2020 forest age map for China with 30 m resolution

Anonymous Referee #2

Thank you for your constructive comments and suggestions on our manuscript entitled "2020 forest age map for China with 30 m resolution" (Manuscript Number: essd-2023-385). We appreciate the time and effort you have invested in reviewing our work. We agree with your assessment of deficiencies in the introduction, method, and discussion sections, and we have addressed these issues comprehensively in the revised version. In the revised manuscript, we have offered a more detailed explanation of the algorithm and its combination with machine learning, addressing any specific concerns you have raised. Also, we have carried out a comprehensive check on the structure and language of the full text. All changes were marked in highlight text in the revised manuscript. Please find our responses to your comments below:

Reply:
1. The introduction should highlight the innovation, why is it LandTendr algorithm instead of CCDC and other algorithms?
Reply: Thanks to the reviewer for pointing this out, we strengthened the innovation of this study in the introduction and added a review of the LandTendr algorithm in predicting the age of forests in the introduction. Please see below or check in Line 70-79.

Disturbance detection approaches, capable of identifying the time of the most recent stand-replacing disturbance, have proven accurate in forest age estimation (Li et al. 2024), which mainly include Landsat-based Detection of Trends in Disturbance and Recovery (LandTrendr) (Kennedy et al. 2010), Continuous Change Detection and Classification (CCDC) (Zhu and Woodcock 2014), the Vegetation Change Tracker (VCT) (Huang et al. 2010), Breaks for Additive Season and Trend (BFAST) (Verbesselt et al. 2010a; Verbesselt et al. 2010b). Among these algorithms, LandTrendr has been recognized for its efficiency in detecting forest disturbances such as fire, deforestation, and urban expansion (de Jong et al. 2021; Rodman et al. 2021). For instance, Li et al (2024) mapped planted forest age using LandTrendr algorithm, demonstrating its efficiency and reliable for forest age mapping. However, these approaches are limited to obtaining forest age in areas with disturbance recorded by remote sensing, thus restricting a comprehensive understanding of forest age structures. Consequently, it is necessary to develop a framework that can provide comprehensive forest age

information on a large scale.

2. Line 86: Why is China divided into these eight vegetation regions?

Reply: Given this topographic, climatic and vegetation species diversity, China harbors eight main vegetation regions (Wu 1980; Zhang ,2007). The age distribution characteristics of different vegetation types vary (Harms et al.2001; Chen et al.2007; Magnani et al.2001; Li et al.2007). Therefore, we have established age models for each vegetation region. The application of eight vegetation regions division is widely applied in the ecological field and generating remote sensing products at national scale (Sun, Sang, and Axmacher 2020; He et al. 2017; Wang et al. 2021; Liu et al. 2022). For the division basis of eight vegetation regions, we have added literature citations in Line 96.

3. Line 104: Why is there a distinction between planted forest and natural forest, what is the purpose of the authors, respectively, and what is the difference between planted forest and natural forest in algorithm.

Reply:

The effects of natural and planted forests on forest age are different. Therefore, we took forest origin (plantation forest and natural forest) as the influence factor of forest age, processed it into dummy variable for quantification, and participated in the machine learning modeling of forest age model.

Different forest origins may lead to different forest age distributions and changes.

Natural and planted forests have distinct age distributions. Natural forests have a complex stand age distribution, with different tree species and heights having different ages, forming a complex canopy structure and species diversity. Planted forests have a simple age distribution, with mostly the same tree species and height, and lower ages, because planted forests are mainly for meeting human economic needs, so they are harvested when the trees reach a certain growth period, making it difficult to form long-lived trees (Guo et al. 2014; Jiao et al. 2003).

4. Line 118: the resolution of Climate data (30 arc-second) is not consistent with it in Table 1.

Reply: 30 arc-seconds is approximately equal to 1 kilometer. The mistake has been corrected according to your detailed professional suggestion in Line 146.

5. Line 123: Why the PC1 gives annual trends in temperature and precipitation, PC2 gives seasonal variations in temperature and precipitation, and PC3 gives precipitation and temperature extremes?

Reply: We apologize for the omission of the PCA principal component analysis results

in the Supplementary section, and have already been supplemented. Please check the supplementary Table 1.

Supplementary Table1 shows the factor score coefficient matrix of the first three components PC1, PC2, PC3, we find that:

(1) PC1 is related to BIO1, BIO2, BIO5, BIO6, BIO9, BIO10, BIO12 and BIO18,which represent annual trends in temperature and precipitation.

(2) PC2 is related to BIO3, BIO4, BIO7, BIO11 and BIO15. These factors mainly reflect seasonal variations in temperature and precipitation

(3) PC3 is related to BIO8, BIO13, BIO14, BIO16, BIO17, and BIO19, and these factors reflect precipitation and temperature extremes.

Therefore, the PC1 gives annual trends in temperature and precipitation, PC2 gives seasonal variations in temperature and precipitation, and PC3 gives precipitation and temperature extremes.

Supplementary Table1 Principal component analysis of 19 bioclimatic variables

Variables	Describe	Principal Component		
		PC1	PC2	PC3
BIO1	Annual Mean Temperature	0.272	0.112	0.224
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))	-0.205	0.177	-0.145
BIO3	Isothermality (BIO2/BIO7) ($\times$100)	-0.129	0.403	-0.123
BIO4	Temperature Seasonality (standard deviation $\times$100)	0.039	-0.476	0.128
BIO5	Max Temperature of Warmest Month	0.266	-0.093	0.261
BIO6	Min Temperature of Coldest Month	0.262	0.181	0.200
BIO7	Temperature Annual Range (BIO5-BIO6)	-0.059	-0.447	0.041
BIO8	Mean Temperature of Wettest Quarter	0.221	0.029	0.447
BIO9	Mean Temperature of Driest Quarter	0.253	0.220	0.097
BIO10	Mean Temperature of Warmest Quarter	0.272	-0.069	0.264
BIO11	Mean Temperature of Coldest Quarter	0.243	0.253	0.168
BIO12	Annual Precipitation	0.272	-0.013	-0.271
BIO13	Precipitation of Wettest Month	0.256	0.077	-0.266
BIO14	Precipitation of Driest Month	0.255	-0.147	-0.257
BIO15	Precipitation Seasonality (Coefficient of Variation)	-0.198	0.286	0.066
BIO16	Precipitation of Wettest Quarter	0.255	0.110	-0.269
BIO17	Precipitation of Driest Quarter	0.248	-0.165	-0.275
BIO18	Precipitation of Warmest Quarter	0.211	0.210	-0.162
BIO19	Precipitation of Coldest Quarter	0.250	-0.124	-0.303

Supplementary Figure1 Scree plot of 19 bioclimatic variables

6. Line 146: The algorithm description and detail of forest age in changed area and forest age in unchanged area is missing.

Reply: Thank you very much for your professional advice, we have adjusted a clear description of the overall algorithm design and objectives. In addition, we have described how the two methods will be integrated in conjunction with each other. Please check below (Line 211-217).

Given the extensive forest coverage in China, it is challenging to handle such large forest area for ML and the LandTrendr algorithm to estimate forest age, even with our vegetation zoning efforts. To enhance the efficiency of forest age estimation and conserve computational resources, we have divided China into 1°×1° grids (see Supplementary Figure 2), limiting ML and LandTrendr algorithms to estimate forest age within each grid. Subsequently, we merge the predictive results from each grid using the Mosaic New Raster tool in ArcGIS Pro 3.0 to obtain nationwide forest age map. Finally, the forest age map estimated through LandTrendr algorithm is applied to update the ML-based results to produce China's forest age data.

7. Line 235: which is the year of China's forest age map? Is it 2020? Use LandTrendr to look back at the age of forests during 1985-2020, but why not

present the age change in the results. Please refer to Xiao, 2023 and Huang, 2023 for details

Reply:We appreciate the reviewer's insightful suggestion. After sorting out the whole method part, we present the corresponding age detected by LandTrendr in Figure 4b, and made description about this result. Please see below or check in Line 260-265.

Based on the optimal MLAs and the LandTrendr change detection algorithm, we have obtained forest age data for China as shown in Figure 4. Figure 4a presents the nationwide distribution of forest age as estimated by MLAs, whereas Figure 4b displays the age distribution from 1985 and 2020 as determined through change detection. The results reveal that reforestation activities from 1985 and 2020 are primarily situated in the southern, southeastern, and northern China, aligning with the findings from Xiao et al (2023). Furthermore, estimates derived from MLAs indicate that old-growth forests are primarily located in the northeast and southwest regions of China.

