for non-canopy). We extracted the Landsat features using the exact acquisition year of the corresponding VHR imagery for temporal consistency. Next, the TCC for each plot was calculated by summing the 100 point-level observations (Fig. 1b). This 30×30 grid design simplifies the estimation process. The total count of canopy numbers directly equals the TCC percentage (e.g., 60 points represent 60% TCC) (Fig. 1c).

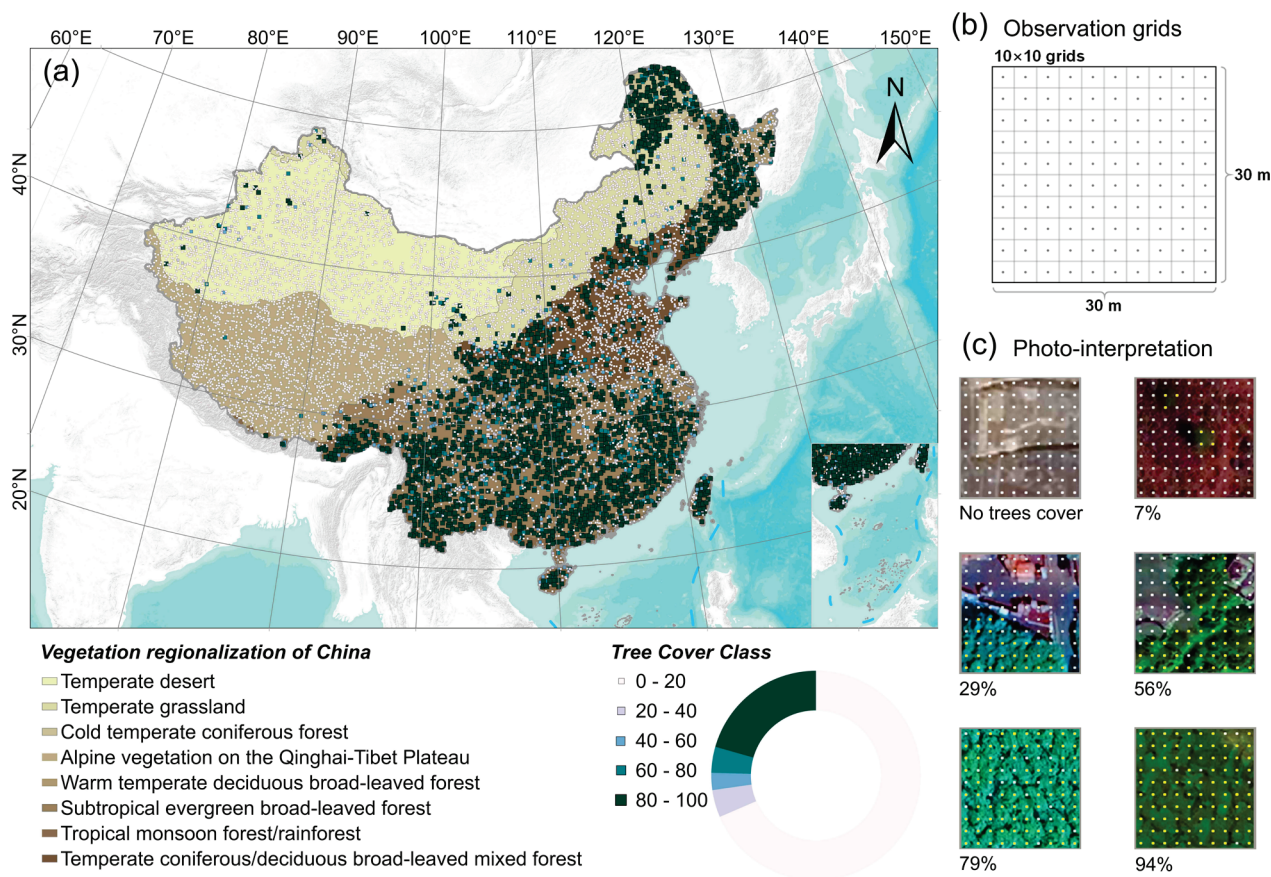


Figure 1. Overview of the reference data. (a) Spatial distribution of tree canopy cover data interpreted from Very-High-Resolution (VHR) images across China's vegetation regionalization (Hou, 2019). (b) Tree canopy cover observation grid layout matching remote sensing image pixels. (c) Principle and calculation method of tree cover area based on photo-interpretation. The yellow dots covered by the canopy, and the white dots not covered by the canopy. The basemap in (a) was designed and developed by Esri | Powered by Esri. For more information, visit https://server.arcgisonline.com/ArcGIS/rest/services/World_Terrain_Base/MapServer (last access: 16 July 2026). Images: © 2025 CNES / Airbus, © 2025 Maxar Technologies.

2.2.2 National forest inventory data

To ensure the reliability of tree canopy cover mapping across China, National Forest Inventory (NFI) data were used as the core reference dataset. These data span four consecutive inventory cycles: the 6th (1999–2003), 7th (2004–2008), 8th (2009–2013), and 9th (2014–2018) NFIs. This standardized dataset was used to systematically assess the performance of the estimation method in quantifying TCC across multiple historical periods.



2.2.3 Spaceborne Lidar derived canopy cover data

To establish a robust and highly reliable reference dataset for the TCC estimation, we utilized the Global Ecosystem Dynamics Investigation (GEDI) Level 2B Canopy Cover product (GEDI02_B) spanning from 2019 to 2024. Specifically, we extracted the directional gap probability-derived canopy cover metric at a 25 m footprint resolution. Given the inherent noise in space-
170 borne 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`). Through these rigorous screening procedures, we obtained a nationwide distribution of GEDI
175 sampling points to support subsequent model training and validation.

2.3 Tree canopy cover mapping based on active learning coupled with machine learning

2.3.1 K-means-based Diversity Active Learning

We employed a K-means-based Diversity Active Learning (KDAL) strategy to optimize the training dataset for the Random Forest (RF) model. The workflow of this strategy is shown in Fig. 2. First, uncertainty-based sampling was used to identify
180 informative samples from the unlabeled pool, where prediction uncertainty was quantified by the variance of the RF ensemble (Beluch et al., 2018). The samples with the highest uncertainty were then retained as a candidate subset. However, uncertainty-driven querying alone may introduce substantial redundancy, since many highly uncertain samples are often concentrated in similar regions of the feature space. To improve sample diversity, we incorporated a K-means-based filtering step after uncertainty screening. The candidate subset was clustered into multiple groups using the K-means algorithm, and the sample closest
185 to each cluster centroid was selected as the representative sample for annotation and subsequent model updating (Demir et al., 2011). This cluster-guided distillation step enabled the selection of samples that were not only highly informative but also structurally distinct, thereby reducing redundancy among queried samples and improving the efficiency of training set expansion.

The proposed KDAL strategy was designed to maintain high information diversity while minimizing the number of ad-
190 ditional training samples required. Since K-means clustering relies on Euclidean distance and is sensitive to variable magnitude, all input features were normalized to the range of 0–1 using min–max scaling prior to clustering. The representative target sample x_{target} within each cluster S_k was determined according to Equation (1).

$$x_{\text{target}} = \underset{x \in S_k}{\operatorname{argmin}} \|x - C_k\|_2^2 \quad (1)$$

where x indicates the feature vector of a candidate sample, and C_k represents the feature vector of the k -th cluster centroid.

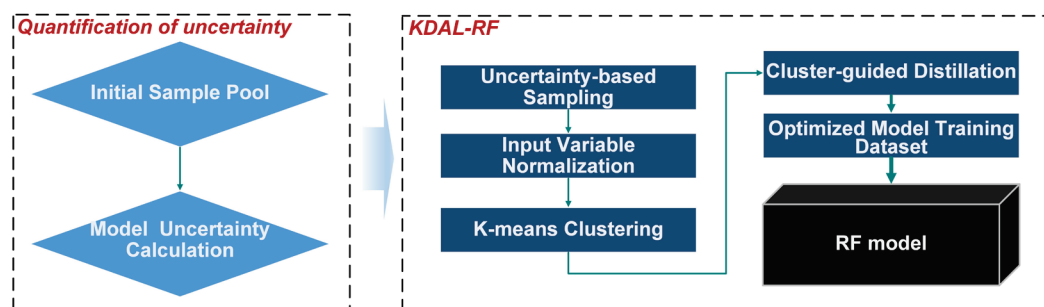


Figure 2. Flowchart of the K-means-based Diversity Active Learning (KDAL) strategy for Random Forest model optimization.

To ensure an unbiased and rigorous evaluation of the AL strategy, we established a statistically balanced independent benchmarking framework, coupled with multiple random restarts to mitigate initialization bias. The evaluation framework was executed in two distinct phases prior to any AL iterations. First, to prevent evaluation metrics from being heavily skewed by the overwhelmingly dominant zero-coverage (TCC=0) background, an independent test set was pre-extracted from the total sample pool. Using an inverse-density stratified sampling technique, this test set was enforced to maintain an absolute uniform distribution across all TCC gradients (0–100%). Subsequently, to evaluate the statistical robustness of the proposed AL algorithm against initial sampling randomness, the entire active learning workflow was repeated across ($N = 10$) independent parallel runs. In each run, a different initial training subset was randomly selected to trigger the iterative querying process, generating N independent learning trajectories and estimation models at each sampling step. The generalization capacity of the models at each AL iteration was evaluated on the frozen balanced benchmark set using two metrics: the coefficient of determination (R^2) and the root mean square error (RMSE).

2.3.2 Random forest

In this study, the Random Forest (RF) regression algorithm (Breiman, 2001) was employed to model the non-linear relationship between the photo-interpreted TCC samples and the satellite-derived predictors. Prior to model training, a rigorous feature selection process was conducted to optimize computational efficiency and mitigate the Hughes phenomenon. Specifically, the relative importance of the initial predictor variables was evaluated using the RF-based feature importance metric (Guyon and Elisseeff, 2003; Hughes, 1968). These variables included Landsat multispectral bands, various vegetation indices, and terrain-related factors. Only the highest-ranked features were retained to construct the annual feature space.

To achieve optimal predictive performance while accommodating the computational constraints of the Google Earth Engine (GEE) platform (e.g., maximum model size limits), the hyperparameters of the RF model were carefully tuned using a grid search approach. Based on the tuning results, the ensemble was configured with 500 decision trees ($n_{tree} = 500$). Other key hyperparameters, such as the number of variables randomly sampled as candidates at each split (m_{try}) and the maximum depth of the trees (max_depth), were optimized to ensure a balance between model accuracy and generalization capability.



220 2.3.3 Construction of Pseudo-plot Samples for Historical Baseline Reconstruction

Because reference samples of TCC for 1975 and 1980 were unavailable for direct model training, this study constructed historical pseudo-plot samples within stable natural even-aged pure stands to support TCC reconstruction for historical baseline periods. The general strategy was to first establish a quantitative relationship between TCC and forest structural and environmental variables using contemporary observations, and then reconstruct the model input variables for the target historical years to predict TCC values, thereby generating pseudo-plot samples for historical baseline mapping. Specifically, a TCC reconstruction model was developed based on contemporary TCC observations in combination with stand, topographic, and climatic variables. The stand variables included stand age and tree density; the topographic variables included elevation, slope, and aspect; and the climatic variables included temperature and precipitation. This model was used to characterize the relationships between TCC and forest structural attributes and environmental conditions, and to provide the basis for historical TCC reconstruction. The design of this modeling scheme is consistent with previous studies, which have shown that canopy cover and related forest structural attributes are jointly constrained by stand development, site conditions, and climatic variability (Lefsky et al., 2002; Hansen et al., 2013; Pretzsch, 2009).

To derive the model inputs for the historical periods, the variables were treated according to their temporal characteristics. Topographic variables were assumed to remain constant over the study period and were therefore used as static predictors. Climatic variables were obtained from historical records for 1975 and 1980. In contrast, stand variables exhibited clear temporal dynamics and thus required retrospective reconstruction. Specifically, stand age for each plot was back-calculated to 1975 and 1980 based on the stand-age time series. Tree density for the corresponding years was then estimated using the established empirical relationship between stand age and stem density, consistent with classical stand development theory and empirical stand-density dynamics in even-aged forests (Pretzsch, 2009; Weiskittel et al., 2011; Tang et al., 1994). Finally, the reconstructed historical stand variables, together with the static topographic variables and contemporaneous climatic variables, were input into the trained TCC reconstruction model to predict pseudo-plot TCC values for 1975 and 1980, which were then used as response variable in the training samples for historical baseline TCC mapping (Fig. 3a).

2.3.4 Accuracy assessment

To quantitatively assess the accuracy of the TCC estimations, five standard statistical metrics were calculated: the Pearson correlation coefficient (r), R^2 , RMSE, and Bias. Specifically, r and R^2 were utilized to evaluate the linear agreement and the proportion of variance explained by the model, respectively. RMSE was employed to quantify the magnitude of estimation errors, while Bias was used to indicate any systematic overestimation or underestimation.

2.3.5 Defining forest extent and tree spatial distribution patterns

To ensure consistency and enable direct comparison between the ChinaTCC30-derived estimates and the NFI statistics, forest extent was delineated using a two-step procedure. In the first step, we adopted the forest definition proposed by Cai et al.



(2024), which closely aligns with the NFI standard. Based on this definition, the initial forest extent was derived from the ChinaTCC30 dataset using an object-based approach. Specifically, an individual 30 m × 30 m Landsat pixel was classified as forest when its TCC exceeded 74.5%, corresponding to 0.067 ha of canopy-covered area within a single pixel. For contiguous groups of neighboring pixels, classification was performed at the object level: a connected patch was identified as forest when its total area exceeded 0.067 ha and its mean TCC was greater than 20%. Accordingly, pixels with TCC below 74.5% could still be included in the forest extent if they belonged to a forest patch satisfying these object-level criteria. In the second step, forest-area estimation was further refined by accounting for mixed pixels along forest boundaries. Specifically, for pixels neighboring the forest extent identified in Step 1, those with canopy cover greater than 10% were treated as forest edges. Forest area in the edge is the product of pixel area and canopy-cover fraction. This refinement reduced the underestimation of forest area caused by boundary-related mixed pixels in 30 m resolution imagery.

2.3.6 TCC mapping approach

This study employed two complementary mapping strategies to generate annual and historical estimates of TCC across China at 30 m and 60 m spatial resolutions, 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 wall-to-wall 30 m TCC maps for 1985–2024 were produced from Landsat imagery and visually interpreted TCC samples.

For the 1975 and 1980 reconstructions, a total of 1,700 MSS scenes were used. Images from 1972–1977 were composited to produce the 1975 baseline, and images from 1978–1984 were composited to produce the 1980 baseline. Remaining gaps were filled using the Global Land Survey (GLS) image collection, a globally composited Landsat 1–4 image set. Pseudo-samples were then generated using a predictive model with observed TCC as the response variable and stand factors, meteorological variables, and topography-related variables as predictors. Historical meteorological variables were retrieved for the target years, while topographic variables were incorporated as static predictors. Together with spectral features extracted from MSS imagery, these pseudo-samples were used for active learning and TCC reconstruction modeling, and the optimized model was applied to generate wall-to-wall TCC maps for 1975 and 1980 (Fig. 3a).

For the annual mapping during 1985–2024, the observed image-interpreted TCC samples and Landsat-derived features were input into the random forest ensemble model within the active learning collaborative random forest framework. (Fig. 3b). To improve model robustness, a five-fold cross-validation strategy was adopted for both the historical reconstruction and the continuous annual mapping. Specifically, the sample dataset was randomly divided into five subsets; in each round, four subsets were used for model training and the remaining subset was used for validation. This procedure produced five independent models, each of which generated a spatial TCC prediction layer for a specific year. The final TCC map was obtained by averaging the five prediction layers on a pixel-wise basis, while the corresponding uncertainty map was derived from the standard deviation among the five predictions. Accordingly, the mean prediction represented the final TCC estimate, whereas the standard deviation provided a spatially explicit measure of model uncertainty (Fig. 3).

