https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

Open Access Earth System Science Data Discussions

2020 forest age map for China with 30 m resolution

Kai Cheng[1], Yuling Chen[1], Tianyu Xiang[2], Haitao Yang[1], Weiyan Liu[3], Yu Ren[1,4], Hongcan Guan[5], Tianyu Hu[6], Qin Ma[7], Qinghua Guo[1,4*]

[1]Institute of Remote Sensing and Geographic Information System, School of Earth and Space Sciences, Peking University, Beijing 100871, China
[2] College of Earth Sciences, Chengdu University of Technology, Chengdu 610059, China
[3] State Forestry and Grassland Administration Key Laboratory of Forest Resources & Environmental Management, Beijing Forestry University, Beijing 100083, China
[4] Institute of Ecology, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China
[5] School of Tropical Agriculture and Forestry, Hainan University, Haikou 570100, China
[6] State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
[7] School of Geography, Nanjing Normal University, Najing 210023, China

Correspondence to: Qinghua Guo (guo.qinghua@pku.edu.cn)

Abstract. A spatially explicit, high-resolution forest age map is critical for quantifying forest carbon stock and carbon sequestration potential. Previous endeavours to estimate forest age in China at national scale mainly concentrated on a sparse resolution or incomplete forest ecosystems because of complex species composition, vast forest areas, insufficient field measurements, and the lack of effective methods. To overcome these limitations, we construct a framework for estimating China's forest age by combining remote-sensing time series analysis with machine learning algorithms based on massive field measurements and remote-sensing dataset. Specifically, the LandTrendr time series analysis is first applied to detect forest disturbances from 1985 to 2020, with the time since the last disturbance serving as a proxy for forest age. Next, for pixels where no disturbance, machine learning algorithms are used to estimate forest age from independent variables, including forest height, climate, terrain, soil, and forest-age field measurements. Finally, MLA models are established for each vegetation division and used to estimate forest ages. Combining these two methods produces a spatially explicit 30 m resolution forest-age map for China in the year of 2020. Validation against independent field plots produces a R^2 from 0.51 to 0.63. Nationally, the average forest age is 56.1 years (standard deviation = 32.7 years), where the Qinghai-Tibet Plateau alpine vegetation zone has the oldest forest with an average of 138.0 years, whereas the forest in the warm temperate deciduous-broadleaf forest vegetation zone averages only 28.5 years. This 30-m-resolution forest-age map provides vital information for accurately understanding the ecological benefits of China's forests and to sustainably manage China's forest resources.

1 Introduction

Forest age provides critical information about forest ecosystem succession and condition, making it essential for understanding the ecological benefits of forests (Lin et al., 2023). China's forests have been significantly disrupted in the past few decades

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

Open Access Earth System Science Data Discussions

due to natural disasters and human activities (Niu et al., 2023), resulting in substantial changes to the forest-age structure. This situation makes it significantly challenging to accurately understand forest ecosystem carbon storage (Pan et al., 2011; Tong et al., 2020). Because of complex species composition, vast forest areas, insufficient field measurements, and the lack of effective methods, existing estimates of the China's forest age on the national scale have concentrated on sparse resolution (Zhang et al., 2017) or incomplete forest ecosystem (Xiao et al., 2023). This brought considerable uncertainties in assessing the carbon sources and sinks in China's forest ecosystem (Piao et al., 2022; Wang et al., 2022). Therefore, there is an urgent need to map time-efficient forest age with high spatial resolution across China.

Currently, China's forest age is acquired mainly through the national forest inventory, which is highly accurate (Xiao et al., 2023) but requires extensive labour and material resources and is time-consuming and costly (Liu et al., 2022). Additionally, most of China's forests are in steep mountainous areas that are difficult to access (Cheng et al., 2023a), which limits the survey range and uneven distribution of field samples, making it difficult to estimate the age of China's forests on a national scale. Thus, using the traditional method of forest inventory is difficult to capture the complete age distribution and spatial characteristics of China's forests in a timely and accurate manner.

Remote sensing technology has been proved effective in estimating forest cover (Su et al., 2020; Tubiello et al., 2023) and forest structure (Yu et al., 2020a; Maltman et al., 2023a) at multiple scales. The opening and sharing of Landsat time series and the development of Google Earth Engine (GEE) cloud-processing platform has facilitated the application of remote sensing to estimate forest age. Several studies have focused on mapping China's forest age; for example, Xiao et al. (2023) mapped the age of China's young forests at 30 m resolution by using continuous change detection and classification (CCDC) method. Yu et al. (2020a) produced a 1-km resolution map of the age for planted forests in China by using the age-height equations. Zhang et al. (2017) proposed a top-down method to generate a 1km stand age map using climate and forest height data. Zhang et al. (2014) mapped a national forest age map with 1 km resolution by using remote-sensing forest height and forest type data. However, the existing China's forest age map at on a national scale has been typically undertaken at coarser spatial resolution (e.g., 1 km), with finer resolution (e.g., 30 m) reserved for young forests. There is still lacking forest age spatial dataset with high resolution covering all of China's forests region.

The use of remote sensing to map forest age primarily encompasses two methodological categories. The first is statistical parametric regression approaches, which estimates forest age through establishing a coherent relationship between remote sensing features and field-collected empirical samples to deduce forest age (Maltamo et al., 2020; Schumacher et al., 2020a). Growth models are one of the most popular parametric models of forest age estimation (Zhang et al., 2014; Zhang et al., 2017; Yu et al., 2020b). However, this type of model is based on tree species, which makes it hard to derive forest age when lacking species information, especially over large scales. The second methodological category is nonparametric machine learning algorithms (MLAs), which are more flexible and can handle complex problems (Alerskans et al., 2022). Currently, the application of MLAs to estimate national forest age has not been widely explored. Some previous studies used a single MLA, such as Random Forest (RF) (Besnard et al., 2021b), to estimate the age of the forest. However, the wide distribution of forests, complex forest types, and diverse terrain and climate conditions in China make applying a single model to determine the

make it difficult to apply a single model to determine the national age of forests

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

national age forests difficult. Therefore, it is necessary to explore the applicability of MLAs in the estimation of forest age in different regions of China.

The objective of the present study is to generate the first China's forest age dataset at 30 m resolution using multi-source datasets based on remote sensing time series analysis and MLAs. Specifically, first, we apply the LandTrendr change detection algorithm to monitor forests disturbed between 1985 and 2020 to estimate their ages. Subsequently, utilizing mainstream machine learning algorithms, we engage in a zone-based exploration to determine the optimal models for forest age estimation in undisturbed areas to estimate forest age. Finally, the resulting forest age map is validated by using forest field measurements and existing remote sensing products. The generated 30-m-resolution forest age map provides critical information to quantify forest carbon storage and to sustainably manage China's forests.

2 Materials and methods

2.1 Dataset and pre-processing

2.1.1 Forest inventory data

The data from China's seventh national forest inventory survey from 2004 to 2008 (http://www.forestry.gov.cn/) were collected to develop models to estimate forest age. The inventory involves systematically and accurately monitoring the national forest resources based on 667m^2sample plots covering the whole country (Ren et al., 2011). The main information collected from the sample plots are tree species, stand age, average tree height, and geographic location. The stand age is determined based on the planting time or is estimated using tree diameter at breast height (Zhang et al., 2017). We totally collected 58,033 field plots ranging in age from 1 to 480 years (Figures 1b and 1c). The mean age of the samples is 34.0 years, with a standard deviation of 29.6 years. The sample plots were distributed across eight vegetation divisions (Figure 1b), each containing at least 436 sample plots for building MLA models to estimate forest age (Figure 1d).