Figure 4. Forest age estimated from LandTrendr (a), MLA (b), and final China's forest age distribution (c) with 30 m resolution.

8. The discussion part of this paper is a little limited, which is all about input data constraints, having no limitations of the algorithm. whether forest age estimation should be combined with forest restoration and other conditions, and whether the parameters of LandTrendr algorithm will affect the detection results.

Reply: We agreed with the reviewer's comments and added the method limitations to the discussion section, including the current limitations of machine learning and limitations specific to the change detection algorithm. Please see below or check in Line 388-400.

Regarding to methods, we combined MLA and disturbance detection approach to derive forest age, for MLA, overfitting is a common challenge, where a model learns the training data too well and fails to generalize to unseen data (Belgiu and Drăguţ 2016). The results presented in Supplementary Table 5 suggest that the constructed forest age models exhibit a certain degree of overfitting, which can cause some errors for forest age estimation. Addressing the issue of overfitting, data augmentation and exploring new deep learning algorithms may be promising directions for further investigation. For LandTrendr approach, it is affected by different parameters such as input bands, vegetation parameters (NBR index), climates, vegetation, terrain and atmospheric conditions (Banskota et al. 2014; Hermosilla et al, 2015; Hua et al, 2021; Huang et al, 2023; Yang et al, 2018). China's unprecedented development has led to extensive land cover changes, making it one of the most intensively managed forest regions globally (Tong et al., 2020). This has resulted in significant forest fragmentation, posing challenges in using NBR and other indices for change detection (Li et al 2024), and creating uncertainty in forest age identification. Furthermore, while the LandTrendr algorithm effectively captures sharp disturbances like fires, clearcutting, and reforestation, it falls short in detecting subtle changes such as silviculture and thinning (Huang et al. 2023; Zhu 2017), This limitation may lead to the omission of young trees and an overestimation of forest age.

References:

Banskota, A., Kayastha, N., Falkowski, M. J., Wulder, M. A., Froese, R. E., & White, J. C. (2014). Forest Monitoring Using Landsat Time Series Data: A Review. Canadian Journal of Remote Sensing, 40(5), 362–384. https://doi.org/10.1080/07038992.2014.987376

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24-31

Chen, H. Y. H., & Popadiouk, R. V. (2002). Changes in spatial patterns with different forest age classes in boreal mixedwood forests. Forest Ecology and Management,

158(3), 283-293.

Chen, Dong, Tatiana V. Loboda, Alexander Krylov, and Peter V. Potapov. Mapping stand age dynamics of the Siberian larch forests from recent Landsat observations, Remote Sensing of Environment, 187: 320-31. https://doi.org/10.1016/j.rse.2016.10.033, 2016.

de Jong, S.M., Shen, Y., de Vries, J., Bijnaar, G., van Maanen, B., Augustinus, P., & Verweij, P. (2021). Mapping mangrove dynamics and colonization patterns at the Suriname coast using historic satellite data and the LandTrendr algorithm. International Journal of Applied Earth Observation and Geoinformation, 97, 102293

Guo, Q., & Ren, H. (2014). Productivity as related to diversity and age in planted versus natural forests. Global Ecology and Biogeography, 23(12), 1461-1471.

Harms, K. E., Condit, R., Hubbell, S. P., & Foster, R. B. (2001). Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. Journal of Ecology, 89(6), 947-959.

He, Bin, Aifang Chen, Weiguo Jiang, and Ziyue Chen. 2017. 'The response of vegetation growth to shifts in trend of temperature in China', Journal of Geographical Sciences, 27: 801-16.

Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., & Hobart, G. W. (2015). Regional detection, characterization, and attribution of annual forest change from 1984 to 2012 using Landsat-derived time-series metrics. Remote Sensing of Environment, 170, 121–132. https://doi.org/10.1016/j.rse.2015.09.004

Hua, J., Chen, G., Yu, L., Ye, Q., Jiao, H., & Luo, X. (2021). Improved Mapping of Long-Term Forest Disturbance and Recovery Dynamics in the Subtropical China Using All Available Landsat Time-Series Imagery on Google Earth Engine Platform. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 2754–2768. https://doi.org/10.1109/JSTARS.2021.3058421

Huang, C., Goward, S.N., Masek, J.G., Thomas, N., Zhu, Z., & Vogelmann, J.E. (2010). An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. Remote Sensing of Environment, 114, 183-198

Huang, Z., Li, X., Du, H., Zou, W., Zhou, G., Mao, F., Fan, W., Xu, Y., Ni, C., Zhang, B., Chen, Q., Chen, J., & Hu, M. (2023). An Algorithm of Forest Age Estimation Based on the Forest Disturbance and Recovery Detection. IEEE Transactions on Geoscience and Remote Sensing, 61, 1–18. https://doi.org/10.1109/TGRS.2023.3322163

Jiao-Jun, Z., Zhi-Ping, F., De-Hui, Z., Feng-qi, J., & Takeshi, M. (2003). Comparison of stand structure and growth between artificial and natural forests of Pinus sylvestiris var. mongolica on sandy land. Journal of Forestry Research, 14(2), 103-111.

Kennedy, R.E., Yang, Z., & Cohen, W.B. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr — Temporal segmentation algorithms. Remote Sensing of Environment, 114, 2897-2910

Li, P., & Duan, Q. (2011). Age structure and spatial patterns of different forest types in the Qinling Mountains, China. Chinese Geographical Science, 21(5), 605-614.

Li, P., Li, H., Si, B., Zhou, T., Zhang, C., & Li, M. (2024). Mapping planted forest age using LandTrendr algorithm and Landsat 5–8 on the Loess Plateau, China. Agricultural and Forest Meteorology, 344, 109795. https://doi.org/10.1016/j.agrformet.2023.109795

Liu, Xiaoqiang, Yanjun Su, Tianyu Hu, Qiuli Yang, Bingbing Liu, Yufei Deng, Hao Tang, Zhiyao Tang, Jingyun Fang, and Qinghua Guo. 2022. 'Neural network guided interpolation for mapping canopy height of China's forests by integrating GEDI and ICESat-2 data', Remote Sensing of Environment, 269: 112844.

Magnani, F., Mencuccini, M., Borghetti, M., Berbigier, P., Berninger, F., Delzon, S.,& Jarvis, P. G. (2007). The human footprint in the carbon cycle of temperate and boreal forests. Nature, 447(7146), 848-850.

Rodman, K.C., Andrus, R.A., Veblen, T.T., & Hart, S.J. (2021). Disturbance detection in landsat time series is influenced by tree mortality agent and severity, not by prior disturbance. Remote Sensing of Environment, 254, 112244

Sun, Siqi, Weiguo Sang, and Jan Christoph Axmacher. 2020. 'China's national nature reserve network shows great imbalances in conserving the country's mega-diverse vegetation', Science of The Total Environment, 717: 137159.

Verbesselt, J., Hyndman, R., Newnham, G., & Culvenor, D. (2010a). Detecting trend and seasonal changes in satellite image time series. Remote Sensing of Environment, 114, 106-115

Verbesselt, J., Hyndman, R., Zeileis, A., & Culvenor, D. (2010b). Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. Remote Sensing of Environment, 114, 2970-2980

Wang, Xiqiang, Rensheng Chen, Chuntan Han, Yong Yang, Junfeng Liu, Zhangwen Liu, Shuhai Guo, and Yaoxuan Song. 2021. 'Soil temperature change and its regional differences under different vegetation regions across China', International Journal of Climatology, 41: E2310-E20.