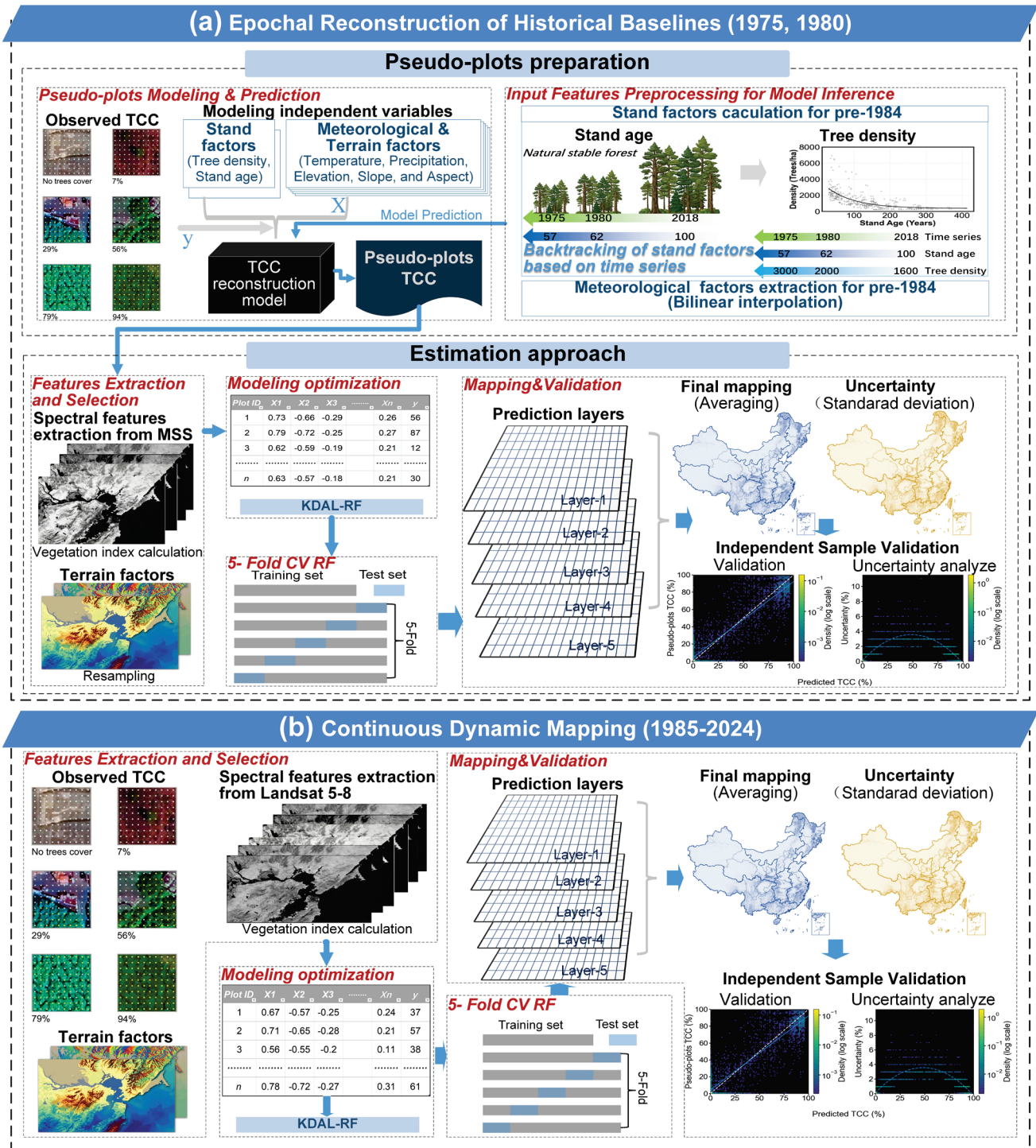


Figure 3. Methodological flowchart of the TCC estimation framework for (a) historical baseline reconstruction in 1975 and 1980 and (b) continuous dynamic mapping from 1985 to 2024. The spectral band layers (Sources: USGS | Powered by USGS). The terrain layers (Sources: NASA | Powered by NASA). Images: © 2025 CNES / Airbus, © 2025 Maxar Technologies.



3 Result

3.1 Construction of pseudo-sample data based on stand factors for the years 1975 and 1980

To construct TCC pseudo-samples for the 1975 and 1980 baseline periods, historical stand variables were first back-casted and then used as inputs to a validated TCC prediction model. To this end, plot-level regression models were developed from pure, even-aged forest inventory plots under comparable site conditions to reconstruct historical stem density. Following comparative evaluation, the best-performing model was retained for each of the nine tree species (Fig. 4a). Consistent with self-thinning theory, stem density decreased nonlinearly with increasing stand age across all tree species. *Larix* spp. was best fitted by a power function, whereas exponential decay functions provided the best fit for the remaining tree species. All retained models were highly significant ($p < 0.001$), with calibration R^2 values of 0.42–0.60 (Fig. 4a; Table S4). Model robustness was further assessed using an independent 20% hold-out test set, yielding R^2 values of 0.57–0.69 and r of 0.77–0.84 (Fig. S1; Table S5).

To enable historical reconstruction, a predictive model for plot-scale TCC was established using visually interpreted TCC as the response variable, with contemporary stand factors, meteorological elements, and topographic variables serving as explanatory variables. A 5-fold cross-validation confirmed the model's accuracy, yielding an R^2 of 0.93, an RMSE of 9.71%, a MAE of 6.02%, and a bias of -0.37% (Fig. 4b). These metrics indicate no systematic estimation errors across the canopy spectrum. Subsequently, the back-casted historical stand factors were input into this validated model to generate historical TCC pseudo-samples. Based on this methodology, a comprehensive pseudo-plot dataset was generated for the 1975 and 1980 baselines (Fig. 4c). Furthermore, to capture the full landscape continuum, zero-value pseudo-plots (0% TCC) were incorporated by randomly generating samples within persistent bare-land and water pixels extracted from the ALCC land cover dataset (Li et al., 2026). Ultimately, this explicitly constructed pseudo-sample dataset serves as the essential response variable for the subsequent remote sensing-based reconstruction of early historical TCC baselines.

3.2 Evaluating variable importance and selection

Feature importance and Pearson correlation analyses were conducted to evaluate the predictive capacity of various variables and to select optimal feature subsets for both the epochal reconstruction models (1975, 1980) (Fig. 5a1, b1) and the continuous mapping models (1985–2024) (Fig. 5a2, b2). Across both periods, variables exhibiting statistically non-significant correlations combined with low relative importance were systematically excluded to minimize data redundancy. Overall, the feature selection process highlighted that visible spectral bands and specific vegetation indices were the primary drivers for tree canopy cover estimation across both modeling frameworks, whereas topographic variables generally provided limited predictive power. Notably, the consistent exclusion of the NIR band across both independent models aligns with findings from previous studies (Carreiras et al., 2006). For the epochal baseline modeling (Fig. 5a1, b1), visible bands (Green, Red) and vegetation indices (MSAVI, RVI, NDVI) exhibited strong correlations and high feature importance. Conversely, the NIR band, DEM, and Aspect were deemed uninformative and subsequently excluded. Consequently, seven features were retained for the historical baseline



reconstruction, comprising two spectral bands (Green and Red), four spectral indices (MSAVI, RVI, NDVI, and DVI), and one topographic factor (Slope). For the continuous mapping modeling (1985–2024) (Fig. 5 a2, b2), 15 features were retained, comprising five spectral bands (B1, B2, B3, B5, B7), nine spectral indices (EVI, NDBI, NDVI, MSAVI, ARVI, RVI, NDSI, MNDWI, NBR), and one terrain factor (Slope).

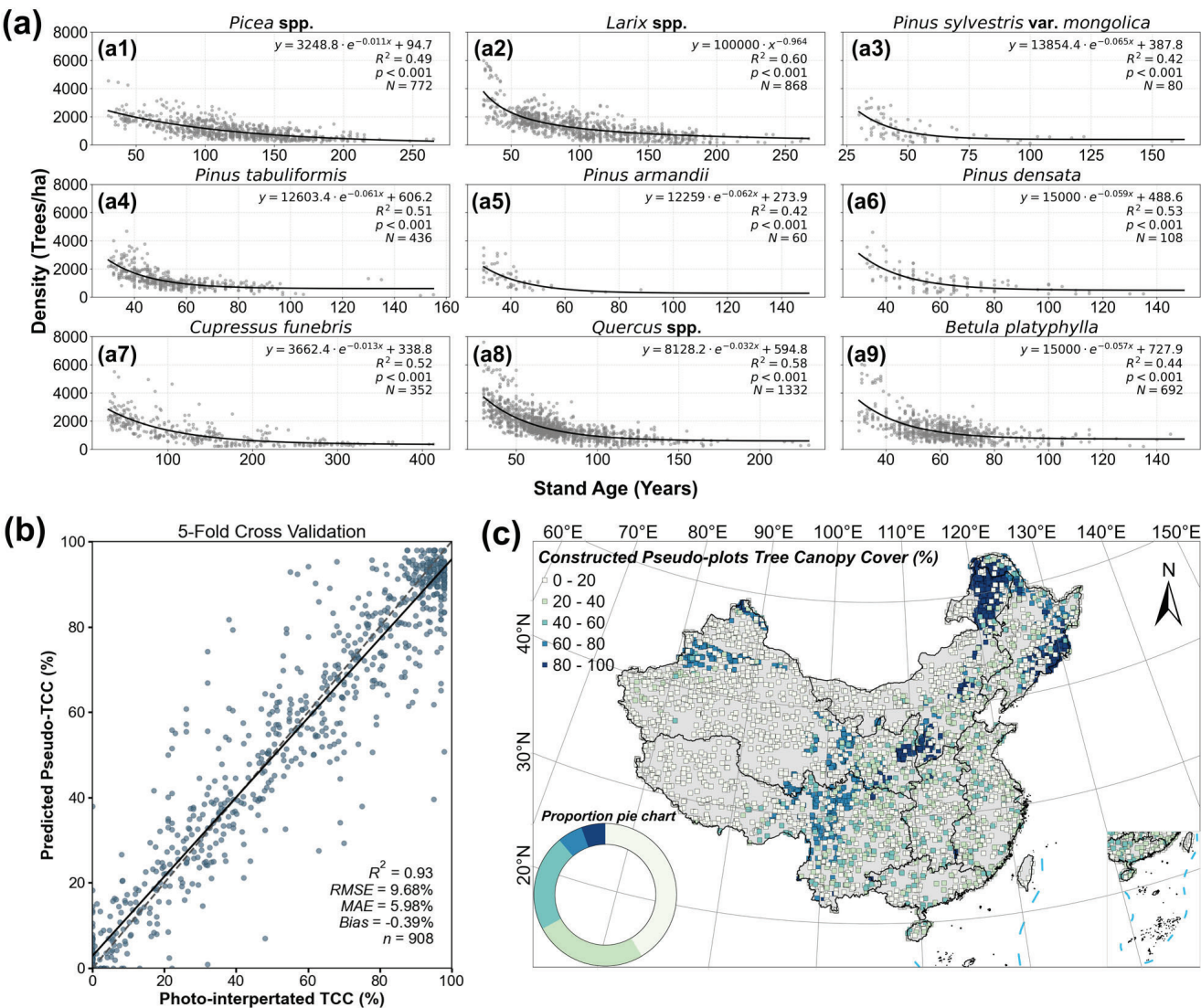


Figure 4. Generation and deduction model validation for historical TCC pseudo-plots. (a) Stand age-density self-thinning curves for nine major tree species/genera (a1-a9), serving as the parameter basis for the deduction algorithm. (b) Performance of the pseudo-plot reconstruction model based on 5-fold cross-validation using photo-interpreted TCC samples. (c) Spatial distribution of the ultimately generated historical pseudo-plots.

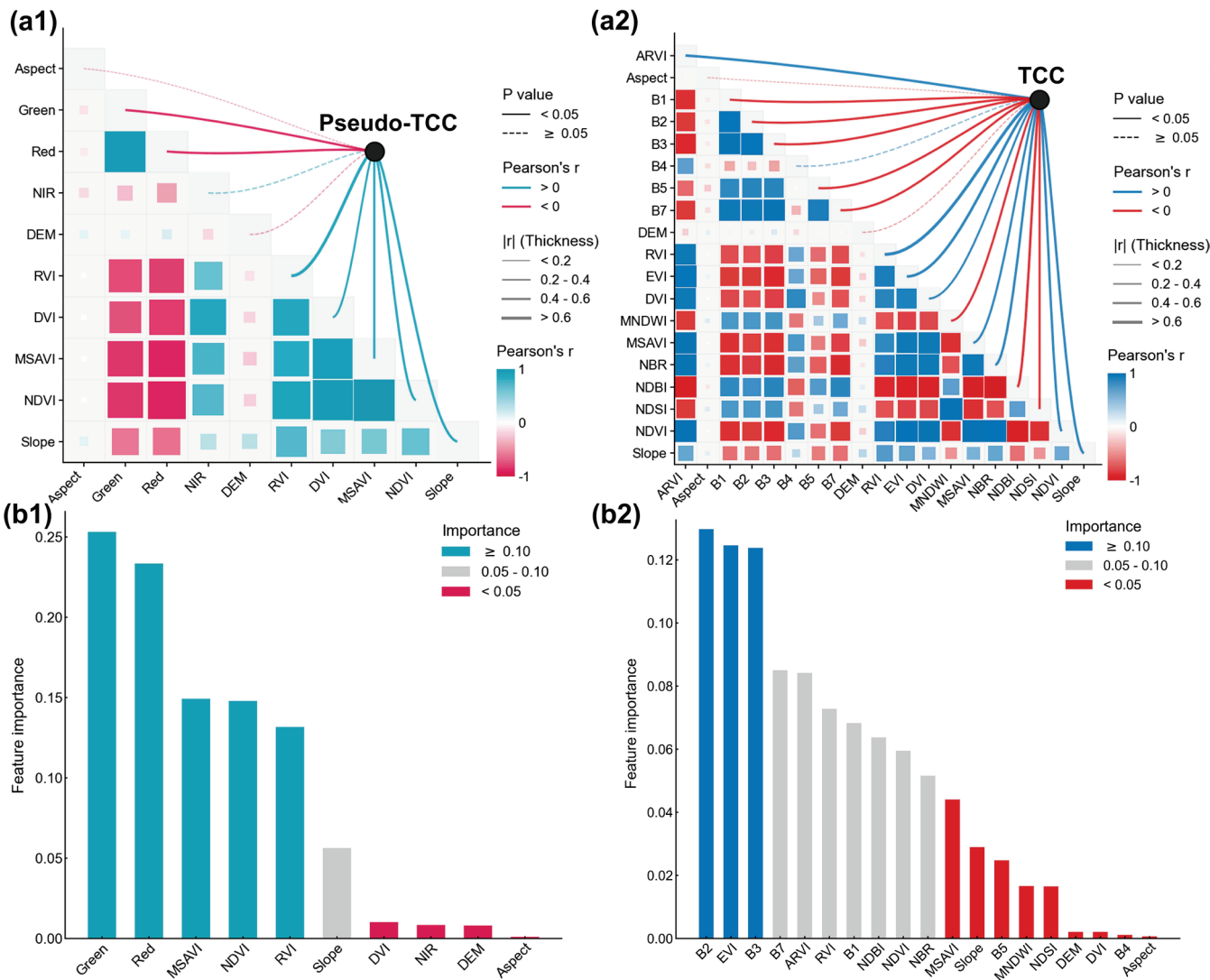


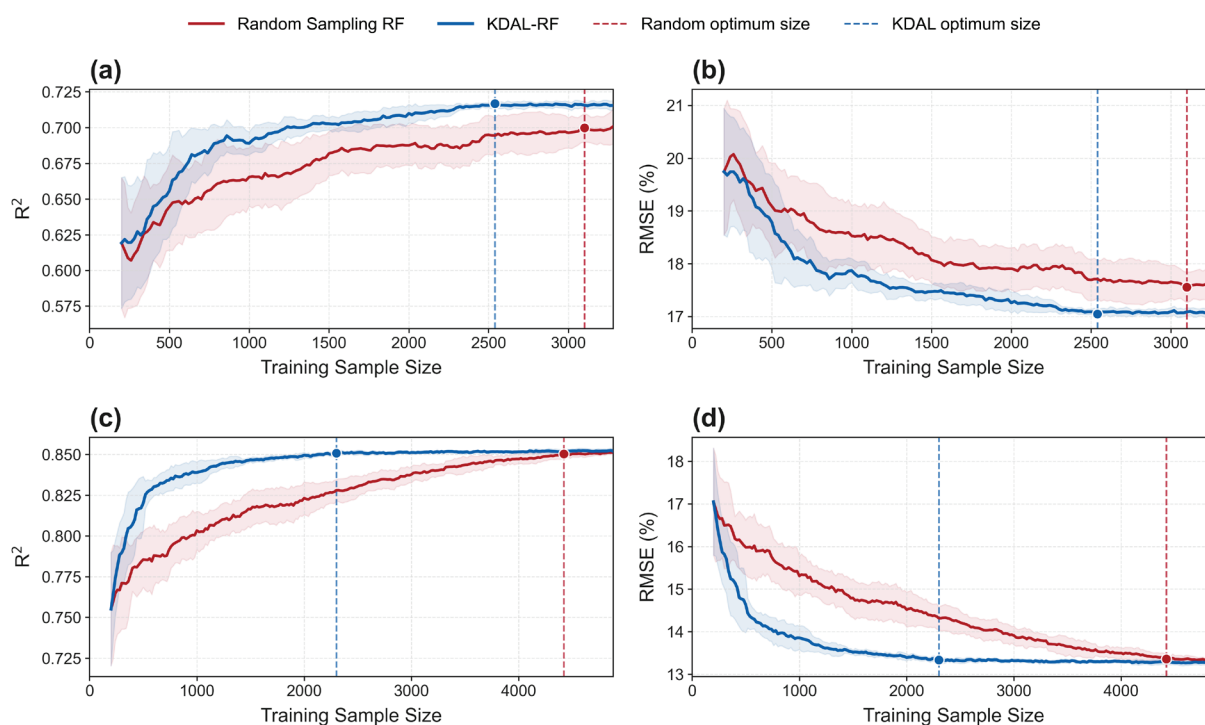
Figure 5. Feature correlation and importance analysis for tree canopy cover inversion models across two periods. The left column (a1, b1) represents the early modeling period based on Landsat MSS features, while the right column (a2, b2) represents the post-1985 modeling period based on Landsat TM/ETM+/OLI features. Panels (a1) and (a2) display the Pearson correlation matrices among predictor variables, along with their correlation strength and significance with the target variable (Pseudo-TCC/TCC). Panels (b1) and (b2) show the feature importance scores derived from the respective models, color-coded by importance thresholds.