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

Open Access Earth System Science Data Discussions

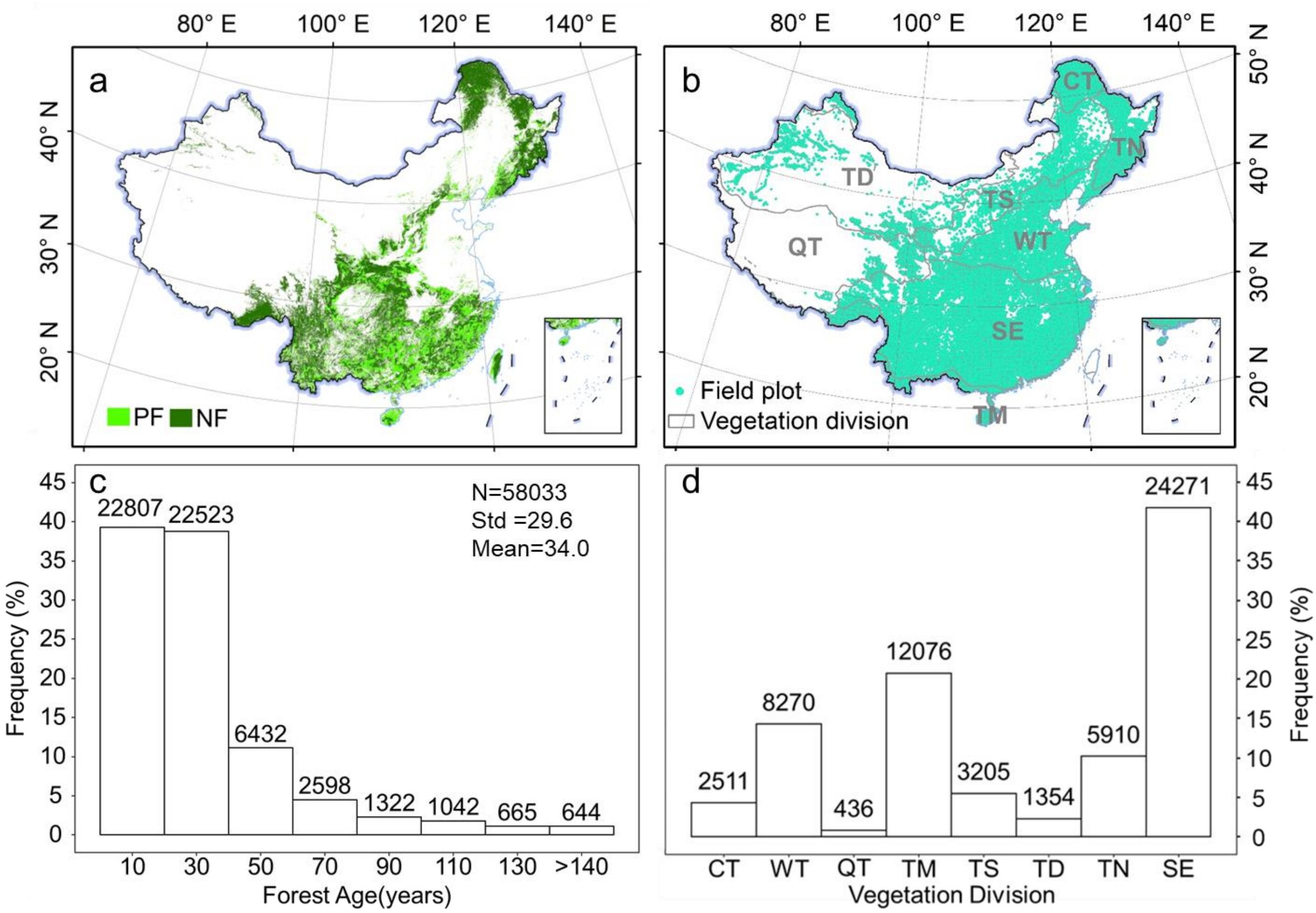


Figure 1. Forest mask and field sample distribution. (a) Planted forest and natural forest mask generated by Cheng et al. (2023). (b) Distribution of field samples over eight vegetation divisions. (c) Frequency distribution of field sample ages. (d) Frequency distribution of field samples for eight vegetation divisions. PF: planted forest, NF: natural forest, CT: Cold Temperate needleleaf forest, WT: Warm Temperate deciduous-broadleaf forest, QT: Qinghai-Tibet Plateau alpine vegetation, TM: Tropical Monsoon forest-rainforest, TS: Temperate Steppe, TD: Temperate Desert, TN: Temperate Needleleaf-broadleaf mixed forest, SE: Subtropical Evergreen broadleaf forest. N: the number of plots, Std: standard deviation, Mean: mean age.

2.1.2 Landsat time-series data

From the GEE platform, we collected Landsat TM, ETM+, OLI Tier 1 surface reflectance images dating from 1985 to 2020 to estimate forest age for disturbed forest regions. All data were atmospherically corrected and processed by the Land Surface Reflectance Code and the Landsat Ecosystem Disturbance Adaptive Processing System algorithms. We removed the clouds or cloud shadows using the C function of the mask algorithm (Du et al., 2023), then we created composited images using a median compositing method for forest regions. Finally, we calculated the normalized burn ratio (NBR) to detect forest disturbance. NBR has been proved effective in numerous studies detecting forest disturbance (Du et al., 2023; Tian et al., 2023). It is calculated as follows by using the near-infrared (NIR) and short-wave infrared (SWIR) bands:

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

$$NBR = \frac{NIR - SWIR}{NIR + SWIR}. \quad \text{(Eq. 1)}$$

2.1.3 Forest mask

This study uses the 2020 dataset of planted and natural forests at 30 m resolution in China (Figure 1a) as a mask for forest age mapping. This dataset is produced by integrating multisource remote-sensing data and a large number of crowdsourced samples, with an overall accuracy of over 80% (Cheng et al., 2023a). In this study, we employ this dataset as a forest mask and utilize a combination of time series change detection algorithms and MLAs to trace the age of these planted and natural forests.

2.1.4 Forest height data

The canopy height data for China was downloaded from the website (https://3decology.org/), which was generated based on deep learning by integrating Global Ecosystem Dynamics Investigation and Ice, Cloud and land Elevation Satellite -2 data. This dataset has a spatial resolution of 30 m and corresponds to 2019. The accuracy of this national forest canopy height data was assessed by comparing three independent validation datasets, indicating high accuracy for the canopy height product by neural network guided interpolation ($R^2 \geq 0.55$, RMSE ≤ 5.5 m) (Liu et al., 2022). Notably, the forest extent used in this dataset is consistent with the forest extent mentioned earlier for planted and natural forests, ensuring spatial consistency when estimating forest age.

2.1.5 Climate data

Climate data were acquired from Worldclim 2.1 (https://worldclim.org/), which offers 19 bioclimatic variables, including temperature and precipitation, with 30 arc-second resolutions. The 19 bioclimatic variables include annual trends, seasonality, and extreme environmental factors in temperature and precipitation. We resampled the 19 GeoTiff (.tif) files to 30 m resolution using a nearest-resampling method for spatial resolution consistency. To reduce the dimension of bioclimatic variables, we applied a principal component analysis to map the 19 bioclimatic variables into a new principal component (PC) space. We use the first three components PC1, PC2, PC3 to represent the climate factors. According to the results of the analysis, PC1 gives annual trends in temperature and precipitation, PC2 gives seasonal variations in temperature and precipitation, and PC3 gives precipitation and temperature extremes (Supplementary Table 1).

2.1.6 Soil data

Soil data were extracted from the harmonized world soil database, V1.2, developed jointly by the Food and Agriculture Organization of the United Nations, the International Institute for Applied Systems, the ISRIC-World Soil Information, the Institute of Soil Science, Chinese Academy of Sciences, and the Joint Research Centre of the European Commission with a resolution of 30 arc-seconds. As per previous studies, soil type and texture were selected from the soil dataset in this study to construct the model to estimate forest age (Besnard et al., 2021a). We also resampled the soil data to 30 m using a nearest-resampling method.

2.1.7 Topographic data

The Shuttle Radar Topography Mission (SRTM) V3 provides global digital elevation data at 30 m resolution and was used in this study to extract topographic variables (Su et al., 2020). Three topographic features, elevation, slope, and aspect, were calculated to estimate forest ages.

Table 1. Descriptions of variables used to estimate the forest age of China.

Data type	Data source	Resolution	Time	Variables
Remote sensing images	Landsat TM/ETM+/OLI	30m	1985–2020	NBR
Forest mask	Planted and natural forest map (Cheng et al., 2023a)	30m	2020	Planted and natural forest
Forest canopy height data	NNGI-Forest Canopy Height	30m	2019	Forest height
Climate data	WorldClim version 2.1 (Fick and Hijmans 2017)	1 km	1970–2000	PC1, PC2, PC3
Soil data	Harmonized World Soil Database V1.2(https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/harmonized-world-soil-database-v12/en/)	30 arc-second	1971–1981	Soil type, soil texture
Topographic data	SRTM DEM	30 m	2000	Elevation, slope, and aspect

2.2 Forest age estimation

To generate the forest age map for China and explore the performance of MLAs to retrieve forest age, we applied two approaches to estimate forest age in China: the LandTrendr detection approach and the MLA-based approach. First, the LandTrendr was applied to detect stand-replacing disturbances based on the Landsat time series images. Second, the MLA-based method estimated ages for undisturbed forest regions within Landsat time series data. Here, we assumed that undisturbed forests over the Landsat record (before 1985) have the similar structural and spectral features to natural forests. Figure 2 shows a detailed flowchart describing the framework for forest age estimation proposed in this study.

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

Open Access Earth System Science Data Discussions

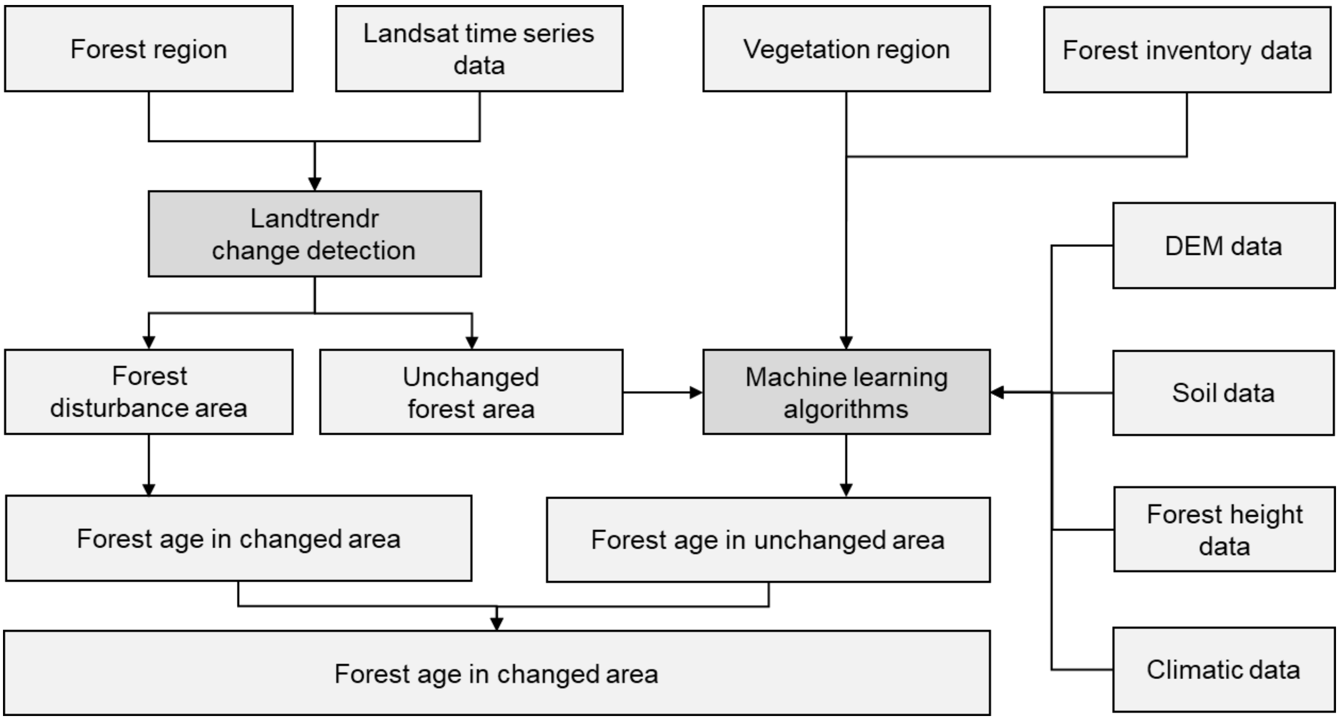


Figure 2. Framework of China's forest age estimation.