Wu, Z.Y., 1980.Vegetation of China. Science Press, Beijing, China

Yang, Y., Erskine, P. D., Lechner, A. M., Mulligan, D., Zhang, S., & Wang, Z. (2018). Detecting the dynamics of vegetation disturbance and recovery in surface mining area via Landsat imagery and LandTrendr algorithm. Journal of Cleaner Production, 178, 353–362. https://doi.org/10.1016/j.jclepro.2018.01.050

Zhang X S.2007. Vegetation map of the People's Republic of China (1:1 000 000) . Geology Press, Beijing, China

Zhu, Z. (2017). Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. ISPRS Journal of Photogrammetry and Remote Sensing, 130, 370-384

Zhu, Z., & Woodcock, C.E. (2014). Continuous change detection and classification of land cover using all available Landsat data. Remote Sensing of Environment, 144, 152-171

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, Nanjing 210023, China

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

Abstract. A high-resolution, spatially explicit forest age map is essential for quantifying forest carbon stocks and carbon sequestration potential. Prior attempts to estimate forest age on a national scale in China have been limited by sparse resolution and incomplete coverage of forest ecosystems, attributed to complex species composition, extensive forest areas, insufficient field measurements, and inadequate methods. To address these challenges, we developed a framework that combines machine learning algorithms (MLAs) and remote sensing time series analysis for estimating the age of China's forests. Initially, we identify and develop the optimal MLAs for forest age estimation across various vegetation divisions based on forest height, climate, terrain, soil, and forest-age field measurements, utilizing these MLAs to ascertain forest age information. Subsequently, we apply the LandTrendr time series analysis to detect forest disturbances from 1985 to 2020, with the time since the last disturbance serving as a proxy for forest age. Ultimately, the forest age data derived from LandTrendr is integrated with the result of MLAs to produce the 2020 forest age map of China. Validation against independent field plots yielded an R^2 ranging from 0.51 to 0.63. On a national scale, the average forest age is 56.1 years (standard deviation of 32.7 years). The Qinghai-Tibet Plateau alpine vegetation zone possesses 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 offers crucial insights for comprehensively understanding the ecological benefits of China's forests and to sustainably manage China's forest resources. The map is available at http://dx.doi.org/10.5281/zenodo.8354262 (Cheng et al., 2023b).

1 Introduction

Forest age is crucial for gaining insights into forest ecosystem succession and condition, thereby playing a pivotal role in comprehending the ecological benefits of forests (Lin et al., 2023). China's forests have undergone significant disruptions due

to natural disasters and human activities over the past few decades, leading to notable changes in the forest age structure (Niu et al., 2023). Consequently, this scenario presents considerable challenges in accurately assessing forest ecosystem carbon storage (Pan et al., 2011; Tong et al., 2020). The complexity of species composition, extensive forest areas, limited field measurements, and ineffective methods have led to existing national-scale estimates of China's forest age focusing on either sparse resolution (Zhang et al., 2017) or partial forest ecosystem coverage (Xiao et al., 2023). This has resulted in significant uncertainties in evaluating the carbon sources and sinks within China's forest ecosystem (Piao et al., 2022; Wang et al., 2022). Therefore, there is an urgent requirement for time-efficient, high-resolution mapping of forest age across China.

At present, China's forest age data is primarily obtained via the national forest inventory, noted for its high accuracy (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, the traditional forest inventory method struggles to accurately and timely capture the complete age distribution and spatial characteristics of China's forests.

Remote sensing technology has demonstrated effectiveness in estimating forest cover (Su et al., 2020; Tubiello et al., 2023) and forest structure (Yu et al., 2020; Maltman et al., 2023) across various scales. The availability and sharing of Landsat time series data, along with the development of Google Earth Engine (GEE) cloud-processing platform have significantly facilitated the application of remote sensing in forest age estimation. Several studies have been conducted to map China's forest age. Xiao et al. (2023) mapped the age of China's young forests at 30 m resolution using time series Landsat imagery. Yu et al. (2020) produced a 1-km resolution map of the age for planted forests in China. Zhang et al. (2017) developed 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 maps are typically undertaken at coarser spatial resolutions (e.g., 1 km), with finer resolutions (e.g., 30 m) being limited to young forests. There remains a lack of high-resolution forest age spatial dataset covering the entire forest region of China.

Statistical models and disturbance detection approaches are two common methods utilized in remote sensing-based forest age estimations. Statistical models deduce forest age by establishing a coherent relationship between remote sensing features and field-collected empirical samples, including parametric regression approaches (Maltamo et al., 2020; Schumacher et al., 2020) and nonparametric machine learning algorithms (MLAs). Growth models represent one of the most widely used parametric models for estimating forest age (Zhang et al., 2014; Zhang et al., 2017; Yu et al., 2020). However, this type of model relies on tree species information, posing challenges in forest age derivation when such data is lacking, particularly at large scales. MLAs has been employed for forest age estimation, owing to their flexibility in addressing complex problems (Alerskans et al., 2022). For examples, Huang et al (2023) integrated random forest (RF) to derived forest age. Chen et al (2016) mapped forest stand age dynamics using RF and Landsat imagery. Nevertheless, the application of MLAs to estimate national forest age has not been deeply explored. Most previous studies used a single MLA, such as Random Forest (RF) (Besnard et al., 2021b), to estimate forest age. The extensive distribution of forests, diverse forest types, and varying terrain and climate

conditions in China make it difficult in using a single model for accurately forest age determination on national scale. Therefore, exploring the applicability of MLAs for forest age estimation in various regions of China is essential.

Disturbance detection approaches, capable of identifying the time of the most recent stand-replacing disturbance, have proven accurate in forest age estimation (Li et al. 2024), which mainly include Landsat-based Detection of Trends in Disturbance and Recovery (LandTrendr) (Kennedy et al. 2010), Continuous Change Detection and Classification (CCDC) (Zhu and Woodcock 2014), the Vegetation Change Tracker (VCT) (Huang et al. 2010), Breaks for Additive Season and Trend (BFAST) (Verbesselt et al. 2010a; Verbesselt et al. 2010b). Among these algorithms, LandTrendr has been recognized for its efficiency in detecting forest disturbances such as fire, deforestation, and urban expansion (de Jong et al. 2021; Rodman et al. 2021). For instance, Li et al (2024) mapped planted forest age using LandTrendr algorithm, demonstrating its efficiency and reliable for forest age mapping. However, these approaches are limited to obtaining forest age in areas with disturbance recorded by remote sensing, thus restricting a comprehensive understanding of forest age structures. Consequently, it is necessary to develop a framework that can provide comprehensive forest age information on a large scale.

The objective of the present study is to generate the first China's forest age dataset at 30 m resolution using multi-source datasets through combining remote sensing time series analysis and MLAs. This involves: 1) identifying the most optimal MLA for age estimation across various vegetation zone in China and estimating the age of China's forests. 2) Utilizing LandTrendr disturbance detection algorithm to identify the most recent forest disturbances from 1985 to 2020, and estimating the forest of these disturbed areas. 3) Using the forest age derived by LandTrendr algorithm to update the result of MLAs to generate China's forest age map, which is then subjected to validation. 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 667 m^2 sample 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 (Liu et al. 2022) (Figure 1b), each containing at least 436 sample plots for building MLA models to estimate forest age (Figure 1d).