3.3 Training and optimization of retrieval models

The proposed active learning strategy (KDAL) substantially improved sample efficiency in both sample-optimization experiments. As the training sample size increased, both methods showed gradual performance improvement, with increasing R^2 and decreasing RMSE (Fig. 6). In the first experiment, each training sample consisted of a pseudo-sample generated through historical retrospective modeling, in which TCC was used as the response variable and Landsat spectral features were used as



340 predictors. 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). KDAL-RF reached a near-stable performance level at about 2,560 samples ($R^2 = 0.72$, RMSE = 17.05%), whereas random-sampling RF required about 3,100 samples to achieve a comparable level, corresponding to a reduction of approximately 17% in training sample demand (Fig. 6a, b). The narrower variability band of KDAL-RF also indicated greater stability across repeated sampling iterations. In
 345 the second experiment, each training sample consisted of a visually interpreted TCC sample paired with Landsat spectral features, with visually interpreted TCC as the response variable and Landsat spectral features as predictors (Fig. 6c, d). KDAL-RF maintained consistently higher R^2 and lower RMSE throughout the training process and reached near-asymptotic performance with only about 2,300 samples, whereas random-sampling RF required about 4,420 samples to achieve a comparable accuracy level ($R^2 = 0.85$, RMSE = 13.33%) (Fig. 6c, d). This corresponds to a reduction of approximately 48% in training
 350 sample demand. Overall, the active learning strategy achieved higher sample efficiency in both experiments, reducing the required training sample size by about 17% in the pseudo-sample-based experiment and by about 48% in the visually interpreted sample experiment (Fig. 6). These results indicate that the main advantage of KDAL lies in accelerating model convergence and reducing training sample requirements through the preferential selection of informative samples.



355 **Figure 6.** Learning curves showing that KDAL-RF improves sample efficiency and accelerates model convergence compared with random-sampling RF across increasing training sample sizes. Panels (a) and (b) show the R^2 and RMSE trends for the pseudo-sample-based experiment, respectively, and panels (c) and (d) show the corresponding results for the visually interpreted sample experiment.



3.4 Mapping accuracy assessment

3.4.1 Independent accuracy evaluation

To assess the reliability and robustness of the KDAL-RF model, two independent accuracy evaluations were performed using 1,549 pseudo samples and 2,692 photo-interpreted validation samples, respectively. Overall, the ChinaTCC30 estimates were in good agreement with the reference data. Against the pseudo-TCC samples (Fig. 7a), the model achieved an R^2 of 0.70, $r = 0.85$, a bias of 1.44%, and an RMSE of 16.95%. When evaluated using the photo-interpreted validation samples (Fig. 7b), model performance improved further, with an R^2 of 0.83, $r = 0.92$, a bias of 0.39%, and an RMSE of 15.14%. In both cases, the fitted regression lines were close to the 1:1 line, indicating a high level of consistency between the predicted and reference TCC values.

Model performance varied across biomes. Lower errors were generally observed in sparsely vegetated regions, where canopy cover was relatively low. Alpine vegetation on the Qinghai–Tibet Plateau and temperate desert exhibited the best performance, with RMSE values of 8.15% and 9.53%, respectively. In contrast, larger errors were found in forest biomes characterized by denser vegetation and more complex canopy structure. The largest RMSE was observed in temperate coniferous and deciduous broad-leaved mixed forest (18.18%), followed by tropical monsoon forest and rainforest (17.72%). Although RMSE differed substantially among biome types, bias remained low overall. These results indicate that the KDAL-RF model maintains stable overall performance across a wide range of vegetation conditions, while exhibiting biome-specific differences in predictive accuracy.

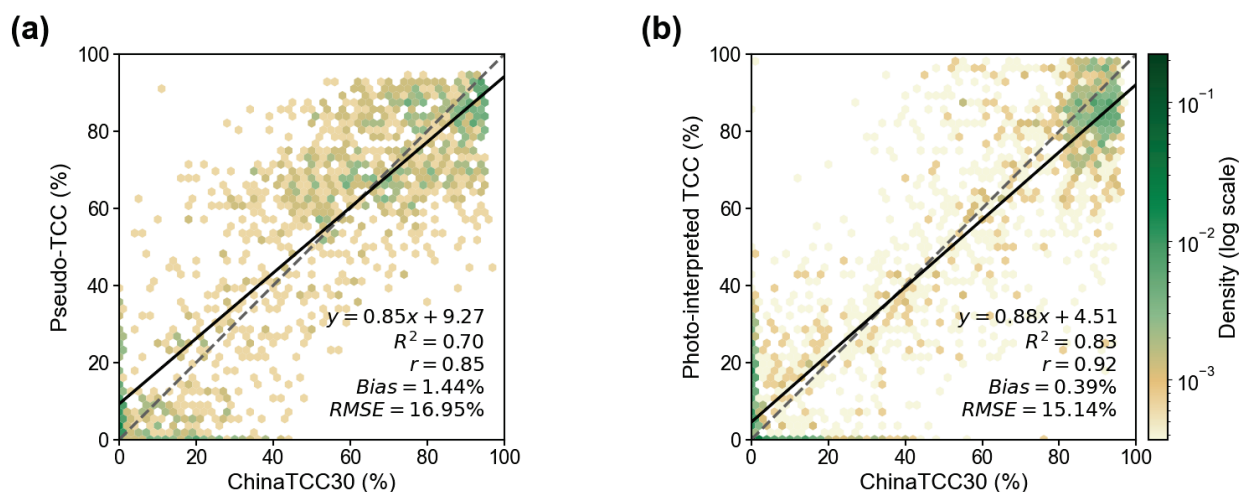


Figure 7. Evaluation of the ChinaTCC30 product developed using K-means-based diversity active learning-Random Forest (KDAL-RF) against independent samples. (a) Scatter plot comparing ChinaTCC30 with the independent pseudo-TCC. (b) Scatter plot comparing ChinaTCC30 with photo-interpreted TCC for the independent test set.



Table 1. Accuracy assessment across ecoregions based on photo-interpreted tree canopy cover

Ecoregions	n	RMSE (%)	Bias (%)
Alpine vegetation on the Qinghai-Tibet Plateau	114	8.15	0.26
Cold temperate coniferous forest	87	15.83	7.28
Subtropical evergreen broad-leaved forest	1272	15.85	2.29
Temperate coniferous and deciduous broad-leaved mixed forest	375	18.18	-2.65
Temperate desert	318	9.53	-2.75
Temperate grassland	256	12.52	-3.88
Tropical monsoon forest and rainforest	106	17.72	0.14
Warm temperate deciduous broad-leaved forest	164	16.05	1.88

3.4.2 External validation using NFI data

To further evaluate the reliability of the ChinaTCC30 product against independent, ground-based measurements, we compared the predicted TCC with observations from the 6th to the 9th NFI (Fig. 8). Overall, the ChinaTCC30 product demonstrates a strong positive correlation with the NFI data across all four inventory periods, with Pearson's r ranging from 0.86 to 0.94. Notably, compared to previous inventory periods, the mapping results exhibit closer agreement with the most recent inventory data. Specifically, the comparison with the 9th NFI yields the highest R^2 (0.82) and the lowest RMSE (14.68%). Furthermore, the linear regression line for this most recent period aligns most closely with the 1:1 line. In contrast, earlier periods, such as the 6th NFI, show relatively lower statistical agreement ($R^2 = 0.66$, RMSE = 18.81%). Additionally, a moderate positive bias (ranging from 3.50% to 7.00%) is observed across all periods. The regression lines and density distributions indicate that the ChinaTCC30 product tends to yield slightly higher TCC estimates compared to the NFI records, particularly in the mid-to-high TCC ranges. Despite this slight systematic deviation, the overall metrics confirm that our product effectively captures the spatial distribution of TCC, showing high consistency with national ground surveys.

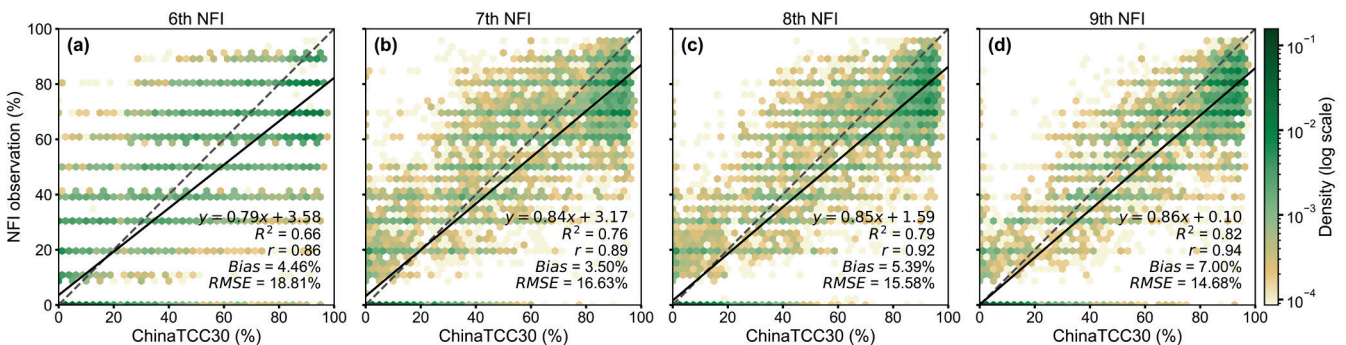


Figure 8. Validation of the ChinaTCC30 product against NFI data across nation four inventory periods (6th–9th).



3.5 TCC temporal dynamics

The long-term TCC maps show clear spatiotemporal dynamics of TCC across China from 1975 to 2024 (Fig. 9). Overall, TCC increased continuously over the study period, revealing a widespread greening trend at the national scale. According to ChinaTCC30, China's forest area has expanded from 1.08 million km² in 1975 to 2.32 million km² in 2024. Spatially, high-TCC areas were mainly distributed in the Northeast, South, and Southwest, while the Northwest remained dominated by low canopy cover (Fig. 9; see Fig. S2-S8 for annual TCC maps). The Spatial-temporal pattern mapping further indicates a gradual increase in both the density and continuity of TCC. Compared with the earlier maps, TCC in 2024 became more extensive in central, southwestern, and eastern China, whereas the major spatial pattern of forest distribution remained broadly stable. This suggests that recent changes were primarily characterized by the enhancement and expansion of existing tree-covered areas rather than by a complete redistribution of forested regions. The center-of-gravity trajectory also reveals a distinct westward 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 westward displacement of the TCC centroid. This pattern is consistent with previous studies showing that many regions with significant forest recovery overlapped with areas targeted by the Natural Forest Conservation Program (NFCP), including the Qinling–Daba Mountain region and parts of Sichuan, where notable gains in forest cover were observed (Viña et al., 2016).

To further characterize the long-term greening trend, a pixel-wise trend analysis of TCC was conducted (Fig. 10). The results indicate that 30.22% of the pixels experienced a significant increase in tree canopy cover, whereas only 2.86% showed a significant decrease. Most pixels (63.95%) remained unchanged, whereas pixels with non-significant gain and loss represented 2.06% and 0.91%, respectively (Fig. 10a). Overall, TCC dynamics in China were dominated by canopy gain and stability. Spatially, significant gains were mainly distributed in Central, South, East, and Southwest China, whereas unchanged areas were concentrated in North and Northwest China (Fig. 10a). Significant loss occurred only in limited and scattered areas. The regional statistics show that South China had the highest proportion of significant gain, followed by Northeast China, while North and Northwest China were characterized primarily by unchanged pixels (Fig. 10b). In all regions, the proportion of significant loss remained low. At the provincial scale, the average rate of change showed clear spatial differences (Fig. 10c). Chongqing, Yunnan, Guizhou, and Guangxi exhibited relatively high positive rates, indicating significant increases in TCC in Southwest China (Fig. 10c). Overall, these results suggest a generally positive trajectory of tree canopy dynamics across China, reflecting the notable effectiveness of afforestation and ecological restoration efforts. 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 reflect 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 reflect 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 broad-leaf forests and secondary vegetation with more homogeneous plantation systems (Ziegler et al., 2009).