2.2.1 LandTrendr detection approach

LandTrendr was designed to detect and analyse changes in surface features, particularly disturbances and recovery processes, and is commonly applied to multispectral remote sensing imagery from the Landsat satellite series to capture long-term forest disturbances (Du et al., 2022). Using LandTrendr to detect forest age involves the following steps:

(1) Time series data transformation. LandTrendr transforms multiple temporal remote-sensing image datasets into a series of indices, such as the NBR.

(2) Breakpoint detection. Using the generated time series indices, LandTrendr retraces from the state in 2020 in search of breakpoints in the time series. These breakpoints signify transition points in the time series, which indicate instances of surface disturbance or recovery.

(3) Age estimation. By pinpointing breakpoints, the time of occurrence for each breakpoint is established. Forest age estimates for the current location are accomplished by subtracting the breakpoint time from the latest time.

LandTrendr was implemented on the GEE platform by using the function of *runLT()* provided by the LT_GEE API (Kennedy et al., 2018). Table 1 lists the main input parameters.

Table 1. Parameters of LandTrendr used in this study.

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

Parameters	Definition	Value
maxSegments	Maximum number of segments to be fitted on the time series	10
spikeThreshold	Threshold for dampening the spikes (1.0 means no dampening)	0.9
vertexCountOvershoot	The initial model can overshoot the maxSegments + 1 vertices by this amount. Later, it will be pruned down to maxSegments + 1	3
preventOneYearRecovery	Prevent segments that represent one-year recoveries	False
recoveryThreshold	If a segment has a recovery rate faster than 1/recovery threshold (in years), then the segment is disallowed	0.25
pvalThreshold	If the *p*-value of the fitted model exceeds this threshold, then the current model is discarded and another one is fit by using the Levenberg–Marquardt optimizer	0.05
bestModelProportion	Takes the model with most vertices that has a *p*-value that is at most this fraction away from the model with the lowest p-value	0.75
minObservationsNeeded	Minimum observations required to perform output fitting	6

2.2.2 Machine learning approach

(1) MLA selection

This study used the following model-screening procedure to explore which model works best for each vegetation division. First, we used the automated machine learning (Auto-ML) open-source Python library LazyPredict to filter for alternative models. LazyRegressor (including 40 MLAs) was used to build stand-age estimation models based on all data, which helps to understand which MLA works well without tuning parameters. The performing models with R^2 greater than 0.60 in each vegetation division were concentrated in thirteen MLAs (Supplementary Table 2). Second, by splitting training data and testing data, the top three MLAs for each vegetation division were determined (Supplementary Table 2). It can be found that the potential optimal models of eight vegetation divisions is concentrated in RF, Gradient Boosting Decision Tree (GBDT), Histogram Gradient Boosting (HistGradientBoost), Light Gradient Boosting Machine (LightGBM), and Categorical Boosting (CatBoost).

RF is an ensemble learning method that combines multiple decision trees (Breiman 2001; Dutta et al., 2020). It leverages the wisdom of crowds to make accurate predictions. RF mitigates overfitting and provides robust results by training each tree on a random subset of the data and features (Lavanya et al., 2017; Guo et al., 2019). GBDT is an ensemble technique that builds

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

a strong predictive model by sequentially training decision trees (Jerome 2001). Each tree corrects the errors of its predecessor (Wei et al., 2019), resulting in a highly accurate and robust model. HistGradientBoost is a variant of GBDT that employs histogram-based techniques. It efficiently approximates data distributions and reduces memory consumption during training. This algorithm is particularly beneficial when dealing with large datasets and complex features (Tesfagergish et al., 2022). LightGBM is a gradient-boosting framework that prioritizes speed and efficiency. It employs a histogram-based approach and parallel computing, making it suitable for large datasets. CatBoost is a new modification gradient boosting algorithm that is designed specifically for handling categorical features. It automatically encodes categorical variables, simplifying the data pre-processing stage. CatBoost is known for its robustness and efficiency, can achieve high accuracy on a small-scale dataset.

We implemented RF, GBDT, and HistGradientBoost by using the Scikit-learn package of Python 3.9.11, while the LightGBM and CatBoost algorithms were constructed by using the lightgbm and catboost packages of Python 3.9.11.

(2) Hyperparameter tuning

Hyperparameter tuning of MLAs is critical in the ML model training process because it significantly enhances the model's performance, generalization capability, and adaptability (Sandha et al., 2020). Bayesian optimization has been selected for hyperparameter tuning due to its complicated derivative evaluation, and nonconvex-function-related features (Mekruksavanich et al., 2022). It is implemented by using Optuna, an open source hyperparameter optimization framework to automate hyperparameter searches (Akiba et al., 2019). The hyperparameters and their searching range in MLAs are listed in Supplementary Table 2.

(3) Model interpretation

Furthermore, we used Shapley Additive explanations (SHAP) values (Lundberg and Lee, 2017; Lundberg et al., 2019), a model-agnostic technique for interpreting ML models, to explore functional correlations between the variables and forest age (Besnard et al. 2021). SHAP derives the Shapely additive contribution values from coalitional game theory (Kim et al. 2023). By examining the contribution of each input variable to the model's output, SHAP can identify the primary drivers of the model's predictions and provide insights into the underlying causes that influence forest age (Sun et al. 2023). The higher the SHAP value, the larger the contribution of the variable. Here we calculated SHAP value using *shap* package in Python.

2.3 Accuracy assessment

We collected three independent data samples to validate the generated forest age map. The first is the forest inventory data independent of training data. The second source involved validation samples obtained from the literature. To ensure the samples collected were representative, we excluded samples with collection dates prior to 2010. The third source was the existing forest age products. As validation metrics, we used the coefficient of determination (R2), the root mean square error (RMSE), the mean absolute error (MAE), and the mean error (ME). These are given mathematically as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2}, \qquad (2)$$

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

Open Access Earth System Science Data Discussions

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \quad (3)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|, \quad (5)$$

$$ME = \frac{\sum_{i=0}^{n}(y_i - \hat{y}_i)}{n}, \quad (6)$$

where y_i is the observed value for the ith analytic tree, $\hat{y}_i$ is the predicted value of the ith observed value, n is the number of trees, and $\overline{y_i}$ is the mean of the observed value.

3 Results

3.1 MLA performance for China's forest age estimation

Through a rigorous hyperparameter-optimization process and independent validation, four distinct MLAs (RF, GBDT, LightGBM, and CatBoost) were selected across eight different vegetation divisions (Table 2). GBDT performed exceptionally well for estimating the forest age of cold temperate needleleaf forest (CT) vegetation zone, producing R^2 of 0.47 and RMSE of 4.95 years (MAE=17.99, ME=-1.86). RF excelled at estimating the forest age of warm temperate deciduous-broadleaf forest (TD) vegetation zone, producing an independent validation R^2 of 0.61 and RMSE of 3.47 years (MAE=9.13, ME=-0.01). CatBoost consistently demonstrated strong performance for the Qinghai-Tibet Plateau alpine vegetation (QT), tropical monsoon forest-rainforest (TM), temperate steppe (TS), temperate desert (TD), and subtropical evergreen broadleaf forest (SE) zones, with R^2 values ranging from 0.57 to 0.85 and RMSE values from 2.04 to 7.65 years. LGBMRegressor was the preferred choice in the temperate needleleaf-broadleaf mixed forest (TN) vegetation division, yielding an R^2 of 0.63 and an RMSE of 4.14 years.

WT ?

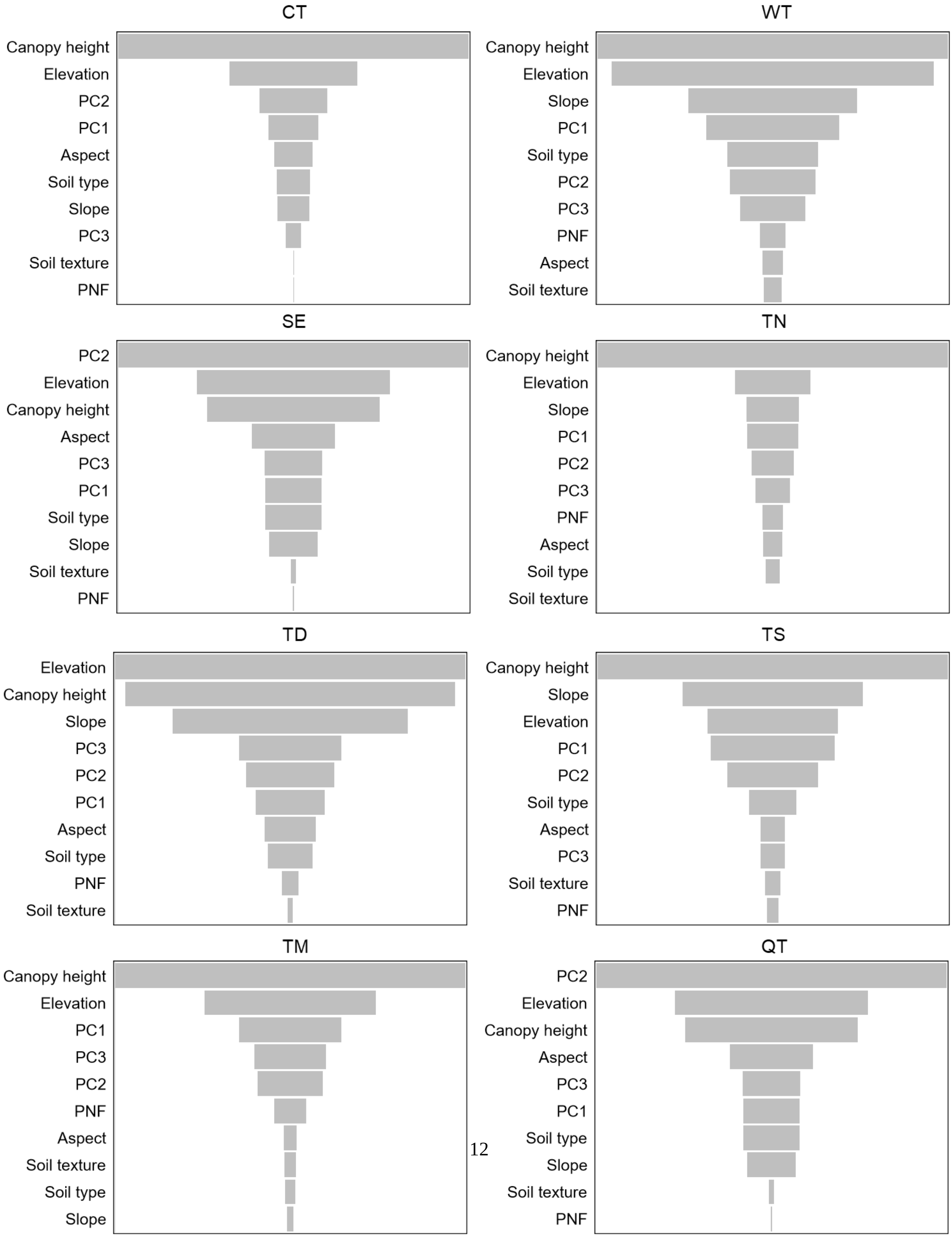
Table 2. MLA for eight vegetation divisions and their validation metrics.