98

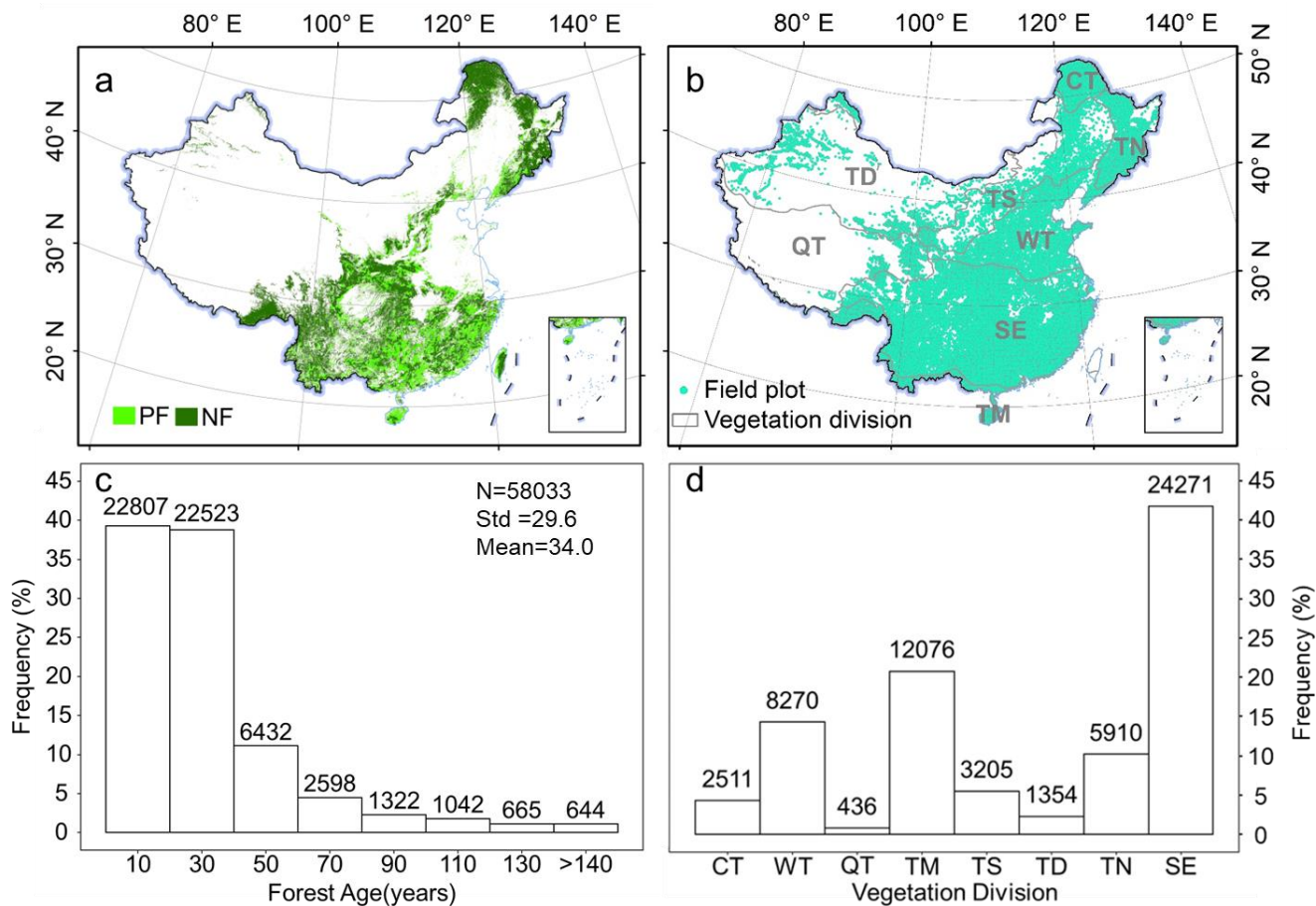



Figure 1. Forest mask and field sample distribution. (a) Planted forest and natural forest mask generated by Cheng et al. (2023a). (b) Distribution of field samples over eight vegetation divisions. (c) Frequency distribution of field sample ages. (d) Frequency distribution of field samples for eight 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

111   disturbance. NBR has been proved effective in numerous studies detecting forest disturbance (Du et al., 2023; Tian et al.,

112   2023). It is calculated as follows by using the near-infrared (NIR) and short-wave infrared (SWIR) bands:

$$NBR = \frac{NIR - SWIR}{NIR + SWIR}.$$

(Eq. 1)

113   2.1.3 Forest mask

114   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

115   mapping. This dataset is produced by integrating multisource remote-sensing data and a large number of crowdsourced samples,

116   with an overall accuracy of over 80% (Cheng et al., 2023a). In this study, we employ this dataset as a forest mask and utilize

117   a combination of time series change detection algorithms and MLAs to trace the age of these planted and natural forests.

118   2.1.4 Forest height data

119   The canopy height data for China was downloaded from the website (https://3decology.org/), which was generated based on

120   deep learning by integrating Global Ecosystem Dynamics Investigation and Ice, Cloud and land Elevation Satellite -2 data.

121   This dataset has a spatial resolution of 30 m and corresponds to 2019. The accuracy of this national forest canopy height data

122   was assessed by comparing three independent validation datasets, indicating high accuracy for the canopy height product by

123   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

124   is consistent with the forest extent mentioned earlier for planted and natural forests, ensuring spatial consistency when

125   estimating forest age.

126   2.1.5 Climate data

127   Climate data were acquired from WorldClim 2.1 (https://worldclim.org/), which offers 19 bioclimatic variables, including

128   temperature and precipitation, with 30 arc-second resolutions. The 19 bioclimatic variables include annual trends, seasonality,

129   and extreme environmental factors in temperature and precipitation. We resampled the 19 GeoTiff (.tif) files to 30 m resolution

130   using a nearest-resampling method for spatial resolution consistency. To reduce the dimension of bioclimatic variables, we

131   applied a principal component analysis to map the 19 bioclimatic variables into a new principal component (PC) space. We

132   use the first three components PC1, PC2, PC3 to represent the climate factors. According to the results of the analysis, PC1

133   gives annual trends in temperature and precipitation, PC2 gives seasonal variations in temperature and precipitation, and PC3

134   gives precipitation and temperature extremes (Supplementary Table 1).

135   2.1.6 Soil data

136   Soil data were extracted from the harmonized world soil database, V1.2, developed jointly by the Food and Agriculture

137   Organization of the United Nations, the International Institute for Applied Systems, the ISRIC-World Soil Information, the

138   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., 2021). 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.

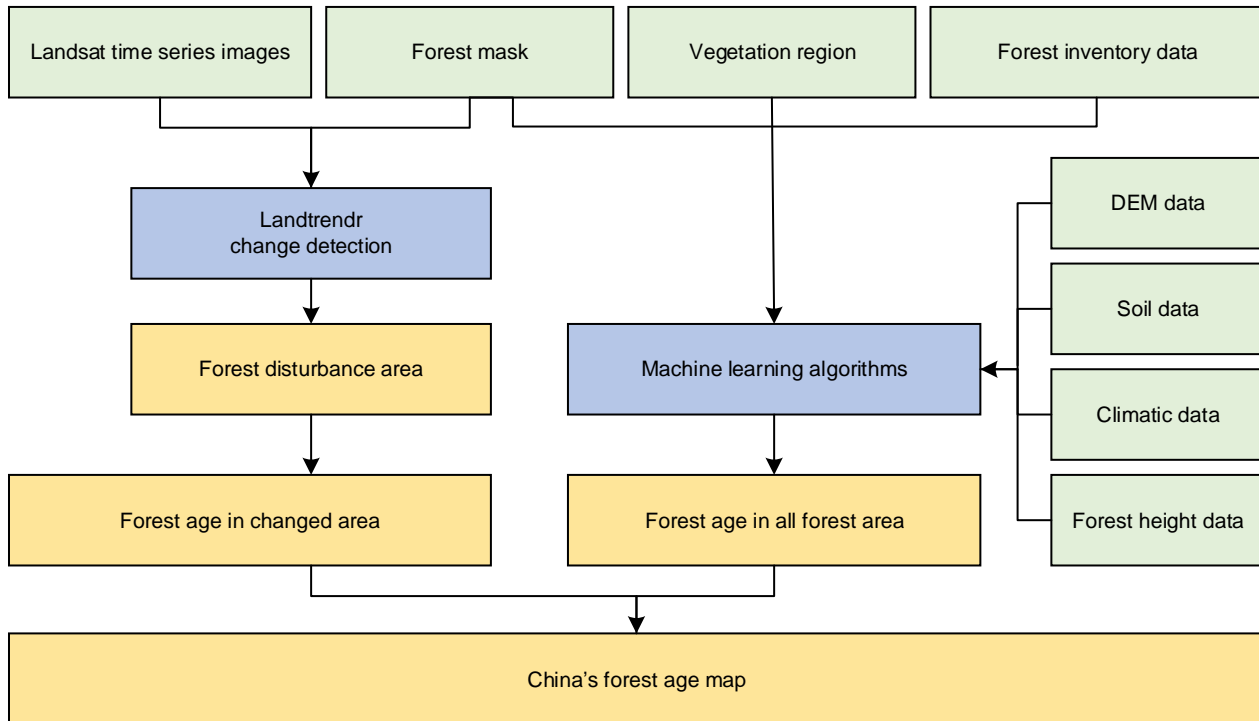
Table 1. Descriptions of variables used to estimate the forest age of China.

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)	30 arc-second	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 MLA-based approach and the LandTrendr disturbance detection approach. First, the MLA-based approach estimates ages for forest regions using forest inventory and multi-source remote sensing data. Second, the LandTrendr algorithm is applied to detect stand-replacing disturbances based on the Landsat time series images. Third, we use the forest age map detected by LandTrendr to update the forest age map derived using MLA-based approach,

154 and generate the China's forest age map with 30 m resolution. Figure 2 shows a detailed flowchart describing the framework

155 for forest age estimation proposed in this study.

156

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

158 **2.2.1** **Machine learning approach**

159 **(1) MLA selection**

160 This study used the following model-screening procedure to explore which model works best for each vegetation division.

161 First, we used the automated machine learning (Auto-ML) open-source Python library LazyPredict to filter for alternative

162 models. LazyRegressor (including 40 MLAs) was used to build stand-age estimation models based on all data, which helps to

163 understand which MLA works well without tuning parameters. The performing models with R^2 greater than 0.60 in each

164 vegetation division were concentrated in thirteen MLAs (Supplementary Table 2). Second, by splitting training data and testing

165 data, the top three MLAs for each vegetation division were determined (Supplementary Table 2). It can be found that the

166 potential optimal models of eight vegetation divisions is concentrated in RF, Gradient Boosting Decision Tree (GBDT),

167 Histogram Gradient Boosting (HistGradientBoost), Light Gradient Boosting Machine (LightGBM), and Categorical Boosting

168 (CatBoost).