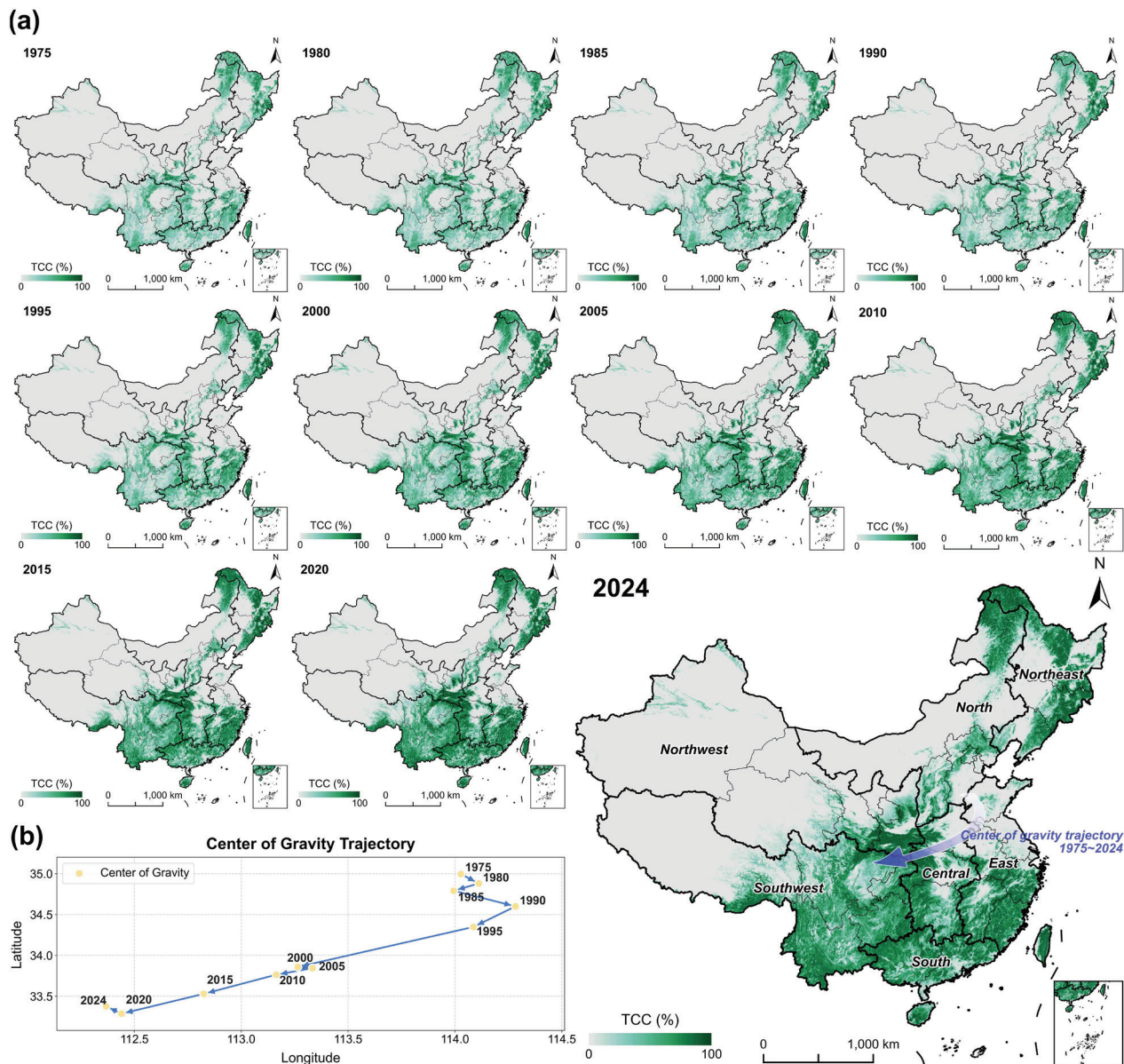


Figure 9. (a) Spatial distribution of TCC in China for 11 representative years from 1975 to 2024 based on ChinaTCC30, and (b) trajectory of the TCC center-of-gravity during the period.

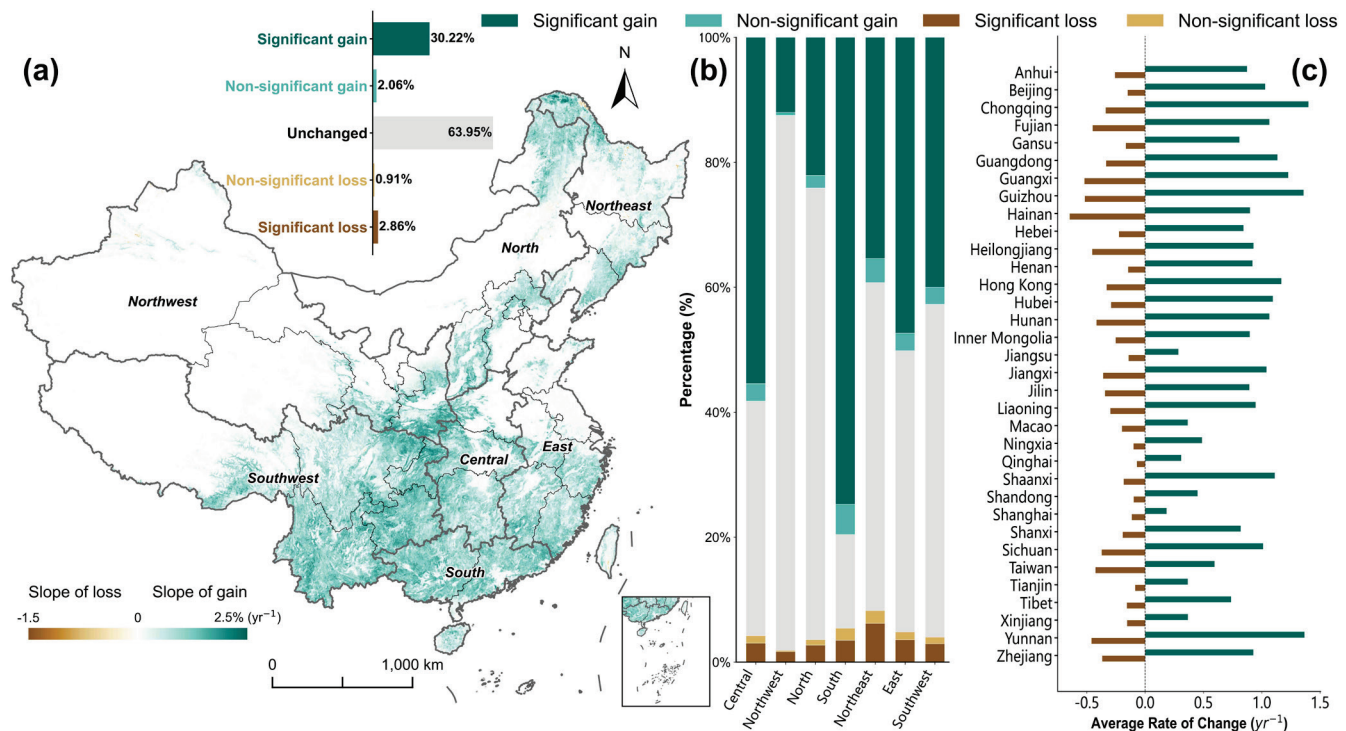


Figure 10. Spatiotemporal trends of ChinaTCC30 in China during 1985–2024. (a) Spatial distribution of the change trend at the pixel level. The inset bar chart shows the national area percentages of five trend categories: significant gain, non-significant gain, unchanged, non-significant loss, and significant loss (significance is defined at $p < 0.05$). (b) Stacked bar chart illustrating the area proportions of the five trend categories across the seven geographic regions of China. (c) Diverging bar chart showing the average rate of change (yr^{-1}) for pixels experiencing significant gain (green bars, right) and significant loss (brown bars, left) aggregated at the provincial level.

To further evaluate the temporal reliability of ChinaTCC30, representative canopy gain and loss pixels were visually inspected using multi-temporal high-resolution imagery (Fig. 11). The results show good agreement between the annual TCC trajectories and the observed land-cover changes. For the gain pixels, the image series show a clear transition from sparse vegetation or open land to dense TCC, which is consistent with the continuous increase in the TCC time series (Fig. 11a). For the loss pixels, the imagery reveals evident canopy disturbance or clearing, and the corresponding TCC trajectories capture these declines well, including both abrupt decreases and more complex fluctuating patterns (Fig. 11b). Overall, the inspection results indicate that ChinaTCC30 can effectively capture both gradual canopy cover gain and canopy loss, further supporting the temporal consistency and reliability of the dataset.

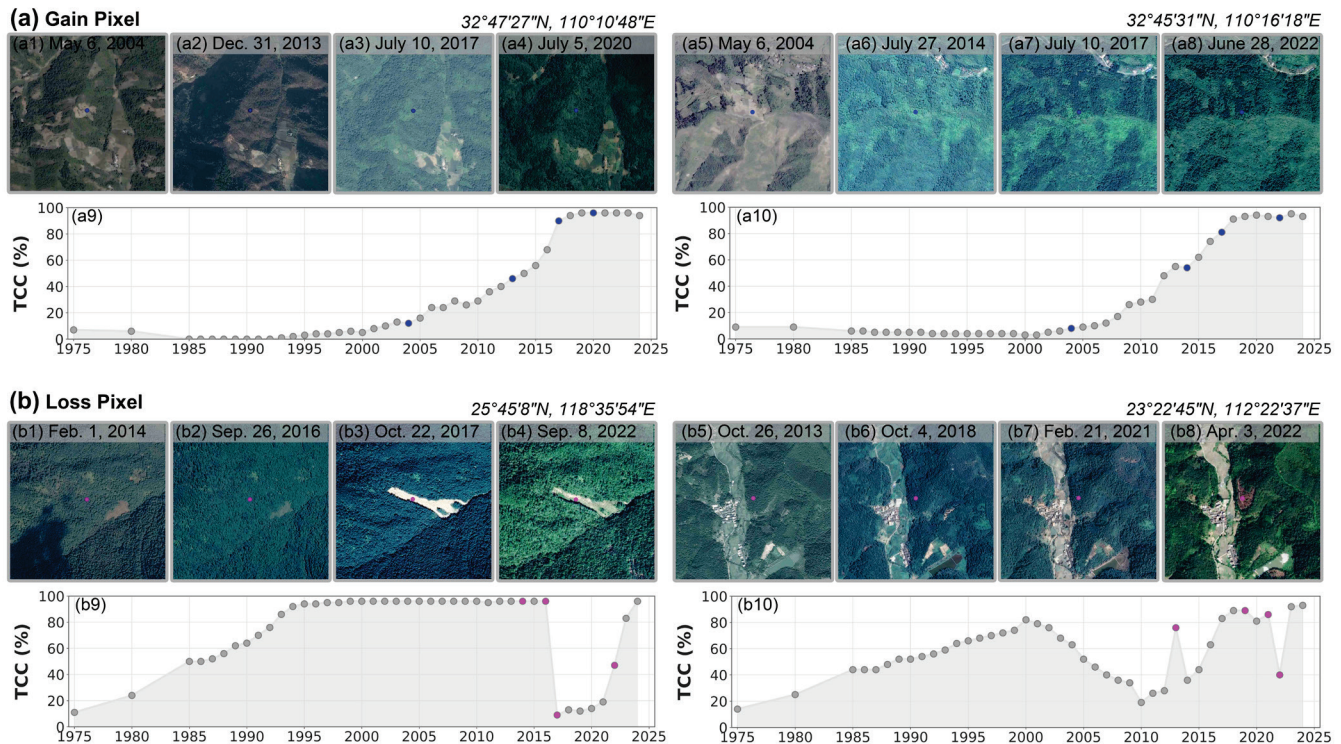


Figure 11. Canopy gain and loss pixels illustrating annual TCC dynamics derived from ChinaTCC30. (a) Gain-pixel examples and (b) loss-pixel examples. Upper panels show multispectral images from multiple years, and lower panels show annual TCC trajectories. Blue and magenta dots mark the target pixels in gain and loss cases, respectively; gray dots denote yearly TCC values, and colored dots indicate the years of the displayed images. Images: © 2025 CNES / Airbus, © 2025 Maxar Technologies.

4 Discussion

4.1 Comparisons to existing mapping products

4.1.1 TCC mapping comparison based on multi-source reference data

To comprehensively evaluate the spatial fidelity and temporal robustness of our ChinaTCC30 product, we first compared ChinaTCC30 with several publicly available tree canopy cover products, including MODIS VCF (250 m), DWTP (10 m), and CATCD (30 m) (see Table S2 for dataset details), using VHR photo-interpreted TCC samples as reference (Fig. 12). We then further validated the product using independent canopy cover estimates derived from GEDI spaceborne LiDAR observations (Fig. 13). Finally, temporal robustness was assessed through comparisons with multi-period canopy cover observations from the NFI (Fig. 14). Together, these analyses provide a systematic assessment of the relative performance, external consistency, and temporal stability of ChinaTCC30.



The comparison with photo-interpreted reference data shows clear differences among the four products (Fig. 12). MODIS VCF had the lowest agreement with the reference samples ($R^2 = 0.46$, $RMSE = 27.05\%$) and showed evident underestimation in areas with high TCC (Fig. 12a). DWTP showed a relatively high correlation ($R^2 = 0.63$, $r = 0.83$), yet it still exhibited a pronounced negative bias of 7.59% (Fig. 12b). CATCD further improved the agreement with the reference data, achieving an R^2 of 0.78, an $RMSE$ of 17.10%, and a slight positive bias of 1.65% (Fig. 12c). In contrast, ChinaTCC30 showed the best agreement with the reference data, with the highest R^2 (0.83) and a near-zero bias (0.39%), and its estimates were less scattered around the 1:1 line (Fig. 12d). Specifically, comparisons against visually interpreted samples indicate that MODIS VCF is limited in capturing high TCC, with estimated values rarely exceeding 90%, which might be due to its relatively coarse spatial resolution (250 m). At this scale, individual pixels with nearly pure forest canopy areas approaching 100% TCC are scarce, especially in area with high spatial heterogeneity (Fig. 16). This mixed-pixel effect constrains the ability of MODIS VCF to effectively characterize stands with high canopy cover, thereby leading to an underrepresentation of saturated TCC conditions. In addition, previous studies have shown that MODIS VCF tends to overestimate TCC in sparsely forested regions while underestimating it in densely forested areas. The underestimation in high-cover regions may be related to saturation of MODIS VCF at around 80%, whereas the overestimation in low-cover regions may result from strong spectral confusion between trees and dense herbaceous vegetation (Sexton et al., 2013). DWTP exhibits a similar pattern of bias. It should be noted that the “tree” band in the DWTP dataset is derived from a land cover product and essentially represents the probability that a pixel is fully covered by trees as estimated from Sentinel-2 imagery, rather than a direct measurement of percentage tree canopy cover (Brown et al., 2022). Comparison with visually interpreted samples shows that DWTP can describe canopy cover at low to intermediate levels, but it still tends to overestimate TCC in sparse forest regions and underestimate it in dense forest areas. Therefore, caution is needed when using either MODIS VCF or DWTP to characterize canopy cover in high-density forest regions. By contrast, ChinaTCC30 shows the strongest agreement with the visually interpreted reference data ($R^2 = 0.83$; $RMSE = 15.14\%$; Fig. 12d). Boxplot analysis stratified by tree canopy cover classes further highlights the advantage of ChinaTCC30 (Fig. 12e). In both sparse-canopy areas (0–20%) and dense-canopy areas (80–100%), ChinaTCC30 exhibits narrower interquartile ranges than the existing datasets, indicating a stronger ability to capture both low and high canopy-cover values.

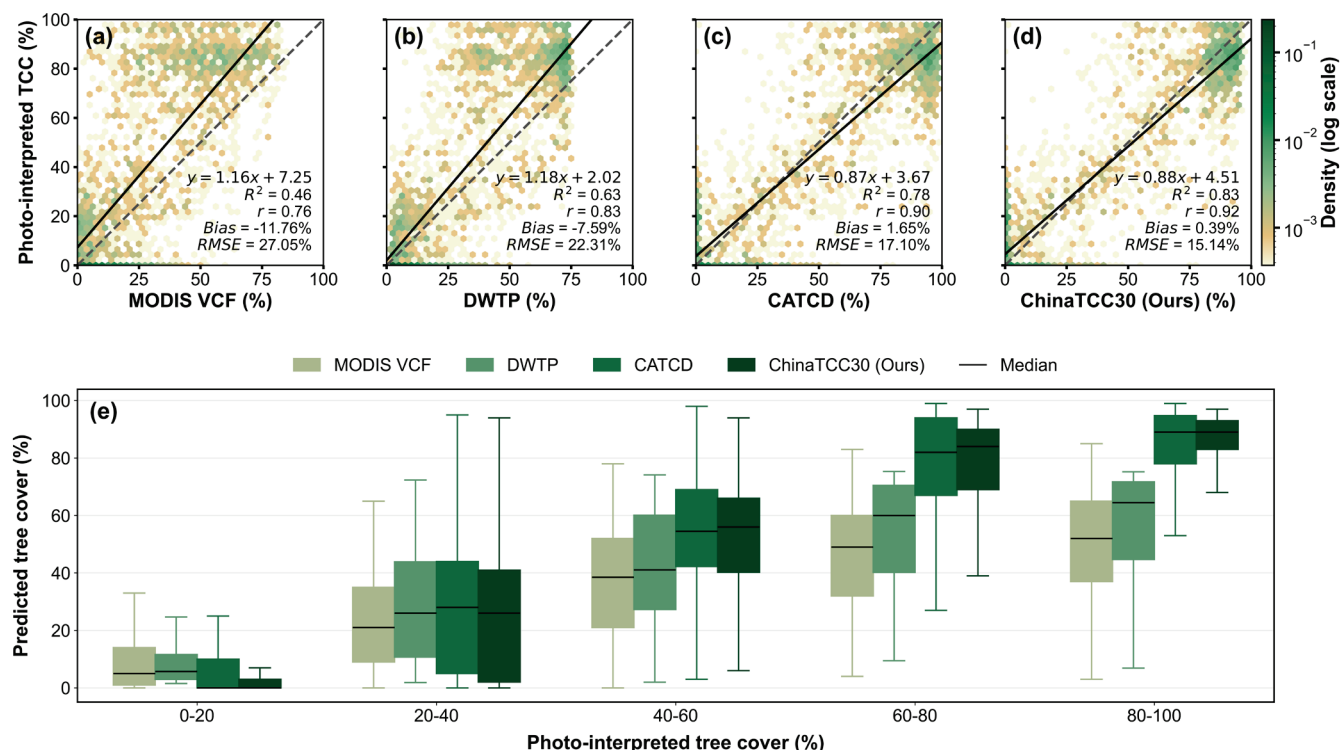


Figure 12. Validation of four tree canopy cover products against photo-interpreted reference data. Scatter density plots comparing the photo-interpreted tree cover with (a) MODIS VCF, (b) DWTP, (c) CATCD, and (d) ChinaTCC30 (Ours).