Vegetation division	Algorithm	R^2	RMSE	MAE	ME
CT	GradientBoost	0.46	4.95	17.99	-1.86
WT	RF	0.61	3.47	9.13	-0.01
QT	CatBoost	0.57	7.65	42.58	10.43
TM	CatBoost	0.85	2.04	1.34	-0.08
TS	CatBoost	0.78	4.16	11.85	-0.87
TD	CatBoost	0.80	5.33	21.02	1.84
TN	LGBM	0.63	4.14	12.78	0.36
SE	CatBoost	0.70	3.49	7.97	0.00

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

We further analysed the factors influencing the forest age estimation in each vegetation division, and the findings are illustrated in Figure 3. While the prioritization of factors affecting forest age estimation varies across different vegetation divisions, canopy height is unquestionably the predominant factor influencing this estimation. Its absolute value is the highest of the CT, WT, TN, TS, and TM vegetation zones (Figure 3). Moreover, it is among the top three most influential factors in all the remaining vegetation zones. Subsequently, topographical conditions assume prominence, with elevation consistently featuring among the top three factors in the SHAP value across all vegetation divisions. In the TD vegetation division, elevation becomes the most influential factor. Climate factors earn third-tier consideration, particularly in the SE vegetation zone, where the impact of PC2 of the climate factors surpasses that of both canopy height and topographical conditions. In the other vegetation divisions, the influence of climate factors generally falls to the mid-range. In contrast, across all eight vegetation divisions, factors related to soil, such as soil type and soil texture, do not exert a pronounced influence on forest age estimation.

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023



233

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

Figure 3. Order of shape values of factors affecting the estimation of forest age in different vegetation zones

3.2 China's forest age map

China's 30 m resolution forest-age map is presented in Figure 4. The mean of the estimated forest age is 56.11 years with a standard deviation of 32.67 years. Geographically, forests in northeast and southwest China are relatively older than those in other regions (Figure 4). At the provincial scale, the average forest age ranges from 3.9 to 116.8 years (Figure 5a, Supplementary Table 5), whereas Qinghai province has the highest mean forest age, and Hong Kong has the lowest mean forest age. Forest ages in Sichuan province are more varied than in other provinces (Figure 5a). On the regional scale, the QT vegetation zones have the oldest forests with an average of 138.0 years, followed by CT (107.6 years), TS (107.0 years), TN (68.3 years), TD (60.3 years), TM (53.0 years), and SE (49.2 years) (Figure 5b, Supplementary Table 6). The WT vegetation zones have the youngest forests (28.5 years).

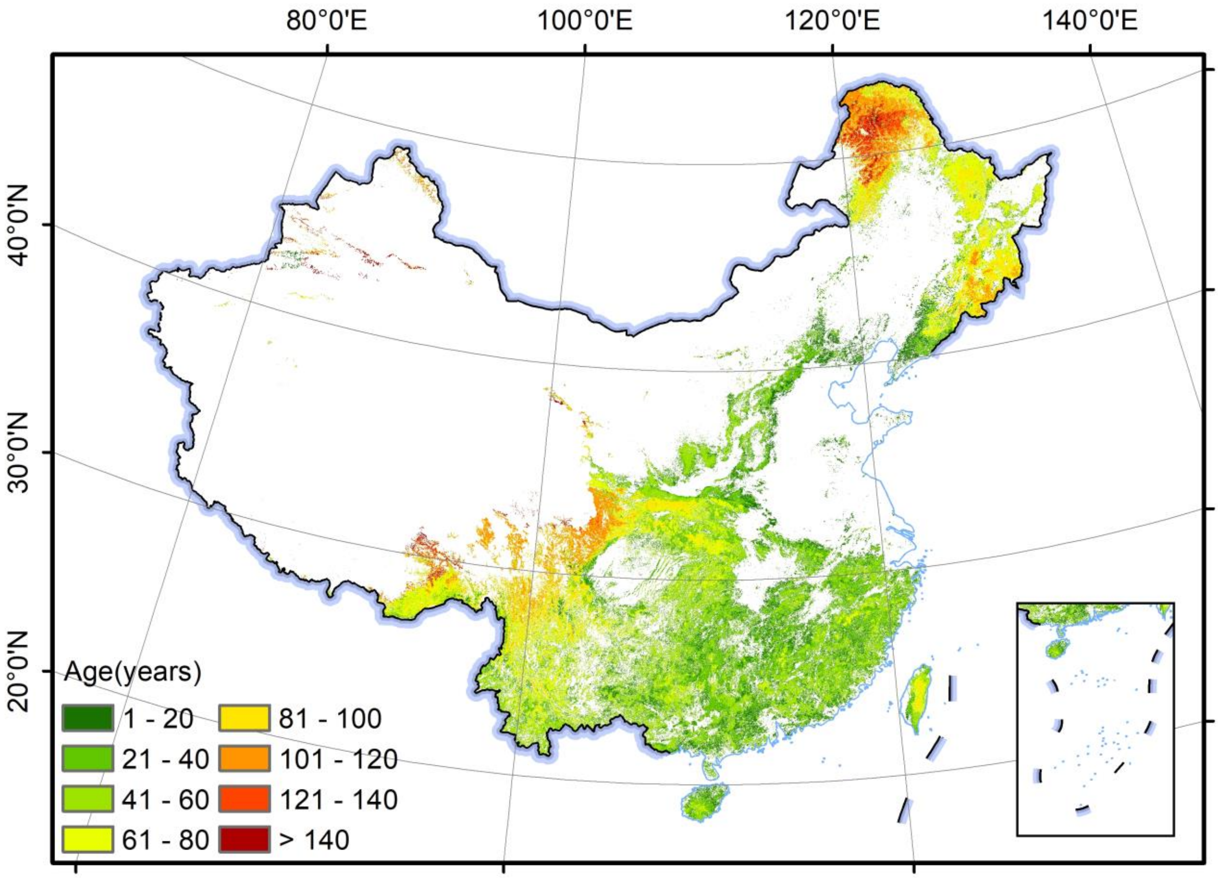


Figure 4. Distribution of China's forest age with 30 m resolution.

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

Open Access Earth System Science Data Discussions

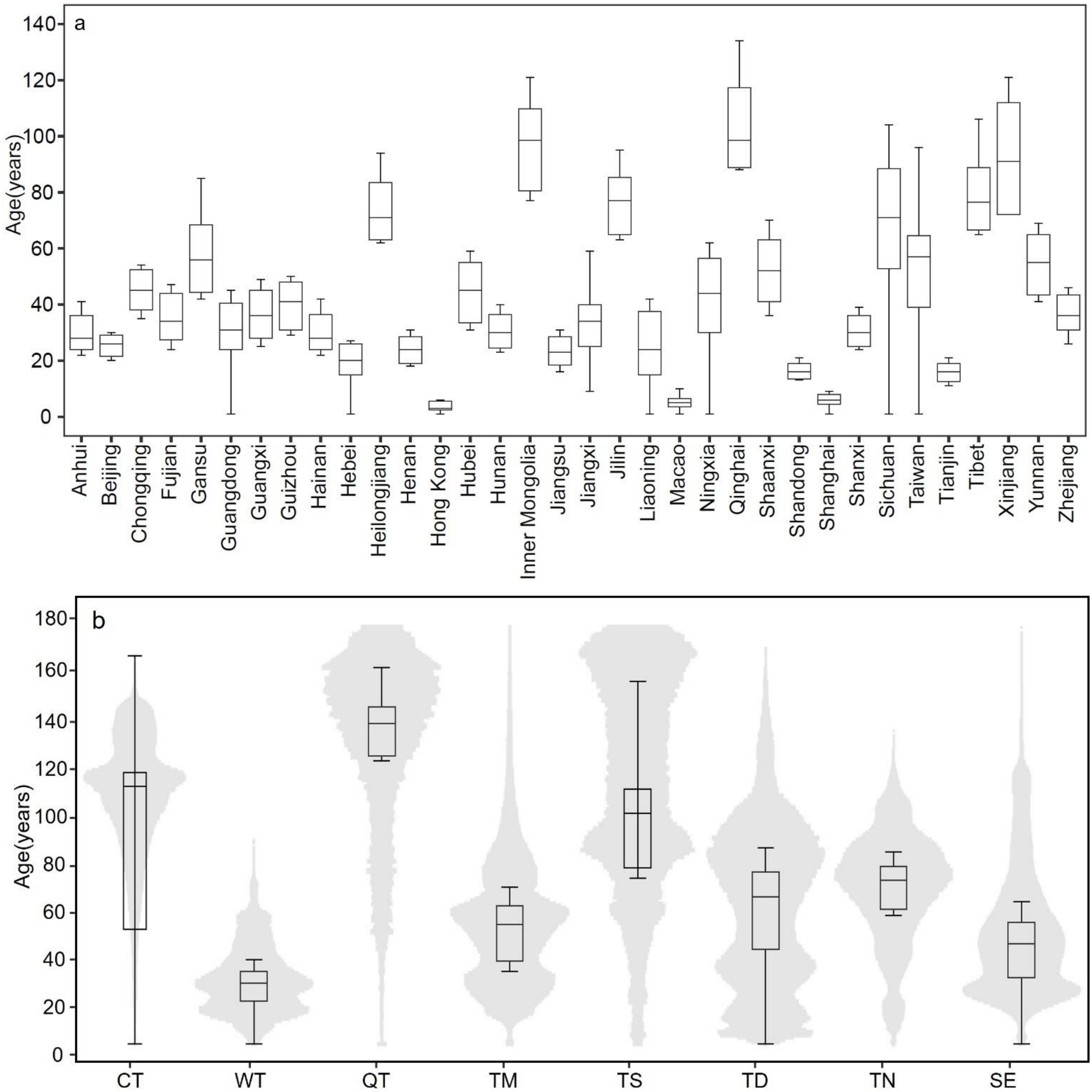


Figure 5. Mean forest age on the (a) provincial and (b) regional scales.

error bars mean?