169 RF is an ensemble learning method that combines multiple decision trees (Breiman 2001; Dutta et al., 2020). It leverages the

170 wisdom of crowds to make accurate predictions. RF mitigates overfitting and provides robust results by training each tree on

171 a random subset of the data and features (Lavanya et al., 2017; Guo et al., 2019). GBDT is an ensemble technique that builds

172 a strong predictive model by sequentially training decision trees (Jerome 2001). Each tree corrects the errors of its predecessor

173 (Wei et al., 2019), resulting in a highly accurate and robust model. HistGradientBoost is a variant of GBDT that employs

174 histogram-based techniques. It efficiently approximates data distributions and reduces memory consumption during training.

175 This algorithm is particularly beneficial when dealing with large datasets and complex features (Tesfagergish et al., 2022).

176 LightGBM is a gradient-boosting framework that prioritizes speed and efficiency. It employs a histogram-based approach and

177 parallel computing, making it suitable for large datasets. CatBoost is a new modification gradient boosting algorithm that is

178 designed specifically for handling categorical features. It automatically encodes categorical variables, simplifying the data pre-

179 processing stage. CatBoost is known for its robustness and efficiency, can achieve high accuracy on a small-scale dataset.

180 We implemented RF, GBDT, and HistGradientBoost by using the Scikit-learn package of Python 3.9.11, while the LightGBM

181 and CatBoost algorithms were constructed by using the lightgbm and catboost packages of Python 3.9.11.

182 **(2) Hyperparameter tuning**

183 Hyperparameter tuning of MLAs is critical in the ML model training process because it significantly enhances the model's

184 performance, generalization capability, and adaptability (Sandha et al., 2020). Bayesian optimization has been selected for

185 hyperparameter tuning due to its complicated derivative evaluation, and nonconvex-function-related features (Mekruksavanich

186 et al., 2022). It is implemented by using Optuna, an open source hyperparameter optimization framework to automate

187 hyperparameter searches (Akiba et al., 2019). The hyperparameters and their searching range in MLAs are listed in

188 Supplementary Table 3.

189 **(3) Model interpretation**

190 Furthermore, we used Shapley Additive explanations (SHAP) values (Lundberg and Lee, 2017; Lundberg et al., 2019), a

191 model-agnostic technique for interpreting ML models, to explore functional correlations between the variables and forest age

192 (Besnard et al. 2021). SHAP derives the Shapely additive contribution values from coalitional game theory (Kim et al. 2023).

193 By examining the contribution of each input variable to the model's output, SHAP can identify the primary drivers of the

194 model's predictions and provide insights into the underlying causes that influence forest age (Sun et al. 2023). The higher the

195 SHAP value, the larger the contribution of the variable. Here we calculated SHAP value using *shap* package in Python.

196 **2.2.2 LandTrendr disturbance detection approach**

197 LandTrendr was designed to detect and analyse changes in surface features, particularly disturbances and recovery processes,

198 and is commonly applied to multispectral remote sensing imagery from the Landsat satellite series to capture long-term forest

199 disturbances (Du et al., 2022). Using LandTrendr to detect forest age involves the following steps:

200 (1) Time series data transformation. LandTrendr transforms multiple temporal remote-sensing image datasets into a series of

201 indices, such as the NBR.

202 (2) Breakpoint detection. Using the generated time series indices, LandTrendr retraces from the state in 2020 in search of

203 breakpoints in the time series. These breakpoints signify transition points in the time series, which indicate instances of surface

204 disturbance or recovery.

205 (3) Age estimation. By pinpointing breakpoints, the time of occurrence for each breakpoint is established. Forest age estimates

206 for the current location are accomplished by subtracting the breakpoint time from the latest time.

207 LandTrendr was implemented on the GEE platform by using the function of *runLT()* provided by the LT_GEE API (Kennedy

208 et al., 2018). Table 1 lists the main input parameters.

209 **Table 1.** Parameters of LandTrendr used in this study.

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

210

211 **2.2.3 China's forest age prediction**

212 Given the extensive forest coverage in China, it is challenging to handle such large forest area for ML and the LandTrendr

213 algorithm to estimate forest age, even with our vegetation zoning efforts. To enhance the efficiency of forest age estimation

214 and conserve computational resources, we have divided China into 1°×1° grids (see Supplementary Figure 2), limiting ML

215 and LandTrendr algorithms to estimate forest age within each grid. Subsequently, we merge the predictive results from each

216 grid using the Mosaic New Raster tool in ArcGIS Pro 3.0 to obtain nationwide forest age map. Finally, the forest age map

217 estimated through LandTrendr algorithm is applied to update the ML-based results to produce China's forest age data.

2.3 Accuracy assessment

2.3.1 Comparison with field samples

We collected field samples through two sources to validate the generated final forest age map. The first is the forest inventory samples independent of training data. The second source involves validation samples obtained from the literatures. To ensure the samples collected were representative, we excluded samples dated before 2010. As validation metrics, we used the coefficient of determination (R^2), the root mean square error (RMSE), the mean absolute error (MAE), and the mean error (ME). These are given mathematically as

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

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

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

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

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.

2.3.2 Comparison with existing forest age data

To make our forest age map more reliable and comparable, we also downloaded global forest age data product produced by Besnard et al. (2021a), which is the only forest age map that can be publicly accessible covering entire China's forests. Then, we resampled our result to the same resolution as this global map, and compared our resultant forest age map with it by assessing their differences in each cell. Additionally, we also collected estimated average forest ages of China by previous studies, and using these statistical numbers to further validate our estimation.

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

239 (WT) vegetation zone, producing an independent validation R^2 of 0.61 and RMSE of 3.47 years (MAE=9.13, ME=-0.01).

240 CatBoost consistently demonstrated strong performance for the Qinghai-Tibet Plateau alpine vegetation (QT), tropical

241 monsoon forest-rainforest (TM), temperate steppe (TS), temperate desert (TD), and subtropical evergreen broadleaf forest (SE)

242 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

243 choice in the temperate needleleaf-broadleaf mixed forest (TN) vegetation division, yielding an R^2 of 0.63 and an RMSE of

244 4.14 years.

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

Vegetation division	Algorithm	R^2	RMSE	MAE	ME
CT	GradientBoost	0.47	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

246

247 We further analysed the factors influencing the forest age estimation in each vegetation division, and the findings are illustrated

248 in Figure 3. While the prioritization of factors affecting forest age estimation varies across different vegetation divisions,

249 canopy height is unquestionably the predominant factor influencing this estimation. Its absolute value is the highest of the CT,

250 WT, TN, TS, and TM vegetation zones (Figure 3). Moreover, it is among the top three most influential factors in all the

251 remaining vegetation zones. Subsequently, topographical conditions assume prominence, with elevation consistently featuring

252 among the top three factors in the SHAP value across all vegetation divisions. In the TD vegetation division, elevation becomes

253 the most influential factor. Climate factors earn third-tier consideration, particularly in the SE vegetation zone, where the

254 impact of PC2 of the climate factors surpasses that of both canopy height and topographical conditions. In the other vegetation

255 divisions, the influence of climate factors generally falls to the mid-range. In contrast, across all eight vegetation divisions,

256 factors related to soil, such as soil type and soil texture, do not exert a pronounced influence on forest age estimation.