Consistency analysis based on GEDI products showed that all canopy cover products were positively correlated with GEDI observations, while ChinaTCC30 exhibited relatively higher agreement (Fig. 13). Specifically, the R^2 values for MODIS VCF, DWTP, CATCD, and ChinaTCC30 were 0.36, 0.48, 0.50, and 0.54, respectively, and the corresponding RMSE values were 21.56%, 19.54%, 19.13%, and 18.24% (see Table S2 for dataset details). Among the products, ChinaTCC30 exhibited the highest R^2 and lowest RMSE ($R^2 = 0.54$; RMSE = 18.24%), demonstrating greater consistency with GEDI-derived canopy cover observations than the other products. Previous studies have shown that GEDI retrieves canopy structure-related metrics from the vertical distribution of full-waveform Lidar return energy at the footprint scale, enabling it to capture continuous variations in canopy cover and ground exposure and making it sensitive to structural heterogeneity within footprints (Lefsky et al., 2002; Dubayah et al., 2020; Hofton et al., 2020). In contrast, products based on coarser grid-based proportion representation methods (e.g., a 7×7 grid-based approach) may be less effective than finer-grid methods in capturing subtle variation in canopy cover and local spatial heterogeneity, which can lead to weaker agreement with GEDI-derived estimates. By contrast, the stronger agreement of ChinaTCC30 with GEDI may be associated with its 10×10 grid-based labeling protocol, which emphasizes the continuous representation of canopy cover proportions during sample annotation, thereby reducing over-smoothing and better preserving the continuity of canopy cover variation.

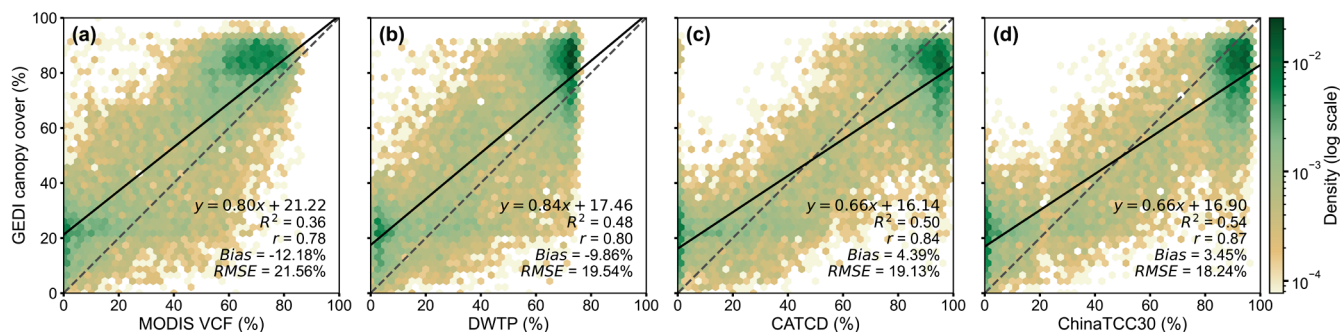


Figure 13. Validation of four tree canopy cover products against GEDI canopy cover. Scatter density plots comparing the GEDI-derived canopy cover with (a) MODIS VCF, (b) DWTP, (c) CATCD, and (d) ChinaTCC30.

Comparisons with data from the 6th to the 9th NFI (NFGA, 2005, 2019) indicate that both CATCD and ChinaTCC30 are able to capture variations in TCC reasonably well, whereas ChinaTCC30 consistently shows higher R^2 of fit and lower error across all inventory periods (Fig. 14). For CATCD, the R^2 value increased gradually from 0.63 in the 6th inventory to 0.74 in the 9th inventory, while the RMSE decreased from 19.75% to 17.65%, indicating an improvement in performance during the later inventory periods. By contrast, the R^2 values of ChinaTCC30 reached 0.66, 0.76, 0.79, and 0.82 for the 6th to the 9th inventories, respectively, all of which were higher than those of CATCD for the corresponding periods. Its RMSE values were 18.81%, 16.63%, 15.58%, and 14.68%, respectively, consistently lower than the corresponding CATCD values of 19.75%, 18.23%, 17.95%, and 17.65%. These results demonstrate that ChinaTCC30 not only has stronger explanatory power, but also provides more stable and accurate estimates of tree canopy cover at the plot scale. Temporally, both products show an increasing trend in R^2 and a decreasing trend in RMSE, suggesting that their agreement with NFI observations improved in the later inventory periods. This pattern may partly reflect the characteristics of the early NFI reference data. In the earlier inventory periods, canopy-related variables were often recorded using class-based reporting, or rounded percentage values, which produced a more discrete distribution of observations and may have weakened the apparent correspondence with continuous remote-sensing estimates (Fig. 13a) (NFGA, 2014). However, ChinaTCC30 maintained consistently higher R^2 and lower RMSE than CATCD across all four inventory periods, indicating that its advantage is temporally robust rather than confined to a specific stage. Overall, compared with CATCD, ChinaTCC30 exhibits superior accuracy and robustness in tree canopy cover estimation, highlighting its potential for regional-scale mapping and long-term dynamic monitoring by combined active learning and random forest.

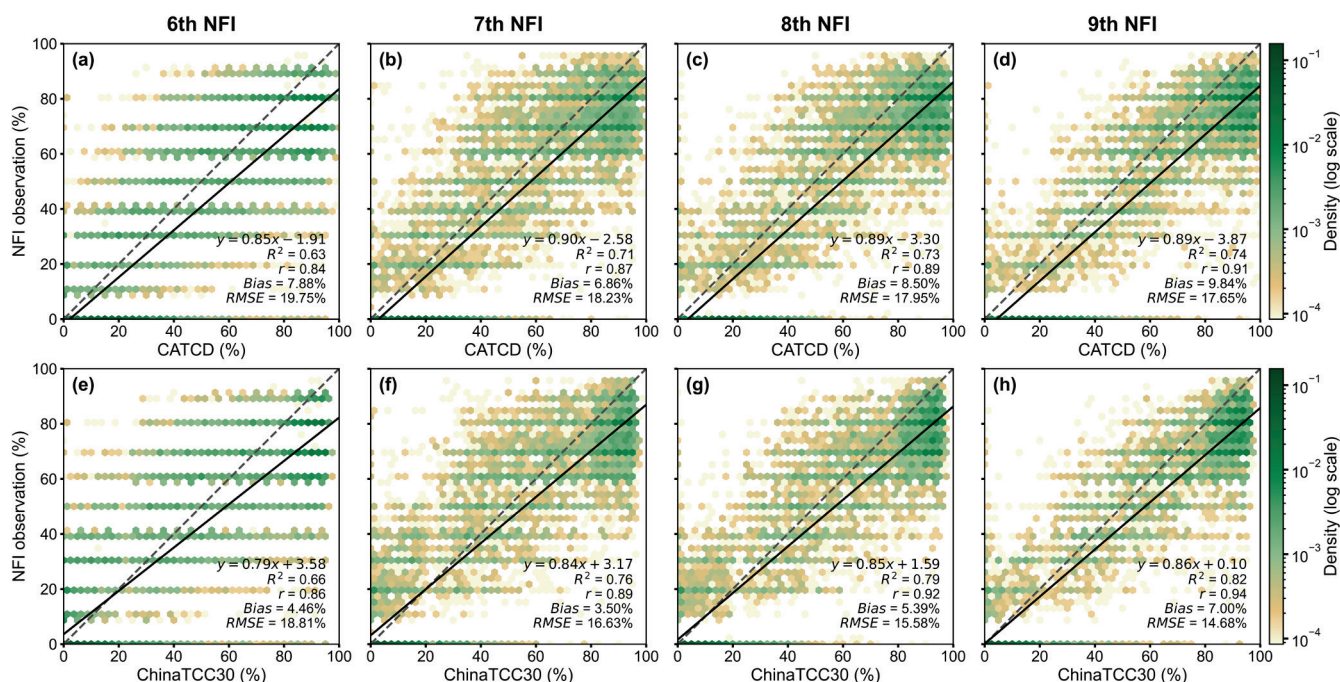


Figure 14. Validation of ChinaTCC30 and CATCD products against NFI observations. Scatter density plots comparing NFI ground observations with CATCD (top row: a–d) and ChinaTCC30 (bottom row: e–h) tree canopy cover estimates. The columns represent four consecutive inventory periods: the 6th (a, e), 7th (b, f), 8th (c, g), and 9th (d, h) NFI.

4.1.2 Static and spatial comparison

We compared ChinaTCC30 at the pixel level with other 30-m tree canopy cover products developed by Hansen et al. (2013), Cai et al. (2024), and Sexton et al. (2013) for the years 2000, 2010, and 2015, respectively (see Table S2 for details of these datasets). The results showed that the mean differences of ChinaTCC30 relative to CATCD, GFC, and GFCC were -1.40%, 2.26%, and 5.80%, respectively (Fig. 15). Overall, ChinaTCC30 showed the highest consistency with CATCD, whereas larger discrepancies were observed with GFC and GFCC (Fig. 15). The difference histograms further showed that most pixel-level differences between ChinaTCC30 and CATCD were concentrated within -5% to 5%, indicating that the two products depicted the national spatial pattern of tree canopy cover in a broadly consistent manner. In contrast, the differences between ChinaTCC30 and both GFC and GFCC were more dispersed and predominantly positive, suggesting that the latter two products generally underestimated tree canopy cover relative to ChinaTCC30 across China.

Although ChinaTCC30 shows high overall agreement with CATCD, some systematic spatial differences remain. In particular, areas with differences ranging from -5 to 0 account for more than 25% of the total area, indicating that TCC values in ChinaTCC30 are generally slightly lower than those in CATCD. Previous studies have shown that different observation methods vary in their sensitivity to canopy gaps and canopy structure, which can in turn affect canopy cover estimates and their remote sensing representation (Jennings et al., 1999; Korhonen et al., 2006; Fiala et al., 2006). In this study, tree canopy cover proportion was interpreted using a 10×10 grid. Compared with a sparser sampling scheme, the higher density of observation



points can more effectively capture continuous variations in canopy cover and locally non-closed areas within a plot, thereby improving the representation of canopy details and sub-pixel heterogeneity. However, this stricter interpretation of canopy cover is also more likely to detect local gaps in partially closed canopies, which may lead to slightly lower estimates than those derived from sparser sampling methods.

555 Compared with CATCD, the differences between ChinaTCC30 and GFC or GFCC were larger and showed stronger regional patterns. Positive differences relative to GFC were concentrated in northern and northeastern China, where pixels with differences greater than 20% accounted for more than 20% of the total, indicating clear underestimation by GFC in northern forest regions and areas with low to moderate TCC. This may be related to its limited ability to detect sparse trees, agroforestry mosaics, and leaf-on canopy conditions in deciduous forests, as well as differences in image acquisition timing,
 560 annual compositing windows, and growing-season image quality.

Positive differences between ChinaTCC30 and GFCC were broader and larger in magnitude, with pixels exceeding 20% widely distributed across major forest regions, indicating stronger underestimation by GFCC in high-cover forests. This is likely related to the derivation of GFCC from the 250-m MODIS VCF product. Previous studies have shown that MODIS VCF tends to overestimate TCC in sparse tree areas and underestimate it in dense forests, with the latter likely caused by
 565 saturation near 80% canopy cover (Sexton et al., 2013).

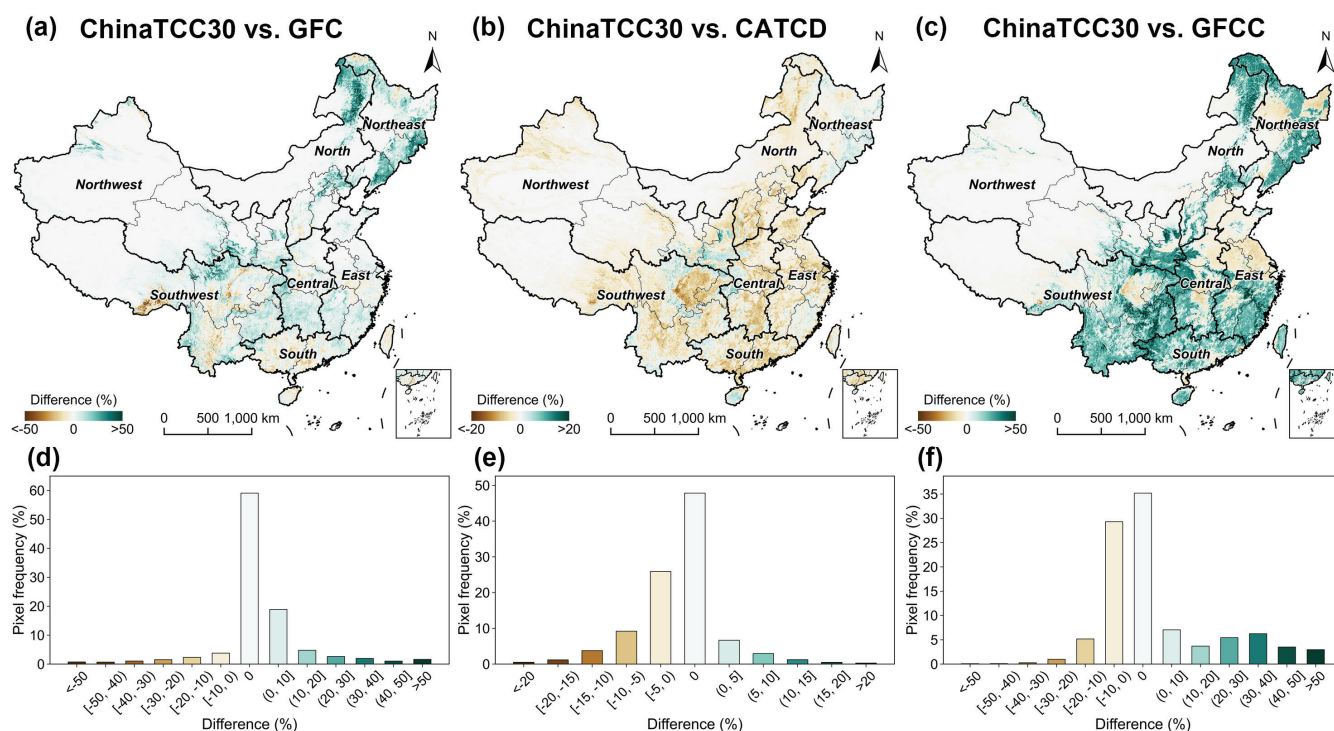


Figure 15. Spatial distribution maps of the differences between ChinaTCC30 and three other products and the corresponding frequency distribution histograms. (a, d) ChinaTCC30 vs. GFC; (b, e) ChinaTCC30 vs. CATCD; and (c, f) ChinaTCC30 vs. GFCC.



570 We compared the spatial performance of different tree canopy cover products across eight representative land-cover sites, including bare land, urban built-up areas, cropland mosaics, and mountainous forests with varying canopy densities (Fig. 16). Overall, MODIS VCF captures regional-scale canopy cover gradients, but its coarse spatial resolution leads to pronounced blocky smoothing in urban and agroforestry transitional areas, limiting its ability to resolve fine-scale tree patches and forest edge structures (Hansen et al., 2003; Sexton et al., 2013). GFCC provides a relatively stable background representation in
575 medium- and high-cover forests, but has limited capacity to preserve boundaries in fragmented landscapes; in addition, striping artifacts and gap-like omissions are evident in some local areas (Fig. 16 6d, 6g, 6h). GFC performs well in densely forested areas, but is relatively less effective in representing continuous canopy cover gradients under low-cover and non-forest backgrounds (Fig. 16 4b–4d, 4h).