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

3.3 Evaluation

2004?

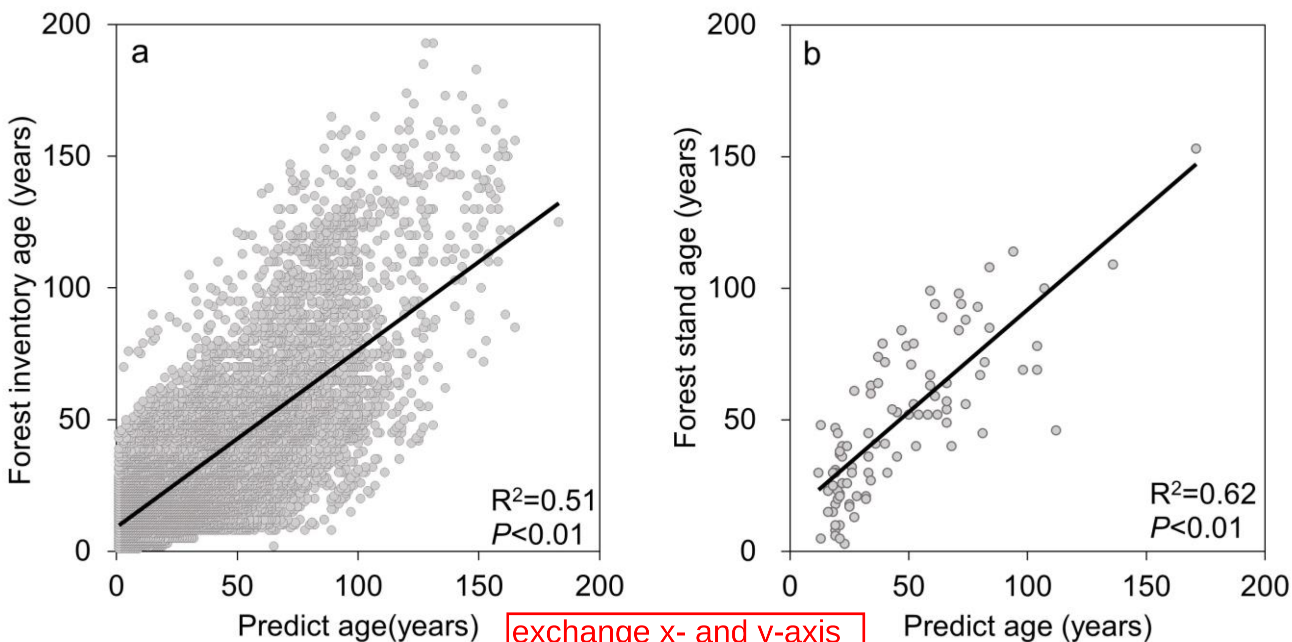
We used three independent data sources to evaluate the final forest age map, including forest inventory samples from 2003 to 2008, field measurements collected from published papers, and existing remote sensing based forest age products.

3.3.1 Comparison with forest inventory samples

We initially validated the forest age estimations by using forest inventory data. The forest inventory samples were acquired from 2003 to 2008. To align with the time frame of the forest age data obtained in this study, we shifted the predicted values corresponding to each sample forward by ~16 years. This strategy allows us to compare them with the inventory-measured forest ages. Figure 6a shows the comparison, which suggests that they have a significant linear relationship with R^2 = 0.51 (Figure 6a).

3.3.2 Comparison with field measurements

We collected 99 field measurements of mean forest stand age after 2010 from published papers (Supplementary Table 7) and compared them with our estimated results. Figure 6b shows that the predicted forest ages also present a significant linear relationship with field measurements, with R^2 = 0.62.



exchange x- and y-axis

Figure 6. Scatter plots of (a) forest inventory age vs predicted forest age for this study and (b) field measurements of forest stand age collected from published papers vs predicted forest.

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

3.3.3 Comparison with existing forest age products

We compared our estimated forest age map with an independent global forest age data product produced by Besnard et al. (2021a). Figure 7 shows the difference between these two maps, which suggests an average difference of 9.7 years. Our mapped forest age is older in northeast regions but younger forests in the middle regions than that from Besnard et al. (2021a) dataset. In addition, we gathered the existing remote sensing based forest age maps over China from published datasets and compared their average forest age with our estimation (Table 3). According to the available data, the average forest age in China ranged from 40 to 43 years between 2000 and 2013, corresponding to approximately 50 to 53 years in 2020. This aligns closely with the average forest age of 56.1 years obtained in this study for the year of 2020, further underscoring the reliability of the forest age mapped in this study.

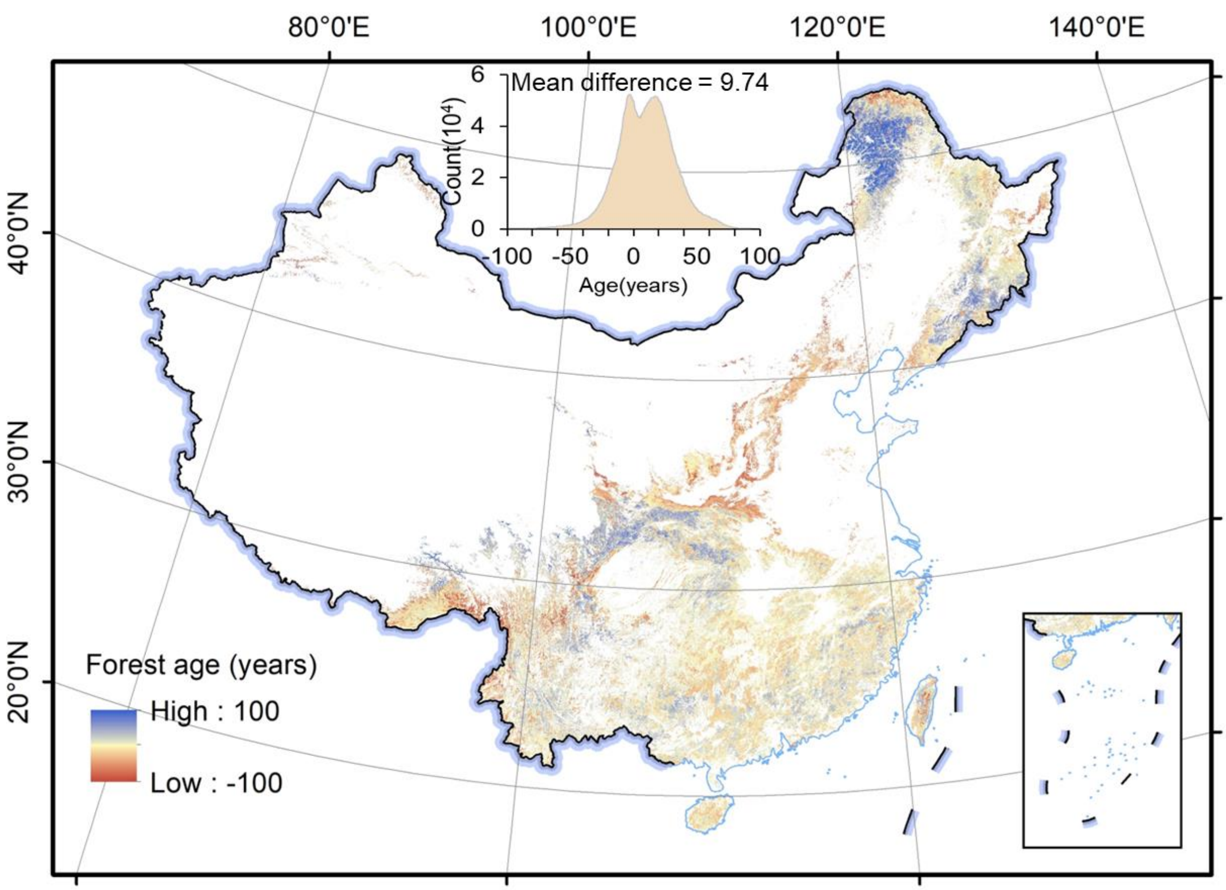


Figure 7. Comparison with global forest age product. The inset at the top left shows the frequency distribution of differences between the global forest age map and our estimated forest age map.

Table 3. China's mean forest age collected from published papers.

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

Source	Mean forest age (years)	Resolution	Mapping year
Zhang et al. (2017)	42.6	1 km	2013
Zhang et al. (2014)	43	1 km	2005
Dai Ming (2011)	40.6	8 km	1998
Wang et al. (2007)	<40	1 km	2001
(Xia et al., 2023)	44.0	1 km	2015
This study	56.1	30 m	2020

4 Discussion

A high-spatial resolution forest age map is an important input for accurately quantifying forest carbon storage and potential. Although several forest age maps for China were generated in the most recent decades, their spatial resolution is coarser, ranging from 1 to 8 km (e.g., Zhang et al., 2014; Zhang et al., 2017), which does not satisfy the application requirements for local-to-regional scales (Xiao et al., 2023). Therefore, we generated a 30 m resolution forest age map of China using remote sensing and inventory data for 2020. Validation against independent forest inventory samples, field measurements collected from published papers, and existing forest age products indicate that the estimated forest age map has R^2 of 0.51 to 0.62, and presented well spatial agreement with the existing forest age product. Such a high-resolution and timely forest age dataset is vital to assess ecological benefits of China's forests and to manage forest resources for sustainable development.