257

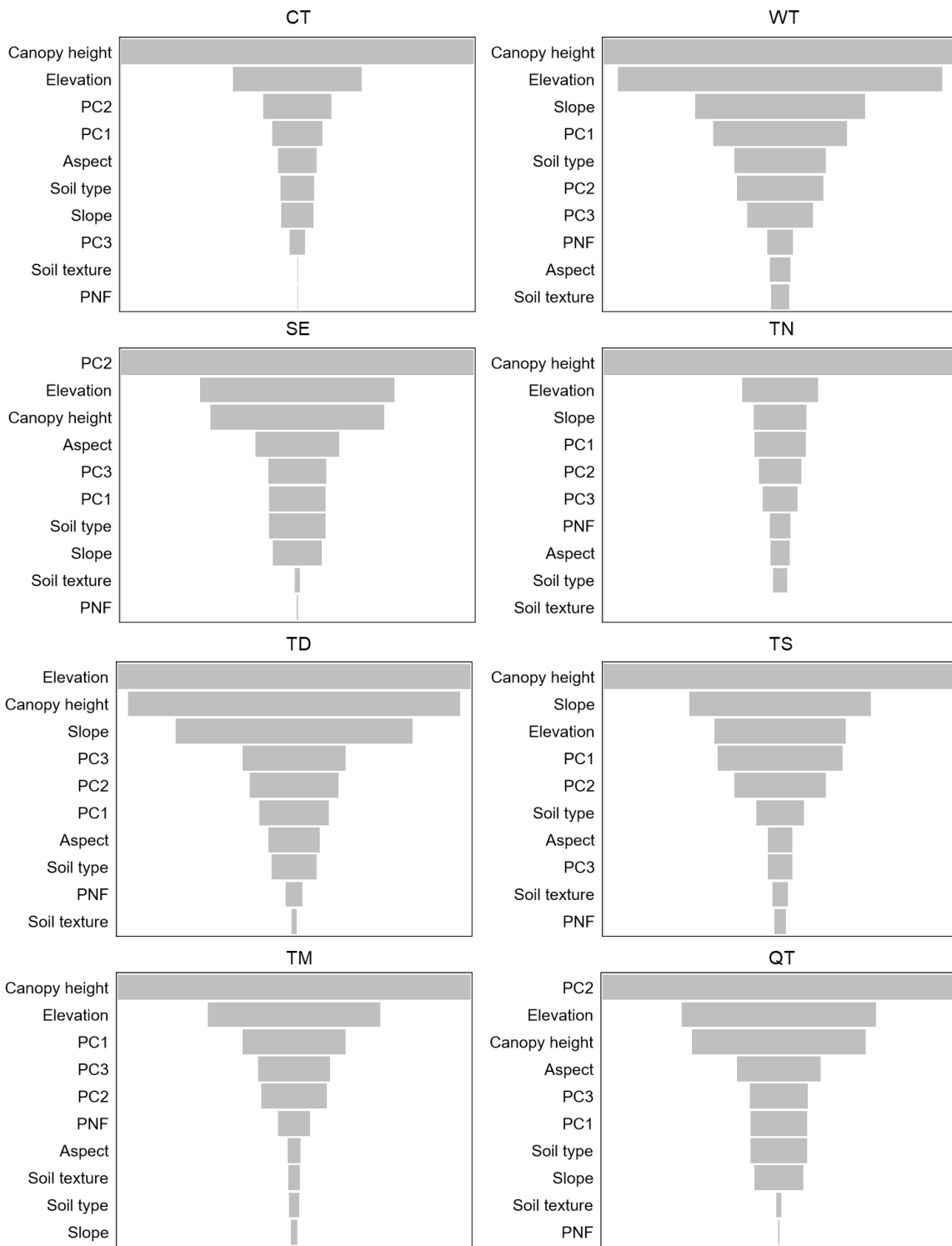



Figure 3. Order of shape values of factors affecting the estimation of forest age in different vegetation zones

3.2 China's forest age map

Based on the optimal MLAs and the LandTrendr change detection algorithm, we have obtained forest age data for China as shown in Figure 4. Figure 4a presents the nationwide distribution of forest age as estimated by MLAs, whereas Figure 4b displays the age distribution from 1985 and 2020 as determined through change detection. The results reveal that reforestation activities from 1985 and 2020 are primarily situated in the southern, southeastern, and northern China, aligning with the findings from Xiao et al (2023). Furthermore, estimates derived from MLAs indicate that old-growth forests are primarily located in the northeast and southwest regions of China.