By contrast, both CATCD and ChinaTCC30 show clearer tree canopy cover patterns across multiple land-cover types
580 and more closely match the landscape configuration in the reference imagery. Their advantages are particularly evident in complex scenes, such as urban greening areas, tree belts within cropland mosaics, fragmented forests, and canopy gaps embedded within dense forests, where they capture the spatial heterogeneity of canopy cover in greater detail. Notably, compared with CATCD, ChinaTCC30 further improves the identification and quantification of tree canopy cover across different land-cover types by incorporating land-use data as prior knowledge in the model. This strategy effectively reduces false-positive
585 canopy signals over bare land (Fig. 16 a7–10), as well as false positives in non-tree areas within impervious urban surfaces and agroforestry mixed systems (Fig. 16 b7–10, c7–10).

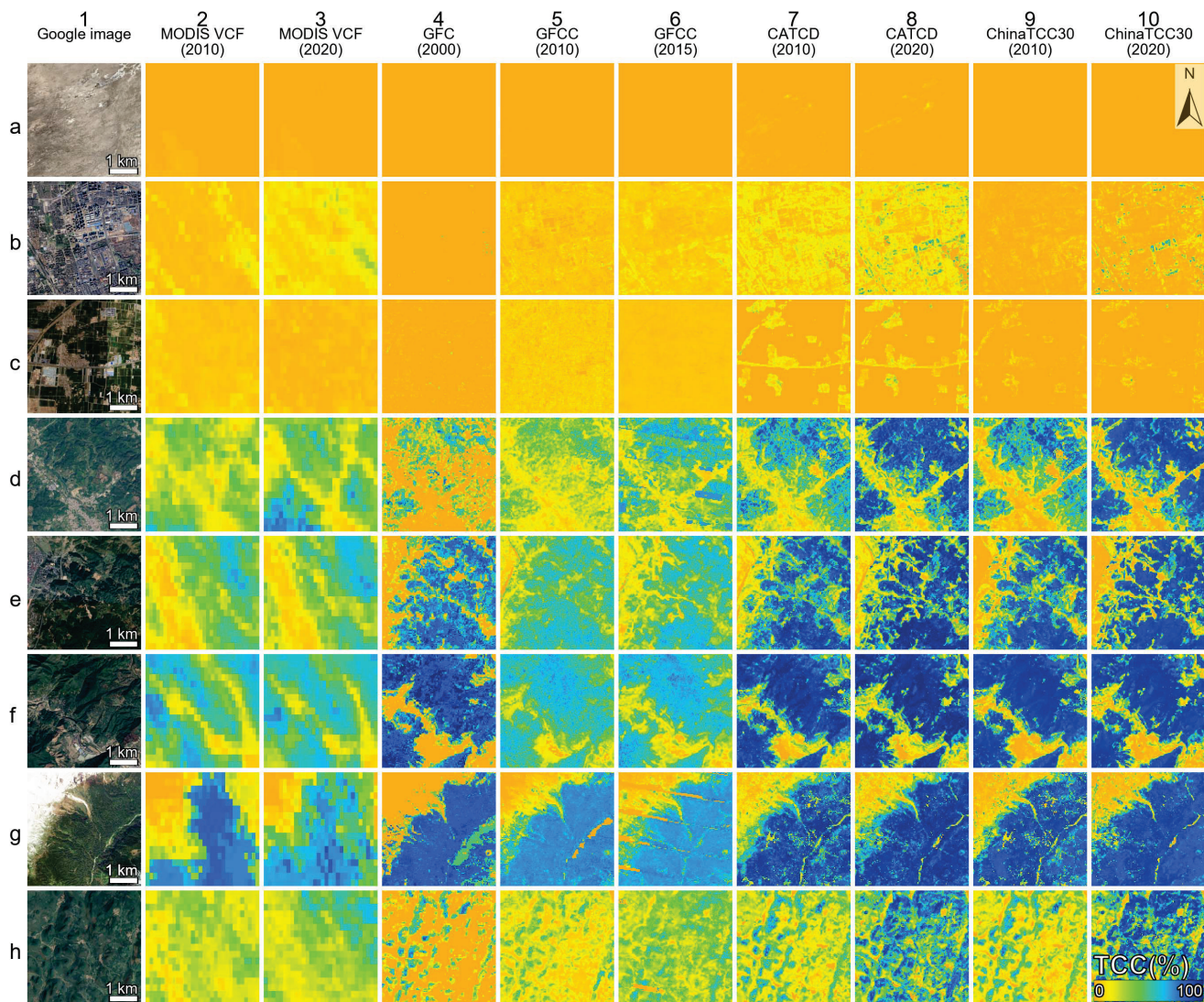


Figure 16. Visual comparison of tree canopy cover (TCC)-related products across eight representative regions. Rows (a–h) represent eight different regions in China. Column (1) shows multispectral images; columns (2)–(3) show the MODIS VCF products for 2010 and 2020, respectively; column (4) shows the GFC product for 2000; columns (5)–(6) show the GFCC products for 2010 and 2015, respectively; columns (7)–(8) show the CATCD products for 2010 and 2020, respectively; and columns (9)–(10) show the ChinaTCC30 products for 2010 and 2020, respectively. Individual panels can be identified by their row–column labels (e.g., a1, d7, h10). Images: © 2025 CNES / Airbus, © 2025 Maxar Technologies.

4.1.3 Dynamic and trend comparison

To further assess the reliability of the results, we compared the temporal change patterns derived from ChinaTCC30 with those from CATCD. The statistics show that, for the period 1985–2024, the proportions of pixels with significantly increased, significantly decreased, and relatively stable canopy cover in ChinaTCC30 were 30.22%, 2.86%, and 63.95%, respectively, whereas the corresponding proportions in CATCD were 33.57%, 4.71%, and 57.30%. The discrepancies between the two



products in the proportions of change may be attributable to differences in calibration periods, reference sample sources, and
 600 model development strategies (Quegan et al., 2019; Santoro et al., 2021). Nevertheless, the two datasets show strong agreement
 in their overall change characteristics, namely that canopy cover remained stable across large parts of China, while the pro-
 portion of increasing areas was substantially greater than that of decreasing areas.

The ChinaTCC30 dataset indicates that forest canopy cover in China experienced a slight decline during 1975–1985,
 followed by an overall and persistent increasing trend during 1985–2024 (Fig. 17). Comparison with the forest inventory data
 605 shows that the early decline is broadly consistent with inventory-based estimates. The increase after 1985 points to a continued
 recovery of Chinese forests over the past several decades and also highlights the substantial effectiveness of forest conservation
 and management measures. The trajectory derived from the ChinaTCC30 dataset is generally consistent with China’s long-
 term afforestation and ecological protection policies (Zeng et al., 2023). In addition, we compared forest area changes esti-
 mated from different datasets to examine differences among products in characterizing forest extent (Fig. 17). In China TCC30,
 610 forest extent was delineated using a two-step strategy (see Sect. 2.3.5 for details). The comparison shows that the forest areas
 estimated from CLCD, GlobeLand30, and ALCC were generally larger than those derived from MODIS VCF, GFCC, CATCD,
 and ChinaTCC30, the latter being continuous canopy-cover products. This difference mainly arises because land-cover prod-
 ucts such as CLCD and GlobeLand30 typically count the full area of any pixel classified as forest, which tends to produce
 inflated forest area estimates (Cai et al., 2024). By contrast, continuous products can more explicitly represent the actual
 615 proportion of canopy cover within each pixel. More specifically, among the land use/land cover datasets, CLCD and ALCC
 were less capable of capturing the persistent increase in forest area across China. Among the compared products, only GFC
 and GLC_FCS30D showed declines in forest area, with GFC declining during 2001–2010 and GLC_FCS30D continuing to
 decline throughout 2001–2022. MODIS VCF and GFCC showed some increasing tendency, but both substantially underesti-
 mated forest area relative to the NFI, with an average underestimation of approximately 45% (Cai et al., 2024). By comparison,
 620 CATCD and ChinaTCC30 not only captured the trajectory of forest change in China more effectively, but also produced forest
 area estimates that were more consistent with the NFI results. Among them, ChinaTCC30 showed better agreement with the
 NFI than CATCD in both the magnitude and temporal range of change.

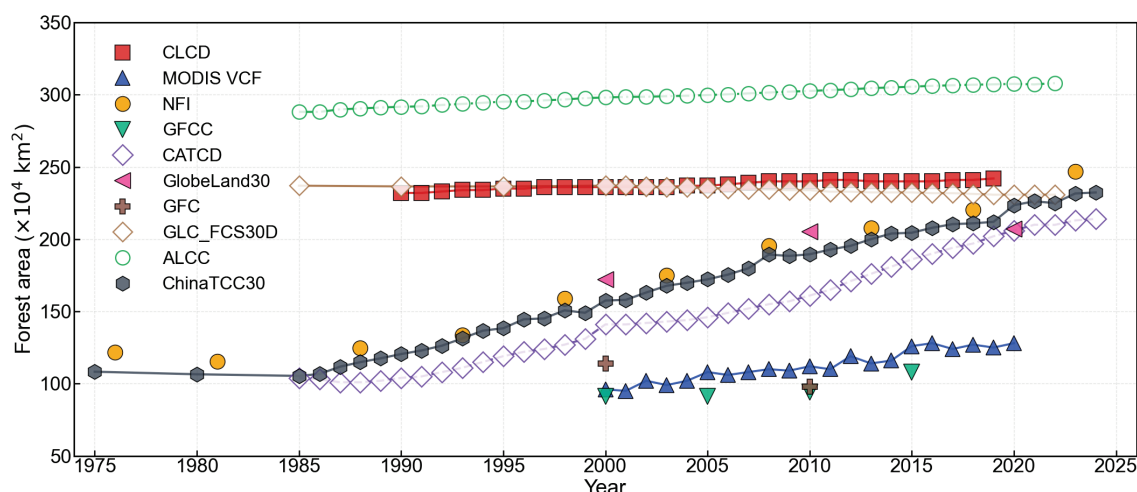


Figure 17. Time-series dynamics of forest area in China derived from multiple datasets. The abbreviations of the datasets and their corresponding full names are as follows: CLCD, China Land Cover Dataset; MODIS VCF, Moderate Resolution Imaging Spectroradiometer Vegetation Continuous Fields; NFI, National Forest Inventory; GFCC, Global Forest Cover Change; CATCD, China Annual Tree Cover Dataset; GlobeLand30, Global Land Cover with 30 m resolution; GFC, Global Forest Change; GLC_FCS30D, global 30 m land-cover product with a fine classification system; ALCC, annual land cover product for China; and ChinaTCC30 (ours).

Our analysis suggests that abrupt changes in TCC observed in some localized areas, primarily in low-TCC regions, may be largely associated with sensor transitions. Although the preprocessed Landsat data had already undergone substantial efforts to improve cross-sensor consistency, the transferability of correction parameters may still be challenged under certain localized conditions (Roy et al., 2016a; Cai et al., 2025a). In addition, the limited temporal representativeness or availability of early-period canopy-cover reference samples may hinder the model from adequately learning the relationship between spectral features and TCC. Therefore, the observed discontinuities are more likely to result from the combined influence of multiple factors, including spectral differences among sensors, interannual fluctuations in image availability and quality, and potential limitations in the temporal correspondence or availability of measurement data used as model inputs during specific periods.

To mitigate such abrupt interannual changes in TCC, we identified pixels with temporally stable land-cover types based on the same-resolution ALCC land-cover dataset and used them as references to correct abrupt canopy-cover discontinuities associated with temporal breakpoints. Specifically, local trends before and after each breakpoint were estimated, and significant jumps observed in stable pixels were interpreted as non-physical shifts caused by changes in sensor characteristics, source-data conditions, or model-transfer uncertainty, rather than actual land-surface change. These discontinuities were then corrected accordingly. This approach is informed by previous studies that employed stable targets for cross-temporal calibration, while further incorporating local temporal trajectory information to identify and correct anomalous breakpoints (Schott et al., 1988; Du et al., 2002; Verbesselt et al., 2010; Zhu and Woodcock, 2014). Our results suggest that this procedure effectively reduces breakpoint biases in canopy-cover mapping, thereby suppressing pseudo-trends and enhancing the robustness of long-term forest dynamics assessments (Fig. 18). The CATCD product exhibits pronounced discontinuities around the 1999–2000



transition, manifested as a sharp and spatially extensive increase in TCC across a large area. Such localized but widespread changes are difficult to reconcile with plausible vegetation dynamics. By contrast, ChinaTCC30 maintains a comparatively stable spatial pattern and annual trajectory. This interpretation is further supported by area statistics: in one representative region, the non-forest area derived from CATCD increased abruptly from nearly zero to more than 40 km² in 2000, whereas the corresponding estimate from ChinaTCC30 remained relatively stable (Fig. 18c, d). Together, these examples demonstrate that, by using temporally stable pixels and local trajectory information to identify anomalous breakpoints, the proposed method can effectively remove non-physical abrupt changes while preserving plausible spatial variability, thereby improving the reliability of long-term canopy-cover mapping and forest-dynamics assessment.

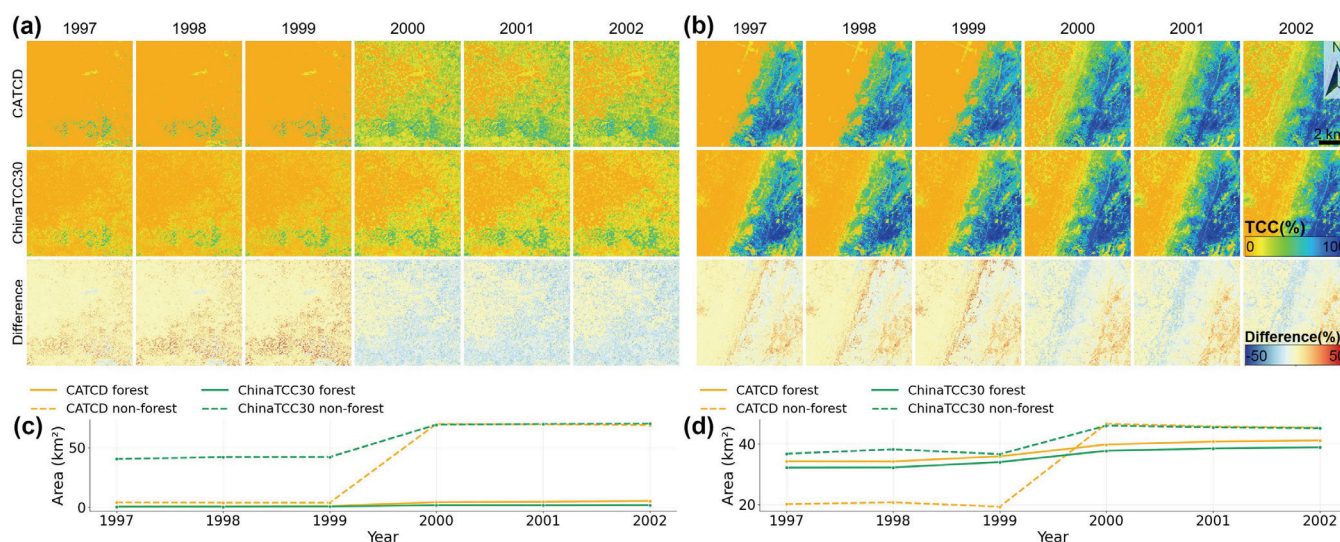


Figure 18. Comparison of annual TCC dynamics derived from CATCD and ChinaTCC30 in two representative regions during 1997–2002: (a, b) annual TCC (%) maps from CATCD and ChinaTCC30, along with their difference maps (ChinaTCC30 – CATCD); (c, d) annual forest and non-forest area statistics for the corresponding regions.

4.2 Limitations and future improvement

ChinaTCC30 provides baseline estimates of TCC for 1975 and 1980, with annual updates from 1985 to 2024. The product was designed to characterize the long-term dynamics of TCC in China as accurately as possible and to represent these changes consistently within a continuous time series. To quantify mapping uncertainty, we used the standard deviation of model cross-validation results as an uncertainty metric during the mapping design stage, thereby reflecting the stability of predictions across regions and periods. Overall, the uncertainty of ChinaTCC30 mainly arises from the following sources.