The generated forest age map indicates that 40.08% of forests are younger than 40 years, 38.11% are 41–80 years old, and 21.81% are over 80 years old. This result indicates that most forests in China are young, which is consistent with the findings of Zhang et al. (2017) and Zhang et al. (2014), even though the specific proportions might vary slightly, which is mainly because they produced forest age distribution data for the year of 2005, whereas our data represents forest age in 2020. Furthermore, similar to Zhang et al. (2017) and Zhang et al. (2014), forests younger than 40 years are primarily in southern and eastern China, whereas forests older than 80 years are predominantly in northeastern and southwestern China (Figure 4). We further analyse the forest age by using China's planted and natural forest mask generated by Cheng et al. (2023) for 2020. The results reveal that the average forest age for planted forests in China is 29.1 years with a standard deviation of 18.2 years, whereas natural forests have an average age of 69.7 years with a standard deviation of 30.6 years. This result aligns closely with the reported 16.5 years for China's planted forests in 2005 (which equates to approximately 31 years in 2020) by Yu et al. (2020b).

This study combines two methods to estimate forest age across China. The first method uses time-series remote sensing imagery and the LandTrendr algorithm to detect pixels that changed within the forest extent from 1985 to 2020. The forest age was estimated according the time since the last disturbance serving as a proxy for forest age. This approach has been

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

extensively used to estimate forest age and is generally acknowledged to be accurate and reliable for detecting disturbance (Hermosilla et al., 2016). For instance, Du et al. (2022) used the LandTrendr algorithm to detect planting times of global planted forests, and Xiao et al. (2023) estimated the forest age of young forests in China since 1984 by using the CCDC time-series algorithm. These successful cases validate the feasibility of using time-series change-detection algorithms to estimate the age of disturbed forests. In this study, we compared our change-detection derived forest age with the age of young forests provided by Xiao et al. (2023) (Figure 8). These two outcomes have a mean difference of −3.79 years (Figure 8a) and have a significant linear relationship with $R^2 = 0.53$ (Figure 8b). However, this approach only detects the forest age of disturbed areas based on Landsat images; for undisturbed forests, we propose using nonparametric MLAs to estimate forest age. Additionally, considering China's abundant forest resources, wide distribution, and complex terrain, this study proposes an approach to study forest-age estimation models based on MLAs.

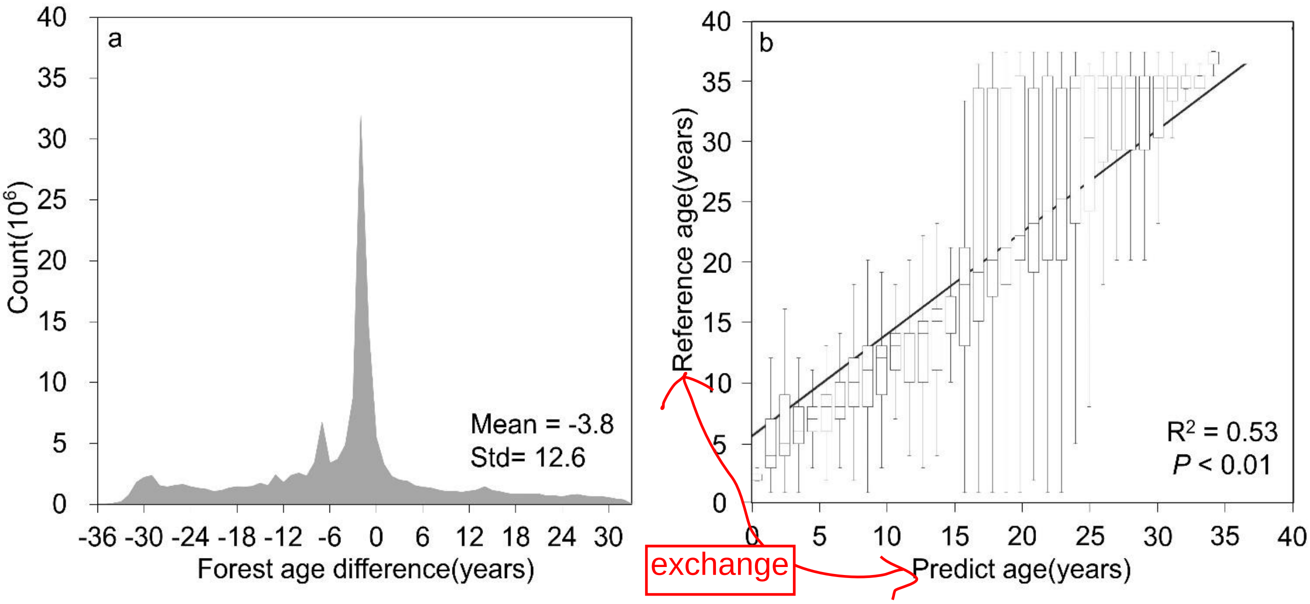


Figure 8. (a) Age difference and (b) linear relationship between estimated forest age and China's Young Forest Age dataset generated by Xiao et al. (2023).

We investigate in-depth the suitability of current mainstream MLAs for estimating forest age. For each vegetation division, we establish the optimal MLA and its optimal parameters (Table 2, Supplementary Table 3). Of the established MLAs, the ensemble learning approaches perform best for both training and evaluation compared with individual-based learners. Several previous studies support the idea that ensemble techniques have achieved better performance than that of its base learners (e. g. Rodriguez et al., 2006; Banfield et al., 2007; Canul-Reich et al., 2007; Rokach, 2009; De Stefano et al., 2011; Matloob et al., 2021). Bagging and boosting are two mainstream ensemble techniques in ensemble learning that combine multiple base models to improve predictive performance. Bagging reduces variance, whereas boosting reduces bias and improves overall model performance (Abbasi et al., 2022). However, most previous studies focus on bagging-based RF models to derive forest structure parameters in remote sensing fields (Simard et al., 2011; Cartus et al., 2012; Montesano et al., 2013; Matasci et al., 2018; Luther et al., 2019; Bolton et al., 2020). The present study highlights that super ensemble learning algorithms based on boosting, including GBDT, LightGBM, and CatBoost, demonstrate higher accuracy in estimating China's forest age compared to the bagging-based RF algorithm. Furthermore, within the current ensemble learning framework, the CatBoost algorithm based on boosting has a clear advantage for estimating forest age in China (Table 2). It produces optimal results in five vegetation zones and is as accurate as the best-performing algorithms in the remaining vegetation zones (Supplementary Table 4). Therefore, we recommend giving priority to the utilization of the CatBoost algorithm in deriving the forest structural parameters in China.

In the process of machine learning modelling for forest age estimation, we selected a total of 10 features, including canopy height, meteorological factors, soil factors, terrain factors, and human activities. Factor analysis indicates that canopy height has significantly influence forest age modelling, which is consistent with previous research, such as Zhang et al. (2017), who estimated forest age in China based on the relationship between canopy height and forest age. The main reason is that canopy height is typically correlated with the growth period (Sharma and Parton, 2007; Schumacher et al., 2020b; Lin et al., 2023). Young trees usually have shorter canopy height and, and as trees age, canopy height gradually increases (Yu et al., 2020b). Therefore, canopy height gives clues about tree age, and many age-estimation models are based on forest height (Lin et al., 2023). Terrain conditions also play important roles in all vegetation zones, especially the elevation and slope features (Figure 2). This is mainly because terrain factors are closely related to vegetation distribution, growth conditions, and hydrological processes (Fernández-Martínez et al., 2014) and affecting forest age estimation (Lin et al., 2008). Climate factors, including temperature and precipitation, also play a significant role in estimating forest age and have been applied to estimate global forest age (Besnard et al., 2021a). Climate elements are most pronounced in the SE and QT vegetation zones because these two zones belong to areas with extreme climates and pronounced seasonal variations (Zhang et al., 2018). The SE region has a warm and humid climate with abundant rainfall (Zhang et al., 2018), which aligns with seasonal growth, making it influential in forest age estimation. The QT region experiences extreme temperature fluctuations, with extremely cold winters and short and cool summers, significantly affecting tree growth rates and cycles (Zhang et al., 2021). Although soil and human activities seem to have a relatively smaller impact in this study, the high accuracy achieved in this study is attributed to the combined contributions of all factors.

Overall, we produce a reliable forest age map for China. This forest age product has been validated by independent field samples and compared against existing datasets with a R^2 ranging from 0.51 to 0.62 (Figure 6). However, it is imperative to acknowledge that intrinsic uncertainties arise from data-related constraints. Primarily, the utilization of forest mask that delineate planted and natural forests introduces an inescapable source of uncertainty, which is particularly high (approximately 10%) in the southern regions of China (Cheng et al., 2023a). Furthermore, the dependence on canopy-height data generated by Liu et al. (2022) as the crucial determinant in forest age estimation (Figure 2) necessitates meticulous consideration (Zhang

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

et al., 2017), giving the uncertainties in the canopy-height data (R^2=0.55) could strongly affect the accuracy in forest age modelling. Finally, when benchmarked against extant products, conspicuous disparities in forest age estimates appear within the northeastern and southwestern regions (Figure 6). These disparities, coupled with insights from forest inventory data, highlight the prevalence of older forests (exceeding 100 years) within these regions (Figures 1 and 4). The unique challenge posed by estimating the age of such older forests, characterized by sluggish growth rates (Maltman et al., 2023b), accentuates the sensitivity to crown height data. Consequently, the uncertainty associated with canopy height data is conspicuously accentuated in these regions.

5 Data availability

The 30 m resolution forest age map of China generated by this study is openly available at https://doi.org/10.5281/zenodo.8354262 (Cheng et al., 2023b). Please contact the authors for more detailed information

6 Conclusion

High-resolution and spatially explicit forest age mapping for China play a crucial role in accurately quantifying the current carbon sequestration of forest ecosystems and its potential in the future. Currently, publicly available China's forest age data suffer from low resolution and incomplete coverage of age ranges, making it difficult to meet the requirements of studies at various spatial scales. Therefore, this study combines time-series analysis of remote sensing imagery with MLAs to create the first 30 m resolution China's forest age map for the year of 2020. Validation against forest inventory data, field measurements, and existing products demonstrates the R^2 values between 0.51 and 0.62. The estimated forest age data reveal an average forest age of 56.1 years for China, with a standard deviation of 32.7 years. This dataset holds significant importance for understanding the carbon source and sink dynamics in China's forest ecosystem.