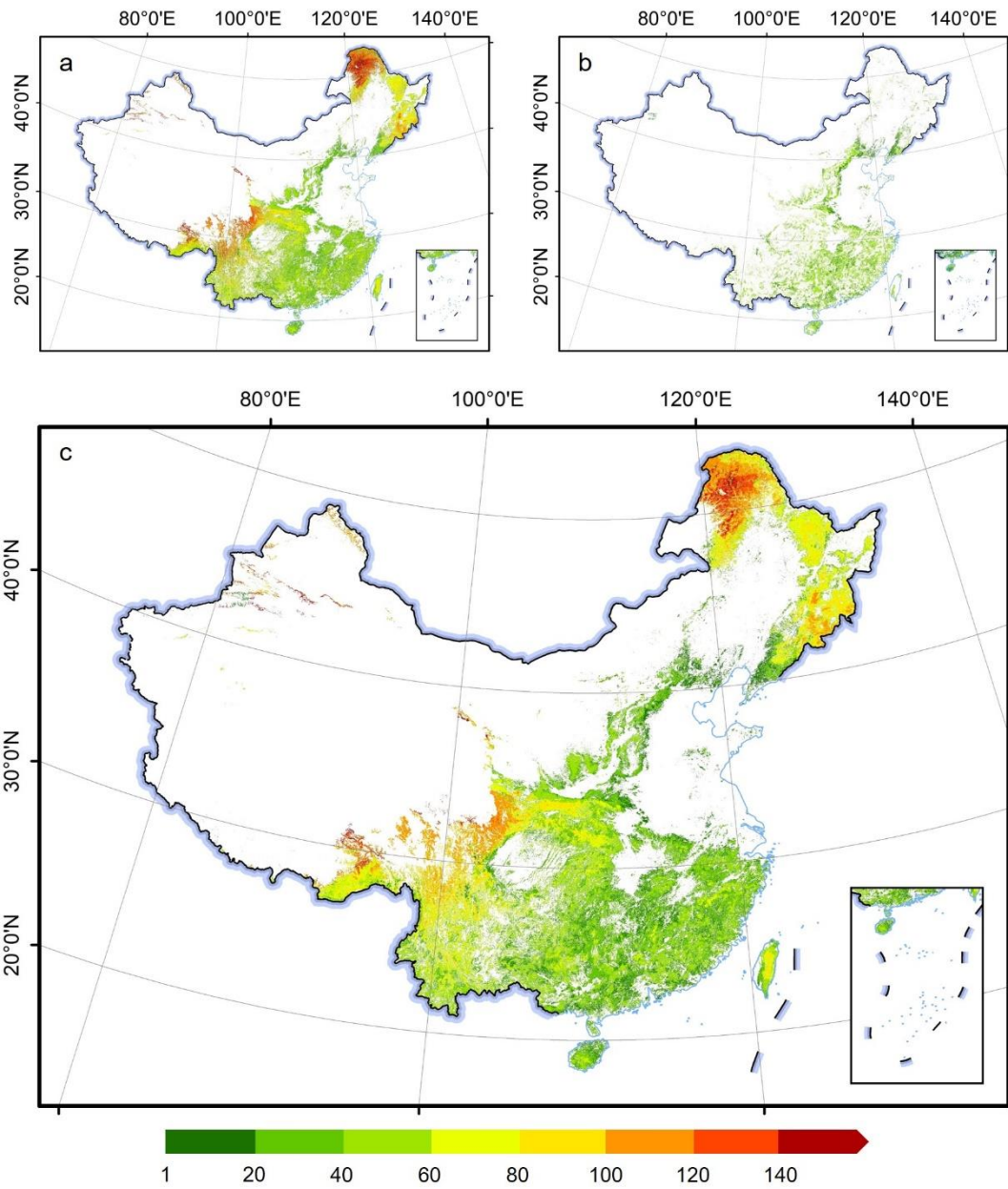
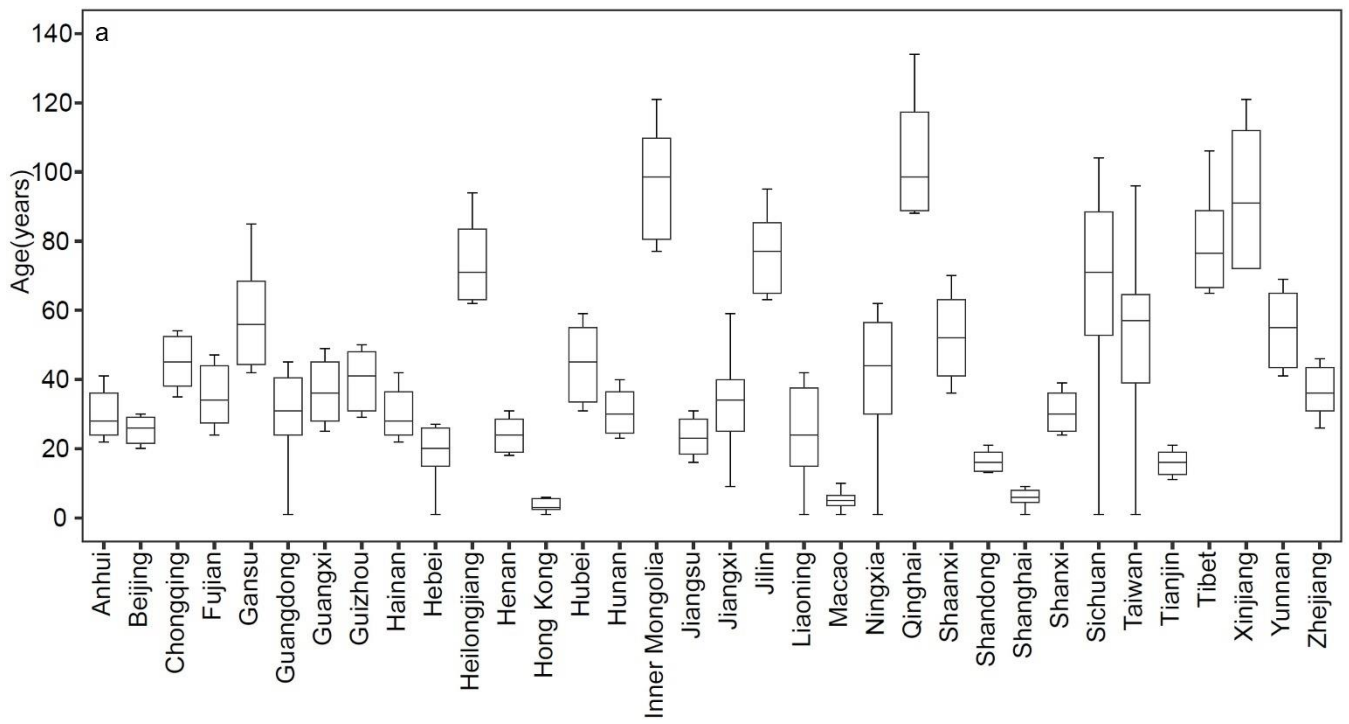
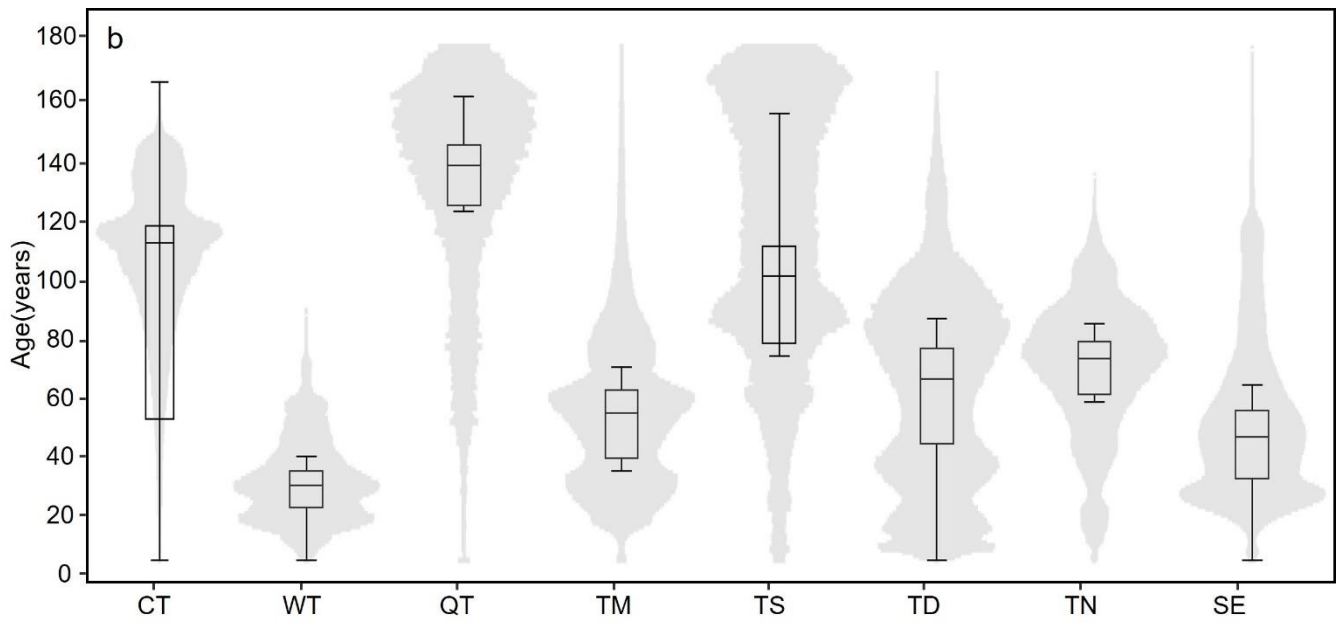
The final forest age map for China obtained in this study is depicted in Figure 4c. Statistically, the mean of the estimated China's forest age is 56.11 years with a standard deviation of 32.67 years. Geographically, forests in northeast and southwest China are relatively older than those in other regions (Figure 4c). At the provincial scale, the average forest age ranges from 3.9 to 116.8 years (Figure 5a, Supplementary Table 6), 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 7). The WT vegetation zones have the youngest forests (28.5 years).

274

Figure 4. Forest age estimated from LandTrendr (a), MLA (b), and final China's forest age distribution (c) with 30 m resolution.

276

277

278 **Figure 5.** (a) Boxplot of China's forest age grouped by provinces (b) Violin plot of the forest age grouped by vegetation divisions.

3.3 Evaluation

3.3.1 Comparison with field 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). We collected 99 field measurements of mean forest stand age after 2010 from published papers (Supplementary Table 8) 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 $R^2 = 0.62$.