First, spectral confusion among different land-cover types is an important source of mapping uncertainty. In areas with complex land surface composition and limited spectral separability, the identification and quantitative estimation of tree canopy cover are inherently more difficult (Foody, 2002; Sexton et al., 2013). To mitigate this issue, land use/land cover data were incorporated into the model inputs to improve the discrimination of tree canopy cover under different background conditions. This strategy effectively reduced false positives of TCC over bare surfaces, as well as the misclassification of non-tree



areas as tree canopy cover in urban impervious surfaces and agroforestry mosaics (Fig. a9–10, b9–10). Nevertheless, in vegetation transition zones where woody and herbaceous cover are intermingled, eliminating such confusion remains difficult because these vegetation types can exhibit similar spectral responses and seasonal characteristics (Hill and Schütt, 2000; Brandt et al., 2020b). This issue is particularly pronounced in high-latitude and arid to semi-arid regions, which therefore tend to exhibit higher uncertainty (Fig. 16c). Second, complex topographic conditions further increase the uncertainty of ChinaTCC30. In mountainous and high-latitude regions, terrain shadow, slope-aspect differences, and local variations in illumination geometry can substantially affect surface reflectance (Teillet et al., 1982; Civco, 1989), thereby weakening the relationship between remotely sensed signals and actual canopy structure. Although standard topographic correction can reduce these effects to some extent, shadowing and radiometric distortion may still introduce local biases in areas with strong topographic relief. Third, differences in Landsat data quality constitute a key source of uncertainty in this long-term product. Landsat is one of the longest-running and most widely used moderate-resolution Earth observation programs and provides an irreplaceable foundation for long-term vegetation monitoring (Wulder et al., 2019). However, Landsat sensors from different periods differ in radiometric quality, geometric accuracy, observation frequency, and the number of available images (Roy et al., 2014; Wulder et al., 2016). These differences are especially important in the earlier period, when data gaps, acquisition conditions, cloud and shadow contamination, and limited image availability during the growing season may have propagated additional uncertainty into annual mapping results (White et al., 2014; Zhu et al., 2015). Therefore, although a unified preprocessing and modeling framework was adopted to maximize temporal consistency, differences in the quality of historical remote sensing data remain an unavoidable source of uncertainty.

In addition, the uncertainty of the early baseline estimates for 1975 and 1980 is closely related to the historical backcasting 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 backcasting 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 backcasting. Although these early results should be regarded as model-constrained historical estimates rather than direct observations, the ecologically constrained backcasting strategy applied only to natural forests provides a reasonable and practical pathway for constructing a long-term continuous tree canopy cover product under conditions of limited contemporaneous observations. This also establishes an important data foundation for investigating changes in TCC across China over the past half century.

5 Data availability

The ChinaTCC30 dataset developed in this study is publicly available from Zenodo and can be accessed at the following addresses. It contains baseline TCC estimates for 1975 and 1980, together with annual TCC products for 1985–2024. For each year, two GeoTIFF files are provided: a TCC raster and a corresponding uncertainty raster. The dataset is distributed in seven



parts: Part 1 (1975, 1980, and 1985–1988), available at <https://doi.org/10.5281/zenodo.21194015>; Part 2 (1989–1994), available at <https://doi.org/10.5281/zenodo.21195911>; Part 3 (1995–2000), available at <https://doi.org/10.5281/zenodo.21199497>; Part 4 (2001–2006), available at <https://doi.org/10.5281/zenodo.21245669>; Part 5 (2007–2012), available at <https://doi.org/10.5281/zenodo.21262725>; Part 6 (2013–2018), available at <https://doi.org/10.5281/zenodo.21263677>; and Part 7 (2019–2024), available at <https://doi.org/10.5281/zenodo.21265832> (Yu et al., 2026a, b, c, d, e, f, g).

6 Conclusion

In this study, we developed an integrated framework combining AL and ML to reconstruct the spatiotemporal distribution of TCC across China for 1975, 1980, and annually from 1985 to 2024, and further quantified its uncertainty. Using TCC samples derived from visual interpretation of high-resolution imagery, the framework achieved robust mapping performance. Validation with multiple independent reference datasets demonstrated the reliability of ChinaTCC30, including visually interpreted samples ($R^2 = 0.83$, RMSE = 15.14%) and multi-period National Forest Inventory (6th – 9th NFI; $R^2 = 0.66 – 0.82$, RMSE = 14.68% – 18.81%). These results confirm that the proposed framework can effectively capture the long-term dynamics of TCC across China while maintaining relatively high accuracy at low mapping cost. The long-term TCC record reveals that China has experienced an overall stable-to-increasing canopy cover trend over the past decades. Approximately 30.22% of the study area showed significant increases in TCC, whereas only 2.86% exhibited significant declines, and 63.95% remained stable. Gravity-center analysis further indicated a southwestward shift in China's TCC distribution, reflecting regional differences in long-term forest recovery. Compared with existing products, ChinaTCC30 not only shows good spatial consistency at the national scale, but also improves the representation of fine-scale spatial heterogeneity and long-term temporal continuity. As a spatially explicit TCC dataset spanning nearly five decades, ChinaTCC30 provides valuable support for low-cost forest monitoring, climate change impact assessment, and the development of forest conservation and management policies.

Supplement

The supplementary materials for this article can be obtained from the submission file.

725 Acknowledgements

This work was supported by China National Science Foundation (No. 32361143869 and 32371663), National Key Research and Development Program of China (No. 2021YFD2200405), Basic Research Program of Jiangsu (BK20250044), Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. KYCX25_1601), and Schmidt Sciences, LLC through the CLARiTy project (ref: G-25-69369).



730 Author contributions

JY: Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing, Visualization. **ZY:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Writing – original draft, Writing – review & editing, Visualization. **WH:** Resources, Software, Visualization. **FZ:** Resources, Software. **SL:** Conceptualization, Funding acquisition.

735 Competing interests

The authors declare no competing interests. Zhen Yu is a member of the Editorial Board for Earth System Science Data but had no role in the peer review, editorial management, or final decision regarding this manuscript.

Declaration of Generative AI and AI-assisted Technologies

740 During the preparation of this manuscript, the authors used ChatGPT (OpenAI) to assist with language editing and improve the readability and clarity of the text. The authors reviewed and edited the output as necessary and take full responsibility for the content of this publication. The use of these tools did not influence the scientific content, interpretation, conclusions, or authorship of the work.

References

- Ahrends, A., Hollingsworth, P. M., Ziegler, A. D., Fox, J. M., Chen, H., Su, Y., and Xu, J.: Current trends of rubber plantation expansion may threaten biodiversity and livelihoods, *Global Environmental Change*, 34, 48-58, <https://doi.org/10.1016/j.gloenvcha.2015.06.002>, 2015.
- Bar-On, Y. M., Phillips, R., and Milo, R.: The biomass distribution on Earth, *Proceedings of the National Academy of Sciences*, 115, 6506-6511, <https://doi.org/10.1073/pnas.1711842115>, 2018.
- 750 Bastin, J.-F., Finegold, Y., Garcia, C., Mollicone, D., Rezende, M., Routh, D., Zohner, C. M., and Crowther, T. W.: The global tree restoration potential, *Science*, 365, 76-79, <https://doi.org/10.1126/science.aax0848>, 2019.
- Berger, K., Rivera Caicedo, J. P., Martino, L., Woher, M., Hank, T., and Verrelst, J.: A survey of active learning for quantifying vegetation traits from terrestrial earth observation data, *Remote Sensing*, 13, 287, <https://doi.org/10.3390/rs13020287>, 2021.
- 755 Brandt, M., Tucker, C. J., Kariryaa, A., Rasmussen, K., Abel, C., Small, J., Chave, J., Rasmussen, L. V., Hiernaux, P., and Diouf, A. A.: An unexpectedly large count of trees in the West African Sahara and Sahel, *Nature*, 587, 78-82, <https://doi.org/10.1038/s41586-020-2824-5>, 2020a.



- Brandt, M., Tucker, C. J., Kariryaa, A., Rasmussen, K., Abel, C., Small, J., Chave, J., Rasmussen, L. V., Hiernaux, P., Diouf, A. A., Kergoat, L., Mertz, O., Igel, C., Gieseke, F., Schöning, J., Li, S., Melocik, K., Meyer, J., Sinno, S., Romero, E., Glennie, E., Montagu, A., Dendoncker, M., and Fensholt, R.: An unexpectedly large count of trees in the West African Sahara and Sahel, *Nature*, 587, 78-82, 10.1038/s41586-020-2824-5, 2020b.
- Breiman, L.: Random Forests, *Machine Learning*, 45, 5-32, 10.1023/A:1010933404324, 2001.
- Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., and Ilyushchenko, S.: Dynamic World, Near real-time global 10 m land use land cover mapping, *Scientific data*, 9, 251, <https://doi.org/10.1038/s41597-022-01307-4>, 2022.
- Bryan, B. A., Gao, L., Ye, Y., Sun, X., Connor, J. D., Crossman, N. D., Stafford-Smith, M., Wu, J., He, C., and Yu, D.: China's response to a national land-system sustainability emergency, *Nature*, 559, 193-204, 2018.
- Burkhart, H. E. and Tomé, M.: Modeling forest trees and stands, Springer Science & Business Media 2012.
- Cai, Y., Li, X., Zhu, P., Nie, S., Wang, C., Liu, X., and Chen, Y.: China Earth Observation Data Cube: The 30-m Seamless Annual Leaf-On Landsat Composites from 1985 to 2023, *Journal of Remote Sensing*, 5, 0698, <https://doi.org/10.34133/remotesensing.0698>, 2025a.
- Cai, Y., Xu, X., Zhu, P., Nie, S., Wang, C., Xiong, Y., and Liu, X.: Unveiling spatiotemporal tree cover patterns in China: The first 30 m annual tree cover mapping from 1985 to 2023, *ISPRS Journal of Photogrammetry and Remote Sensing*, 216, 240-258, <https://doi.org/10.1016/j.isprsjprs.2024.08.001>, 2024.
- Cai, Y., Zhu, P., Li, X., Liu, X., Chen, Y., Shen, Q., Xu, X., Zhang, H., Nie, S., and Wang, C.: Dynamics of China's forest carbon storage: the first 30 m annual aboveground biomass mapping from 1985 to 2023, *Earth System Science Data*, 17, 6993-7018, <https://doi.org/10.5194/essd-17-6993-2025>, 2025b.
- Carreiras, J. M., Pereira, J. M., and Pereira, J. S.: Estimation of tree canopy cover in evergreen oak woodlands using remote sensing, *Forest ecology and management*, 223, 45-53, <https://doi.org/10.1016/j.foreco.2005.10.056>, 2006.
- Chen, C., Park, T., Wang, X., Piao, S., Xu, B., Chaturvedi, R. K., Fuchs, R., Brovkin, V., Ciais, P., and Fensholt, R.: China and India lead in greening of the world through land-use management, *Nature sustainability*, 2, 122-129, 2019.
- Civco, D. L.: Topographic normalization of Landsat Thematic Mapper digital imagery, <https://doi.org/10.5555/19910647252>, 1989.
- Cohen, W. B., Yang, Z., and Kennedy, R.: Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync—Tools for calibration and validation, *Remote Sensing of Environment*, 114, 2911-2924, <https://doi.org/10.1016/j.rse.2010.07.010>, 2010.
- Du, Y., Teillet, P. M., and Cihlar, J.: Radiometric normalization of multitemporal high-resolution satellite images with quality control for land cover change detection, *Remote sensing of Environment*, 82, 123-134, [https://doi.org/10.1016/s0034-4257\(02\)00029-9](https://doi.org/10.1016/s0034-4257(02)00029-9), 2002.



- Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., and Luthcke, S.:
 790 The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography, *Science of remote sensing*, 1, 100002, <https://doi.org/10.1016/j.srs.2020.100002>, 2020.
- Estévez, J., Salinero-Delgado, M., Berger, K., Pipia, L., Rivera-Caicedo, J. P., Woher, M., Reyes-Muñoz, P., Tagliabue, G., Boschetti, M., and Verrelst, J.: Gaussian processes retrieval of crop traits in Google Earth Engine based on Sentinel-2 top-of-atmosphere data, *Remote sensing of environment*, 273, 112958, <https://doi.org/10.1016/j.rse.2022.112958>, 2022.
- 795 FAO: Global forest resources assessment 2020: Main report, Food and Agriculture Organization of the United Nations, <https://doi.org/10.4060/ca9825en>, 2020.
- Fensholt, R., Langanke, T., Rasmussen, K., Reenberg, A., Prince, S. D., Tucker, C., Scholes, R. J., Le, Q. B., Bondeau, A., and Eastman, R.: Greenness in semi-arid areas across the globe 1981–2007—an Earth Observing Satellite based analysis of trends and drivers, *Remote sensing of environment*, 121, 144–158, [10.1016/j.rse.2012.01.017](https://doi.org/10.1016/j.rse.2012.01.017), 2012.
- 800 Fiala, A. C., Garman, S. L., and Gray, A. N.: Comparison of five canopy cover estimation techniques in the western Oregon Cascades, *Forest ecology and management*, 232, 188–197, <https://doi.org/10.1016/j.foreco.2006.05.069>, 2006.
- Foody, G. M.: Status of land cover classification accuracy assessment, *Remote sensing of environment*, 80, 185–201, [https://doi.org/10.1016/S0034-4257\(01\)00295-4](https://doi.org/10.1016/S0034-4257(01)00295-4), 2002.
- Guo, B., Fang, M., Yang, L., Guo, T., Ma, C., Hu, X., Guo, Z., Ma, Z., Li, Q., and Wang, Z.: Remapping carbon storage
 805 change in retired farmlands on the Loess Plateau in China from 2000–2021 in high spatiotemporal resolution, *Earth System Science Data*, 18, 429–441, 2026.
- Guyon, I. and Elisseeff, A.: An introduction to variable and feature selection, *Journal of machine learning research*, 3, 1157–1182, 2003.
- Hansen, M., DeFries, R., Townshend, J., Carroll, M., Dimiceli, C., and Sohlberg, R.: Global percent tree cover at a spatial
 810 resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm, *Earth Interactions*, 7, 1–15, [https://doi.org/10.1175/1087-3562\(2003\)007%3C0001:GPTCAA%3E2.0.CO;2](https://doi.org/10.1175/1087-3562(2003)007%3C0001:GPTCAA%3E2.0.CO;2), 2003.
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., and Loveland, T. R.: High-resolution global maps of 21st-century forest cover change, *science*, 342, 850–853, 2013.
- Higgs, E., Falk, D. A., Guerrini, A., Hall, M., Harris, J., Hobbs, R. J., Jackson, S. T., Rhemtulla, J. M., and Throop, W.: The
 815 changing role of history in restoration ecology, *Frontiers in Ecology and the Environment*, 12, 499–506, 2014.
- Hill, J. and Schütt, B.: Mapping Complex Patterns of Erosion and Stability in Dry Mediterranean Ecosystems, *Remote Sensing of Environment*, 74, 557–569, [https://doi.org/10.1016/S0034-4257\(00\)00146-2](https://doi.org/10.1016/S0034-4257(00)00146-2), 2000.
- Hofton, M., Blair, J. B., Story, S., and Yi, D.: Algorithm theoretical basis document (ATBD), 2020.
- Hou, X.: 1:1 million vegetation map of China, National Tibetan Plateau Data Center [dataset], 2019.
- 820 Huang, L., Liu, J., Shao, Q., and Xu, X.: Carbon sequestration by forestation across China: Past, present, and future, *Renewable and Sustainable Energy Reviews*, 16, 1291–1299, <https://doi.org/10.1016/j.rser.2011.10.004>, 2012.