Author contributions

KC, YC and QG designed the research. KC, YC, TX performed the analysis, KC and YC wrote the paper. QG, HY, and QM supervised and reviewed the paper. HG and YR reviewed the manuscript. QG, KC, YC, WL collected the field measurements and existing remote sensing products. KC and YC contributed equally to this work.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgements

We would like to thank the editor and the two reviewers for their valuable comments.

Financial support

This research has been supported by the National Key Research and Development Program of China (grant no. 2022YFF130203), the International Research Center of Big Data for Sustainable Development Goals (grant no. CBAS2022GSP06), the National Natural Science Foundation of China (grant no. 42371329 and 31971575).

References

Abbasi, E., M. R. Alavi Moghaddam, and E. Kowsari. A systematic and critical review on development of machine learning based-ensemble models for prediction of adsorption process efficiency. J. Clean. Prod., 379:134588, https://doi.org/10.1016/j.jclepro.2022.134588, 2022

Akiba, T., S. Sano, T. Yanase, T. Ohta, and M. Koyama. Optuna: A Next-generation Hyperparameter Optimization Framework. Pages 2623–2631 Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Association for Computing Machinery, Anchorage, AK, USA. https://doi.org/10.48550/arXiv.1907.10902, 2019.

Alerskans, E., A.-S. P. Zinck, P. Nielsen-Englyst, and J. L. Høyer. Exploring machine learning techniques to retrieve sea surface temperatures from passive microwave measurements. Remote Sens Environ., 281:113220, https://doi.org/10.1016/j.rse.2022.113220, 2022.

Banfield, R. E., L. O. Hall, K. W. Bowyer, and W. P. Kegelmeyer. A Comparison of Decision Tree Ensemble Creation Techniques. IEEE Trans. Pattern Anal. Mach. Intell., 29:173-180, https://doi.org/10.1109/TPAMI.2007.250609, 2007.

Besnard, S., S. Koirala, M. Santoro, U. Weber, J. Nelson, J. Gutter, B. Herault, J. Kassi, A. N'Guessan, C. Neigh, B. Poulter, T. Zhang, and N. Carvalhais. Mapping global forest age from forest inventories, biomass and climate data. Earth Syst. Sci. Data., 13:4881-4896, https://doi.org/10.5194/essd-13-4881-2021, 2021a.

Bolton, D. K., P. Tompalski, N. C. Coops, J. C. White, M. A. Wulder, T. Hermosilla, M. Queinnec, J. E. Luther, O. R. van Lier, R. A. Fournier, M. Woods, P. M. Treitz, K. Y. van Ewijk, G. Graham, and L. Quist. Optimizing Landsat time series length for regional mapping of lidar-derived forest structure. Remote Sens Environ., 239:111645, https://doi.org/10.1016/j.rse.2020.111645, 2020.

Breiman, L. Random Forests. Machine Learning 45:5-32, https://doi.org/10.1023/A:1010933404324, 2001.

Canul-Reich, J., L. Shoemaker, and L. O. Hall. Ensembles of Fuzzy Classifiers. Pages 1-6 in 2007 IEEE International Fuzzy Systems Conference. https://doi.org/10.1109/FUZZY.2007.4295345, 2007.

Cartus, O., J. Kellndorfer, M. Rombach, and W. Walker. Mapping Canopy Height and Growing Stock Volume Using Airborne Lidar, ALOS PALSAR and Landsat ETM+. Remote Sens., 4:3320-3345, https://doi.org/10.3390/rs4113320, 2012.

Cheng, K., Y. Su, H. Guan, S. Tao, Y. Ren, T. Hu, K. Ma, Y. Tang, and Q. Guo. Mapping China's planted forests using high resolution imagery and massive amounts of crowdsourced samples. ISPRS J. Photogramm. Remote Sens., 196:356-371, https://doi.org/10.1016/j.isprsjprs.2023.01.005, 2023a.

Cheng, K, Y, Chen, T, Xiang, H, Yang, W, Liu, Y, Ren, H, Guan, T, Hu, Q, Ma, and Qinghua Guo. 2020 forest age map for China with 30 m resolution (1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.8354262, 2023b.

Dai, M., T. Zhou, L. Yang, and G. Jia. Spatial pattern of forest ages in China retrieved from national-level inventory and remote sensing imageries. GEOGRAPHICAL RESEARCH 30:172-184, https://doi.org/10.11821/yj2011010017, 2011 (in Chinese).

De Stefano, C., F. Fontanella, G. Folino, and A. S. di Freca. A Bayesian Approach for Combining Ensembles of GP Classifiers. Pages 26-35. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21557-5_5, 2011.

Du, Z., L. Yu, J. Yang, D. Coomes, K. Kanniah, H. Fu, and P. Gong. Mapping Annual Global Forest Gain From 1983 to 2021 With Landsat Imagery. IEEE J-STARS., 16:4195-4204, https://doi.org/10.1109/JSTARS.2023.3267796, 2023.

Du, Z., L. Yu, J. Yang, Y. Xu, B. Chen, S. Peng, T. Zhang, H. Fu, N. Harris, and P. Gong. A global map of planting years of plantations. Scientific Data., 9:141, https://doi.org/10.1038/s41597-022-01260-2, 2022.

Dutta, K. K., S. S. A, A. Victor, A. G. Nathu, M. A. Habib, and D. Parashar. Kannada Alphabets Recognition using Decision Tree and Random Forest Models. Pages 534-541 in 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS). https://doi.org/10.1109/ICISS49785.2020.9315972, 2020.

Fernández-Martínez, M., S. Vicca, I. A. Janssens, S. Luyssaert, M. Campioli, J. Sardans, M. Estiarte, and J. Peñuelas. Spatial variability and controls over biomass stocks, carbon fluxes, and resource-use efficiencies across forest ecosystems. Trees. 28:597-611, https://doi.org/10.1007/s00468-013-0975-9, 2014.

Guo, Y., Y. Zhou, X. Hu, and W. Cheng. Research on Recommendation of Insurance Products Based on Random Forest. Pages 308-311 in 2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI). https://doi.org/10.1109/MLBDBI48998.2019.00069, 2019.

Hermosilla, T., M. A. Wulder, J. C. White, N. C. Coops, G. W. Hobart, and L. B. Campbell. Mass data processing of time series Landsat imagery: pixels to data products for forest monitoring. Int J Digit Earth., 9:1035-1054, https://doi.org/10.1080/17538947.2016.1187673, 2016.

Jerome, H. F. Greedy function approximation: A gradient boosting machine. Ann Stat., 29:1189-1232, https://doi.org/10.1214/aos/1013203451, 2001.

Kennedy, R. E., Z. Yang, N. Gorelick, J. Braaten, L. Cavalcante, W. B. Cohen, and S. Healey. Implementation of the LandTrendr Algorithm on Google Earth Engine. Remote Sens., 10:691, https://doi.org/10.3390/rs10050691, 2018.

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

Kim, Hyunglok, Wade Crow, Xiaojun Li, Wolfgang Wagner, Sebastian Hahn, and Venkataraman Lakshmi. True global error maps for SMAP, SMOS, and ASCAT soil moisture data based on machine learning and triple collocation analysis. Remote Sens Environ., 298: 113776, https://doi.org/10.1016/j.rse.2023.113776, 2023.

Lavanya, K., S. Bajaj, P. Tank, and S. Jain. Handwritten digit recognition using hoeffding tree, decision tree and random forests — A comparative approach. Pages 1-6 in 2017 International Conference on Computational Intelligence in Data Science (ICCIDS). https://doi.org/10.1109/ICCIDS.2017.8272641, 2017.

Lin, G., B. Xia, Z. Zeng, and W. Huang. The Relationship between NDVI, Stand Age and Terrain Factors of Pinus elliottii Forest. Pages 232-236 in 2008 International Workshop on Education Technology and Training & 2008 International Workshop on Geoscience and Remote Sensing. https://doi.org/10.1109/ETTandGRS.2008.302, 2008.

Lin, X., R. Shang, J. M. Chen, G. Zhao, X. Zhang, Y. Huang, G. Yu, N. He, L. Xu, and W. Jiao. High-resolution forest age mapping based on forest height maps derived from GEDI and ICESat-2 space-borne lidar data. Agric For Meteorol., 339:109592, https://doi.org/10.1016/j.agrformet.2023.109592, 2023.

Liu, X., Y. Su, T. Hu, Q. Yang, B. Liu, Y. Deng, H. Tang, Z. Tang, J. Fang, and Q. Guo. Neural network guided interpolation for mapping canopy height of China's forests by integrating GEDI and ICESat-2 data. Remote Sens Environ., 269, https://doi.org/10.1016/j.rse.2021.112844, 2022.

Lundberg, S. and Lee, S.-I. A Unified Approach to Interpreting Model Predictions, arXiv [preprint], arXiv:1705.07874, https://doi.org/10.48550/arXiv.1705.07874, 2017.

Lundberg, S. M., Erion, G. G., and Lee, S.-I. Consistent Individualized Feature Attribution for Tree Ensembles, arXiv [preprint], arXiv:1802.03888, https://doi.org/10.48550/arXiv.1802.03888, 2019

Luther, J. E., R. A. Fournier, O. R. van Lier, and M. Bujold. Extending ALS-Based Mapping of Forest Attributes with Medium Resolution Satellite and Environmental Data. Remote Sens., 11:1092, https://doi.org/10.3390/rs11091092, 2019.

Maltamo, M., H. Kinnunen, A. Kangas, and L. Korhonen. Predicting stand age in managed forests using National Forest Inventory field data and airborne laser scanning. For. Ecosyst., 7:44, https://doi.org/10.1186/s40663-020-00254-z, 2020.

Maltman, J. C., T. Hermosilla, M. A. Wulder, N. C. Coops, and J. C. White. Estimating and mapping forest age across Canada's forested ecosystems. Remote Sens Environ., 290:113529, https://doi.org/10.1016/j.rse.2023.113529, 2023a.

Matasci, G., T. Hermosilla, M. A. Wulder, J. C. White, N. C. Coops, G. W. Hobart, and H. S. J. Zald. Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots. Remote Sens Environ., 209:90-106, https://doi.org/10.1016/j.rse.2017.12.020, 2018.

Matloob, F., T. M. Ghazal, N. Taleb, S. Aftab, M. Ahmad, M. A. Khan, S. Abbas, and T. R. Soomro. Software Defect Prediction Using Ensemble Learning: A Systematic Literature Review. IEEE Access, 9:98754-98771, https://doi.org/10.1109/ACCESS.2021.3095559, 2021.

Mekruksavanich, S., P. Jantawong, N. Hnoohom, and A. Jitpattanakul. Hyperparameter Tuning in Convolutional Neural Network for Face Touching Activity Recognition using Accelerometer Data. Pages 101-105 in 2022 Research, Invention,

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

and Innovation Congress: Innovative Electricals and Electronics (RI2C). https://doi.org/10.1109/RI2C56397.2022.9910262, 2022.

Montesano, P. M., B. D. Cook, G. Sun, M. Simard, R. F. Nelson, K. J. Ranson, Z. Zhang, and S. Luthcke. Achieving accuracy requirements for forest biomass mapping: A spaceborne data fusion method for estimating forest biomass and LiDAR sampling error. Remote Sens Environ., 130:153-170, https://doi.org/10.1016/j.rse.2012.11.016, 2013.

Niu, Y., V. Squires, and A. Jentsch. Risks of China's increased forest area. Science, 379:447-448, https://doi.org/10.1126/science.adg0210, 2023.

Pan, Y., R. A. Birdsey, J. Fang, R. Houghton, P. E. Kauppi, W. A. Kurz, O. L. Phillips, A. Shvidenko, S. L. Lewis, J. G. Canadell, P. Ciais, R. B. Jackson, S. W. Pacala, A. D. McGuire, S. Piao, A. Rautiainen, S. Sitch, and D. Hayes. A Large and Persistent Carbon Sink in the World. Forests. Science, 333:988-993, https://doi.org/10.1126/science.1201609, 2011.

Piao, S., Y. He, X. Wang, and F. Chen. Estimation of China's terrestrial ecosystem carbon sink: Methods, progress and prospects. Sci. China Earth Sci., 65:641 – 651, https://doi.org/10.1007/s11430-021-9892-6, 2022.

Ren, Y., Wei, X., Zhang, L., Cui, S., Chen, F., Xiong, Y., & Xie, P. Potential for forest vegetation carbon storage in Fujian Province, China, determined from forest inventories. Plant Soil., 345, 125-140, https://doi.org/10.1007/s11104-011-0766-2, 2011.

Rodriguez, J. J., L. I. Kuncheva, and C. J. Alonso. Rotation Forest: A New Classifier Ensemble Method. IEEE Trans. Pattern Anal. Mach. Intell., 28:1619-1630, https://doi.org/10.1109/TPAMI.2006.211, 2006.

Rokach, L. Taxonomy for characterizing ensemble methods in classification tasks: A review and annotated bibliography. Computational Statistics & Data Analysis, 53:4046-4072, https://doi.org/10.1016/j.csda.2009.07.017, 2009.

Sandha, S. S., M. Aggarwal, I. Fedorov, and M. Srivastava. Mango: A Python Library for Parallel Hyperparameter Tuning. Pages 3987-3991 in ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), https://doi.org/10.1109/ICASSP40776.2020.9054609, 2020.

Schumacher, J., M. Hauglin, R. Astrup, and J. Breidenbach. Mapping forest age using National Forest Inventory, airborne laser scanning, and Sentinel-2 data. For. Ecosyst., 7:60, https://doi.org/10.1186/s40663-020-00274-9, 2020a.

Sharma, M., and J. Parton. Height–diameter equations for boreal tree species in Ontario using a mixed-effects modelling approach. For. Ecol. Manag., 249:187-198, https://doi.org/10.1016/j.foreco.2007.05.006, 2007.

Simard, M., N. Pinto, J. B. Fisher, and A. Baccini. Mapping forest canopy height globally with spaceborne lidar. J. Geophys. Res. Biogeosci., 116, https://doi.org/10.1029/2011JG001708, 2011.

Su, Y., Q. Guo, T. Hu, H. Guan, S. Jin, S. An, X. Chen, K. Guo, Z. Hao, Y. Hu, Y. Huang, M. Jiang, J. Li, Z. Li, X. Li, X. Li, C. Liang, R. Liu, Q. Liu, H. Ni, S. Peng, Z. Shen, Z. Tang, X. Tian, X. Wang, R. Wang, Z. Xie, Y. Xie, X. Xu, X. Yang, Y. Yang, L. Yu, M. Yue, F. Zhang, and K. Ma. An updated Vegetation Map of China (1:1000000). Sci. Bull., 65:1125-1136, https://doi.org/10.1016/j.scib.2020.04.004, 2020.

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

Sun, Bochao, Wenjun Cui, Gaoyang Liu, Biao Zhou, and Weijian Zhao. A hybrid strategy of AutoML and SHAP for automated and explainable concrete strength prediction. Case Stud. Constr. Mater., 19: e02405, https://doi.org/10.1016/j.cscm.2023.e02405, 2023.

Tesfagergish, S. G., J. Kapočiūtė-Dzikienė, and R. Damaševičius. Zero-Shot Emotion Detection for Semi-Supervised Sentiment Analysis Using Sentence Transformers and Ensemble Learning. Appl. Sci., 12(17), 8662, https://doi.org/10.3390/app12178662, 2022.

Tian, L., L. Liao, Y. Tao, X. Wu, and M. Li. Forest Age Mapping Using Landsat Time-Series Stacks Data Based on Forest Disturbance and Empirical Relationships between Age and Height. Remote Sens., 15:2862, https://doi.org/10.3390/rs15112862, 2023.

Tong, X., M. Brandt, Y. Yue, P. Ciais, M. Rudbeck Jepsen, J. Penuelas, J.-P. Wigneron, X. Xiao, X.-P. Song, S. Horion, K. Rasmussen, S. Saatchi, L. Fan, K. Wang, B. Zhang, Z. Chen, Y. Wang, X. Li, and R. Fensholt. Forest management in southern China generates short term extensive carbon sequestration. Nat. Commun., 11, https://doi.org/10.1038/s41467-019-13798-8, 2020.

Tubiello, F. N., G. Conchedda, L. Casse, P. Hao, G. De Santis, and Z. Chen. A new cropland area database by country circa 2020. Earth Syst. Sci. Data Discuss., 2023:1-33, https://doi.org/10.5194/essd-2023-211, 2023.

Wang, S., J. M. Chen, W. M. Ju, X. Feng, M. Chen, P. Chen, and G. Yu. Carbon sinks and sources in China's forests during 1901–2001. J. Environ. Manage., 85:524-537, https://doi.org/10.1016/j.jenvman.2006.09.019, 2007.

Wang, Y., X. Wang, K. Wang, F. Chevallier, D. Zhu, J. Lian, Y. He, H. Tian, J. Li, J. Zhu, S. Jeong, and J. G. Canadell. The size of the land carbon sink in China. Nature, 603:E7-E9, https://doi.org/10.1038/s41586-021-04255-y, 2022.

Wei, Z., Y. Meng, W. Zhang, J. Peng, and L. Meng. Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the Tibetan Plateau. Remote Sens Environ., 225:30-44, https://doi.org/10.1016/j.rse.2019.02.022, 2019.

Xia, J., X. Xia, Y. Chen, R. Shen, Z. Zhang, B. Liang, J. Wang, and W. Yuan. Reconstructing Long-Term Forest Age of China by Combining Forest Inventories, Satellite-Based Forest Age and Forest Cover Data Sets. J. Geophys. Res. Biogeosci., 128: e2023JG007492, https://doi.org/10.1029/2023JG007492, 2023.

Xiao, Y., Q. Wang, X. Tong, and P. M. Atkinson. Thirty-meter map of young forest age in China. Earth Syst. Sci. Data., 15:3365-3386, https://doi.org/10.5194/essd-15-3365-2023, 2023.

Yu, Z., H. R. Zhao, S. R. Liu, G. Y. Zhou, J. Y. Fang, G. R. Yu, X. L. Tang, W. T. Wang, J. H. Yan, G. X. Wang, K. P. Ma, S. G. Li, S. Du, S. J. Han, Y. X. Ma, D. Q. Zhang, J. X. Liu, S. Z. Liu, G. W. Chu, Q. M. Zhang, and Y. L. Li. Mapping forest type and age in China's plantations. Sci. Total Environ., 744. https://doi.org/10.1016/j.scitotenv.2020.140790, 2020b.

Zhang H, J. Y， , Shen X, Li G, and Z. D. Rising Air Temperature and Its Asymmetry Under Different Vegetation Regions in China. Sci. Geol. Sin., 38 (2): 272-283, https://doi.org/10.13249/j.cnki.sgs.2018.02.014, 2018.

https://doi.org/10.5194/essd-2023-385
Preprint. Discussion started: 10 October 2023

Zhang, C., W. Ju, J. M. Chen, D. Li, X. Wang, W. Fan, M. Li, and M. Zan. Mapping forest stand age in China using remotely sensed forest height and observation data. J. Geophys. Res. Biogeosci., 119:1163-1179, https://doi.org/10.1002/2013JG002515, 2014.

Zhang, Y., Y. Yao, X. Wang, Y. Liu, and S. Piao. Mapping spatial distribution of forest age in China. Earth Space Sci., 4:108-116, https://doi.org/10.1002/2016EA000177, 2017.

Zhang, Z., F. Zhang, L. Wang, A. Lin, and L. Zhao. Biophysical climate impact of forests with different age classes in mid- and high-latitude North America. For. Ecol. Manag., 494:119327. https://doi.org/10.1016/j.foreco.2021.119327, 2021.