- Hughes, G.: On the mean accuracy of statistical pattern recognizers, *IEEE Transactions on Information Theory*, 14, 55-63, 10.1109/TIT.1968.1054102, 1968.
- Jennings, S., Brown, N., and Sheil, D.: Assessing forest canopies and understorey illumination: canopy closure, canopy cover
 825 and other measures, *Forestry*, 72, 59-74, <https://doi.org/10.1093/forestry/72.1.59>, 1999.
- Kolluru, V., John, R., Chen, J., Henebry, G. M., Xiao, J., Shinde, R., Kussainova, M., and Ganzorig, U.: Leveraging sUAS-Sentinel-2 synergy for cross-scale mapping of canopy cover and aboveground biomass across Mongolia and Kazakhstan, *Remote Sensing of Environment*, 336, 115302, <https://doi.org/10.1016/j.rse.2026.115302>, 2026.
- Korhonen, L., Korhonen, K. T., Rautiainen, M., and Stenberg, P.: Estimation of forest canopy cover: a comparison of field
 830 measurement techniques, <https://doi.org/10.14214/sf.315>, 2006.
- Lefsky, M. A., Cohen, W. B., Parker, G. G., and Harding, D. J.: Lidar remote sensing for ecosystem studies: Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists, *BioScience*, 52, 19-30, [https://doi.org/10.1641/0006-3568\(2002\)052\[0019:LRSFES\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2002)052[0019:LRSFES]2.0.CO;2), 2002.
- 835 Li, B., Liu, X., Zhuang, H., Shi, Q., Zeng, L., Cai, Y., Zhang, H., Cai, Y., Wu, C., and Xu, X.: ALCC: Temporally Consistent Annual Land Cover Maps over China from 1985 to 2022 Based on an Ensemble Change Detection Method, *Journal of Remote Sensing*, 6, 1029, <https://doi.org/10.34133/remotesensing.1029>, 2026.
- Liu, J., Li, S., Ouyang, Z., Tam, C., and Chen, X.: Ecological and socioeconomic effects of China's policies for ecosystem services, *Proceedings of the National academy of Sciences*, 105, 9477-9482, <https://doi.org/10.1073/pnas.0706436105>, 2008.
- 840 Lu, F., Hu, H., Sun, W., Zhu, J., Liu, G., Zhou, W., Zhang, Q., Shi, P., Liu, X., and Wu, X.: Effects of national ecological restoration projects on carbon sequestration in China from 2001 to 2010, *Proceedings of the National Academy of Sciences*, 115, 4039-4044, <https://doi.org/10.1073/pnas.1700294115>, 2018.
- Maxwell, A. E., Warner, T. A., and Fang, F.: Implementation of machine-learning classification in remote sensing: An applied review, *International journal of remote sensing*, 39, 2784-2817, 2018.
- 845 NFGA (National Forestry and Grassland Administration): National Technical Guidelines for Continuous Forest Resource Inventory, Beijing, China, 2014.
- NFGA (National Forestry and Grassland Administration): Forest Resources Statistics of China (1999-2003), Beijing, China, 2005.
- NFGA (National Forestry and Grassland Administration): Forest Resources Statistics of China (2014-2018), Beijing, China,
 850 2019.
- NFGSDC: Forest and grass resources and ecological status in China in 2021, Beijing, 2022.
- Pan, Y., Birdsey, R. A., Phillips, O. L., and Jackson, R. B.: The structure, distribution, and biomass of the world's forests, *Annual review of ecology, evolution, and systematics*, 44, 593-622, <https://doi.org/10.1146/annurev-ecolsys-110512-135914>, 2013.



- 855 Piao, S., Wang, X., Park, T., Chen, C., Lian, X., He, Y., Bjerke, J. W., Chen, A., Ciais, P., Tømmervik, H., Nemani, R. R., and Myneni, R. B.: Characteristics, drivers and feedbacks of global greening, *Nature Reviews Earth & Environment*, 1, 14-27, 10.1038/s43017-019-0001-x, 2020.
- Pretzsch, H.: Forest dynamics, growth, and yield: A review, analysis of the present state, and perspective, *Forest Dynamics, Growth and Yield: From Measurement to Model*, 1-39, 2008.
- 860 Pretzsch, H.: Forest dynamics, growth and yield, Springer, <https://doi.org/10.1007/978-3-540-88307-4>, 2009.
- Pugh, T. A., Lindeskog, M., Smith, B., Poulter, B., Arneeth, A., Haverd, V., and Calle, L.: Role of forest regrowth in global carbon sink dynamics, *Proceedings of the National Academy of Sciences*, 116, 4382-4387, <https://doi.org/10.1073/pnas.1810512116>, 2019.
- Qin, S., Wang, H., Rogers, C., Bermúdez, J., Lourenço, R. B., Zhang, J., Li, X., Chau, J., Tompalski, P., and Gonsamo, A.: Aboveground biomass mapping of Canada with SAR and optical satellite observations aided by active learning, *ISPRS Journal of Photogrammetry and Remote Sensing*, 226, 204-220, <https://doi.org/10.1016/j.isprsjprs.2025.05.022>, 2025.
- 865 Quegan, S., Le Toan, T., Chave, J., Dall, J., Exbrayat, J.-F., Minh, D. H. T., Lomas, M., D'alessandro, M. M., Paillou, P., and Papathanassiou, K.: The European Space Agency BIOMASS mission: Measuring forest above-ground biomass from space, *Remote Sensing of Environment*, 227, 44-60, <https://doi.org/10.1016/j.rse.2019.03.032>, 2019.
- 870 Roy, D. P., Kovalskyy, V., Zhang, H., Vermote, E. F., Yan, L., Kumar, S., and Egorov, A.: Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity, *Remote sensing of Environment*, 185, 57-70, <https://doi.org/10.1016/j.rse.2015.12.024>, 2016a.
- Roy, D. P., Kovalskyy, V., Zhang, H. K., Vermote, E. F., Yan, L., Kumar, S. S., and Egorov, A.: Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity, *Remote sensing of Environment*, 185, 57-70, <https://doi.org/10.1016/j.rse.2015.12.024>, 2016b.
- 875 Roy, D. P., Wulder, M. A., Loveland, T. R., C.E. W., Allen, R. G., Anderson, M. C., Helder, D., Irons, J. R., Johnson, D. M., Kennedy, R., Scambos, T. A., Schaaf, C. B., Schott, J. R., Sheng, Y., Vermote, E. F., Belward, A. S., Bindschadler, R., Cohen, W. B., Gao, F., Hipple, J. D., Hostert, P., Huntington, J., Justice, C. O., Kilic, A., Kovalskyy, V., Lee, Z. P., Lyburner, L., Masek, J. G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R. H., and Zhu, Z.: Landsat-8: Science and product vision for terrestrial global change research, *Remote Sensing of Environment*, 145, 154-172, <https://doi.org/10.1016/j.rse.2014.02.001>, 2014.
- 880 Santoro, M., Cartus, O., Carvalhais, N., Rozendaal, D., Avitabile, V., Araza, A., De Bruin, S., Herold, M., Quegan, S., and Rodríguez-Veiga, P.: The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations, *Earth System Science Data*, 13, 3927-3950, <https://doi.org/10.5194/essd-13-3927-2021>, 2021.
- 885 Schott, J. R., Salvaggio, C., and Volchok, W. J.: Radiometric scene normalization using pseudoinvariant features, *Remote sensing of Environment*, 26, 1-16, [https://doi.org/10.1016/0034-4257\(88\)90116-2](https://doi.org/10.1016/0034-4257(88)90116-2), 1988.
- Settles, B.: Active learning literature survey, <http://digital.library.wisc.edu/1793/60660>, 2009.



- Sexton, J. O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K. M., Channan, S., and DiMiceli, C.: Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error, *International Journal of Digital Earth*, 6, 427-448, 2013.
- 890 Song, X.-P., Hansen, M. C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E. F., and Townshend, J. R.: Global land change from 1982 to 2016, *Nature*, 560, 639-643, <https://doi.org/10.1038/s41586-018-0411-9>, 2018.
- Tang, S., Meng, C. H., Meng, F.-R., and Wang, Y. H.: A growth and self-thinning model for pure even-age stands: theory and applications, *Forest Ecology and Management*, 70, 67-73, [https://doi.org/10.1016/0378-1127\(94\)90075-2](https://doi.org/10.1016/0378-1127(94)90075-2), 1994.
- 895 Teillet, P. M., Guindon, B., and Goodenough, D. G.: On the Slope-Aspect Correction of Multispectral Scanner Data, *Canadian Journal of Remote Sensing*, 8, 84-106, 10.1080/07038992.1982.10855028, 1982.
- Tuia, D., Ratle, F., Pacifici, F., Kanevski, M. F., and Emery, W. J.: Active learning methods for remote sensing image classification, *IEEE Transactions on Geoscience and Remote Sensing*, 47, 2218-2232, <https://doi.org/10.1109/TGRS.2008.2010404>, 2009.
- 900 Tuia, D., Volpi, M., Copa, L., Kanevski, M., and Munoz-Mari, J.: A survey of active learning algorithms for supervised remote sensing image classification, *IEEE Journal of Selected Topics in Signal Processing*, 5, 606-617, <https://doi.org/10.1109/JSTSP.2011.2139193>, 2011.
- Verbesselt, J., Hyndman, R., Newnham, G., and Culvenor, D.: Detecting trend and seasonal changes in satellite image time series, *Remote sensing of Environment*, 114, 106-115, <https://doi.org/10.1016/j.rse.2009.08.014>, 2010.
- 905 Verrelst, J., Berger, K., and Rivera-Caicedo, J. P.: Intelligent sampling for vegetation nitrogen mapping based on hybrid machine learning algorithms, *IEEE Geoscience and Remote Sensing Letters*, 18, 2038-2042, <https://doi.org/10.1109/LGRS.2020.3014676>, 2020.
- Verrelst, J., Dethier, S., Rivera, J. P., Munoz-Mari, J., Camps-Valls, G., and Moreno, J.: Active learning methods for efficient hybrid biophysical variable retrieval, *IEEE Geoscience and Remote Sensing Letters*, 13, 1012-1016, 2016.
- 910 Viña, A., McConnell, W. J., Yang, H., Xu, Z., and Liu, J.: Effects of conservation policy on China's forest recovery, *Science advances*, 2, e1500965, <https://doi.org/10.1109/LGRS.2016.2560799>, 2016.
- Weiskittel, A. R., Hann, D. W., Kershaw Jr, J. A., and Vanclay, J. K.: *Forest growth and yield modeling*, John Wiley & Sons 2011.
- White, J. C., Wulder, M. A., Hobart, G. W., Luther, J. E., Hermosilla, T., Griffiths, P., Coops, N. C., Hall, R. J., Hostert, P., Dyk, A., and Guindon, L.: Pixel-Based Image Compositing for Large-Area Dense Time Series Applications and Science, *Canadian Journal of Remote Sensing*, 40, 192-212, 10.1080/07038992.2014.945827, 2014.
- Woodcock, C. E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S. N., Helder, D., and Helmer, E.: Free access to Landsat imagery, *SCIENCE VOL 320*: 1011, 2008.
- Wulder, M. A., White, J. C., Loveland, T. R., Woodcock, C. E., Belward, A. S., Cohen, W. B., Fosnight, E. A., Shaw, J., Masek, J. G., and Roy, D. P.: The global Landsat archive: Status, consolidation, and direction, *Remote Sensing of Environment*, 185, 271-283, <https://doi.org/10.1016/j.rse.2015.11.032>, 2016.
- 920



- Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., Allen, R. G., Anderson, M. C., Belward, A. S., Cohen, W. B., Dwyer, J., Erb, A., Gao, F., Griffiths, P., Helder, D., Hermosilla, T., Hipple, J. D., Hostert, P., Hughes, M. J., Huntington, J., Johnson, D. M., Kennedy, R., Kilic, A., Li, Z., Lymburner, L., McCorkel, J., Pahlevan, N., Scambos, T. A., Schaaf, C., Schott, J. R., Sheng, Y., Storey, J., Vermote, E., Vogelmann, J., White, J. C., Wynne, R. H., and Zhu, Z.: Current status of Landsat program, science, and applications, *Remote Sensing of Environment*, 225, 127-147, <https://doi.org/10.1016/j.rse.2019.02.015>, 2019.
- 925 Yu, J., Yu, Z., Han, W., Zhang, F., and Liu, S.: ChinaTCC30: An annual 30 m tree canopy cover dataset for China from the 1970s to 2024 (Part 3: 1995-2000), Zenodo [dataset], <https://doi.org/10.5281/zenodo.21199497>, 2026a.
- 930 Yu, J., Yu, Z., Han, W., Zhang, F., and Liu, S.: ChinaTCC30: An annual 30 m tree canopy cover dataset for China from the 1970s to 2024 (Part 1: 1975, 1980, and 1985-1988), Zenodo [dataset], <https://doi.org/10.5281/zenodo.21194015>, 2026b.
- Yu, J., Yu, Z., Han, W., Zhang, F., and Liu, S.: ChinaTCC30: An annual 30 m tree canopy cover dataset for China from the 1970s to 2024 (Part 2: 1989-1994), Zenodo [dataset], <https://doi.org/10.5281/zenodo.21195911>, 2026c.
- 935 Yu, J., Yu, Z., Han, W., Zhang, F., and Liu, S.: ChinaTCC30: An annual 30 m tree canopy cover dataset for China from the 1970s to 2024 (Part 7: 2019-2024), Zenodo [dataset], <https://doi.org/10.5281/zenodo.21265832>, 2026d.
- Yu, J., Yu, Z., Han, W., Zhang, F., and Liu, S.: ChinaTCC30: An annual 30 m tree canopy cover dataset for China from the 1970s to 2024 (Part 6: 2013-2018), Zenodo [dataset], <https://doi.org/10.5281/zenodo.21263677>, 2026e.
- Yu, J., Yu, Z., Han, W., Zhang, F., and Liu, S.: ChinaTCC30: An annual 30 m tree canopy cover dataset for China from the 1970s to 2024 (Part 5: 2007-2012), Zenodo [dataset], <https://doi.org/10.5281/zenodo.21262725>, 2026f.
- 940 Yu, J., Yu, Z., Han, W., Zhang, F., and Liu, S.: ChinaTCC30: An annual 30 m tree canopy cover dataset for China from the 1970s to 2024 (Part 4: 2001-2006), Zenodo [dataset], <https://doi.org/10.5281/zenodo.21245669>, 2026g.
- Yu, J., Lai, H., Xu, L., Luo, S., Zhou, W., Song, H., Xi, L., and Shu, Q.: Estimation of forest canopy cover by combining ICESat-2/ATLAS data and geostatistical method/co-kriging, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 1824-1838, <https://doi.org/10.1109/JSTARS.2023.3340429>, 2023.
- 945 Zeng, W., Chen, X., and Yang, X.: Estimating changes of forest carbon storage in China for 70 years (1949–2018), *Scientific Reports*, 13, 16864, <https://doi.org/10.1038/s41598-023-44097-4>, 2023.
- Zhang, C., Su, Y., Liu, L., Wu, J., Huang, G., Li, X., Bi, C., Yan, W., and Laforteza, R.: Seasonal and long-term dynamics in forest microclimate effects: global pattern and mechanism, *npj climate and atmospheric science*, 6, 116, <https://doi.org/10.1038/s41612-023-00442-y>, 2023.
- 950 Zhu, Z. and Woodcock, C. E.: Continuous change detection and classification of land cover using all available Landsat data, *Remote sensing of Environment*, 144, 152-171, <https://doi.org/10.1016/j.rse.2014.01.011>, 2014.
- Zhu, Z., Wang, S., and Woodcock, C. E.: Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images, *Remote sensing of Environment*, 159, 269-277, <https://doi.org/10.1016/j.rse.2014.12.014>, 2015.



- 955 Zhu, Z., Piao, S., Myneni, R. B., Huang, M., Zeng, Z., Canadell, J. G., Ciais, P., Sitch, S., Friedlingstein, P., and Arneeth, A.:
Greening of the Earth and its drivers, *Nature climate change*, 6, 791-795, 2016.
- Ziegler, A. D., Fox, J. M., and Xu, J.: The rubber juggernaut, *Science*, 324, 1024-1025,
<https://doi.org/10.1126/science.1173833>, 2009.
- 960 Ziter, C. D., Pedersen, E. J., Kucharik, C. J., and Turner, M. G.: Scale-dependent interactions between tree canopy cover and
impervious surfaces reduce daytime urban heat during summer, *Proceedings of the National Academy of Sciences*, 116,
7575-7580, <https://doi.org/10.1073/pnas.1817561116>, 2019.