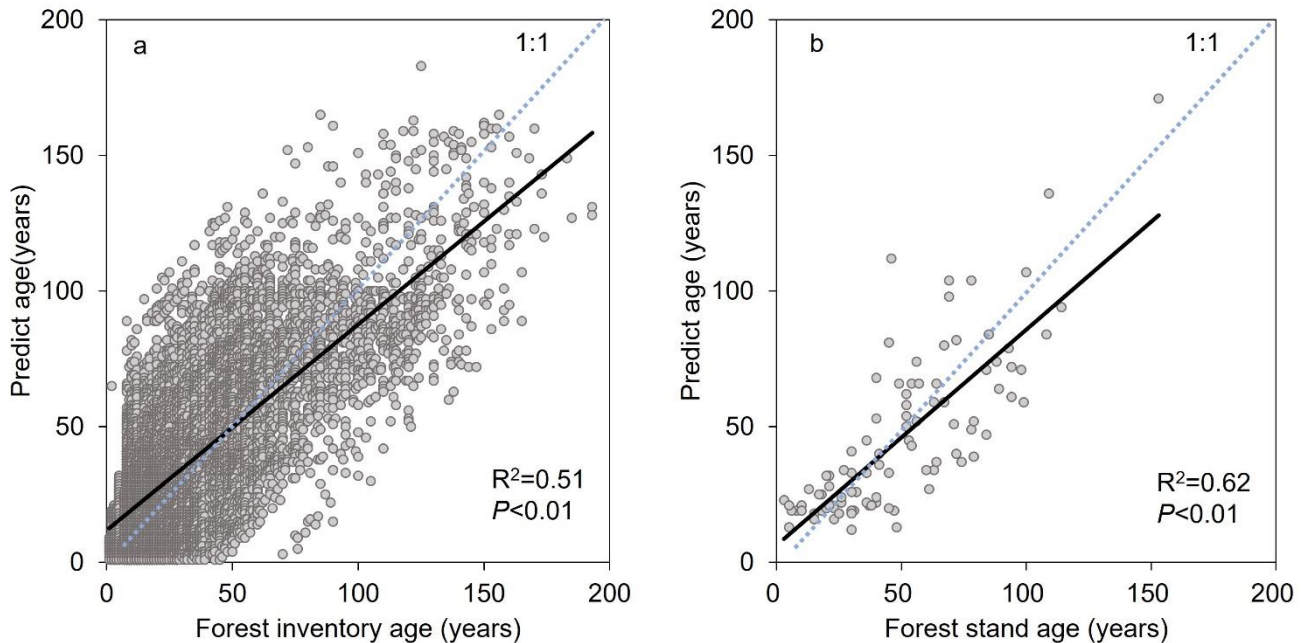
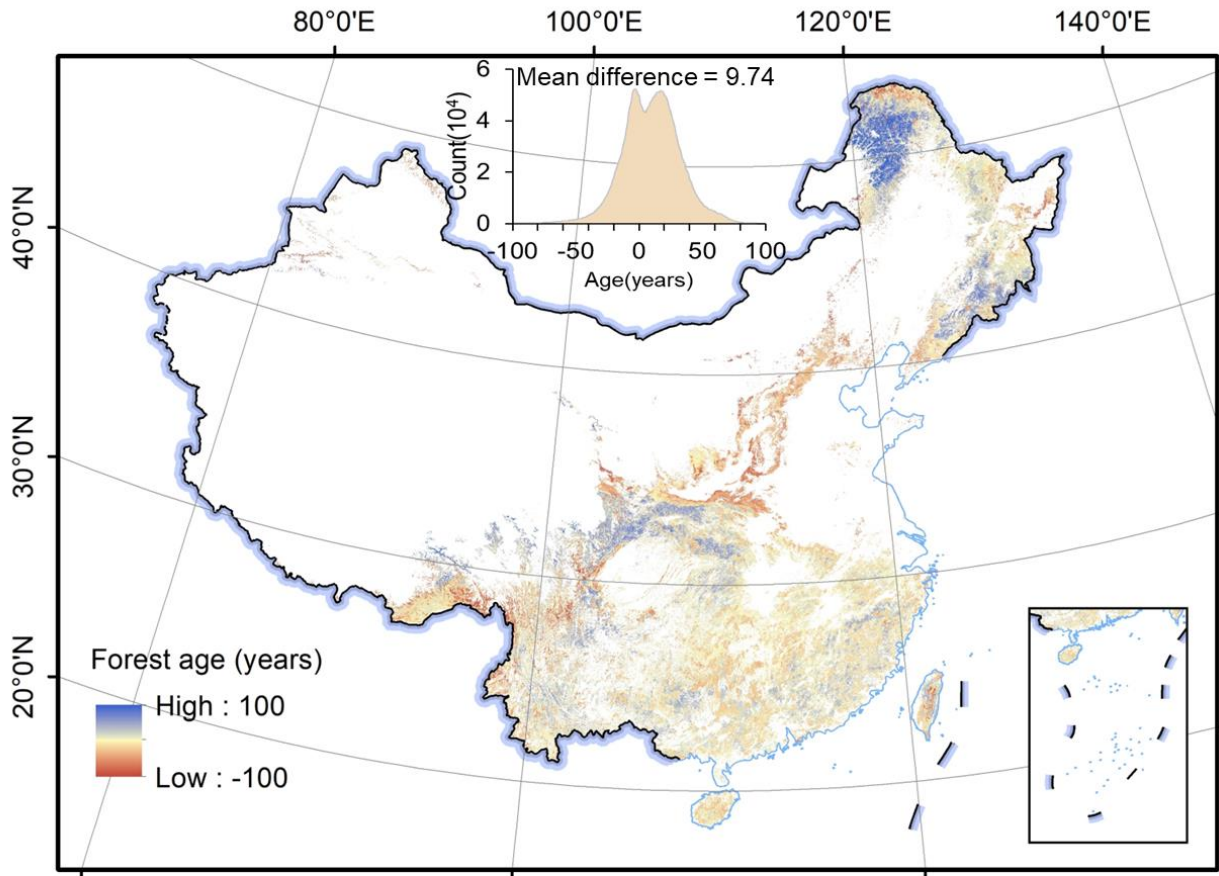



Figure 6. Scatter plots of (a) forest inventory age vs predicted forest age for this study and (b) field measurements of forest stand age collected from published papers vs predicted forest.

3.3.2 Comparison with existing forest age map

Figure 7 shows the difference between our estimation and existing global forest age map, 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 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

298 closely with the average forest age of 56.1 years obtained in this study for the year of 2020, further underscoring the reliability
299 of the forest age mapped in this study.

300
301 **Figure 7.** Comparison with global forest age product. The inset at the top left shows the frequency distribution of differences between the
302 global forest age map and our estimated forest age map.

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

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

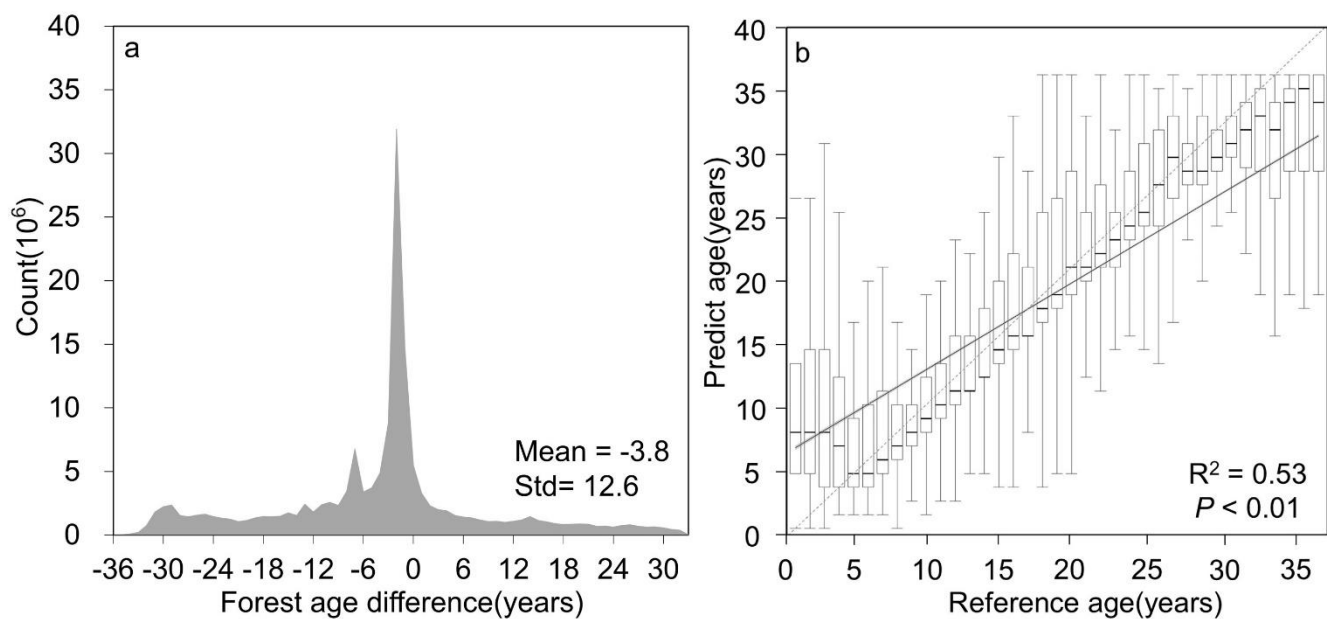
A high-spatial resolution forest age map is an important input for accurately quantifying forest carbon storage and potential. Although several forest age maps for China were generated in the most recent decades, their spatial resolution is coarser, ranging from 1 to 8 km (e.g., Zhang et al., 2014; Zhang 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. (2023a) 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. (2020).

This study combines two methods to estimate forest age across China. We first 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 4). 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 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 5). 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., 2020; Lin et al., 2023). Young trees usually have shorter canopy height and, and as trees age, canopy height gradually increases (Yu et al., 2020). 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., 2021). 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.

The second 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 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).

371

374

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, there is still a slight overestimation of younger forest and an underestimation of older forest compared with validation samples, which is mainly related to dataset and methods used in this study. In terms of dataset, 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 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 7). These disparities, coupled with insights from forest inventory data, highlight the prevalence of older forests (exceeding 100 years) within these regions (Figures 4). The unique challenge posed by estimating the age of such older forests, characterized by sluggish growth rates (Maltman et al., 2023), accentuates the sensitivity to crown height data. Consequently, the uncertainty associated with canopy height data is conspicuously accentuated in these regions. Regarding to methods, we combined MLA and disturbance detection approach to derive forest age, for MLA, overfitting is a common challenge, where a model learns the training data too well and fails to generalize to unseen data (Belgiu and Drăguţ 2016). The results presented in Supplementary Table 5 suggest that the constructed forest age models exhibit a certain degree of overfitting, which can cause some errors for forest age estimation. Addressing the issue of overfitting, data

augmentation and exploring new deep learning algorithms may be promising directions for further investigation. For LandTrendr approach, it is affected by different parameters such as input bands, vegetation parameters (NBR index), climates, vegetation, terrain and atmospheric conditions (Banskota et al. 2014; Hermosilla et al, 2015; Hua et al, 2021; Huang et al, 2023; Yang et al, 2018). China's unprecedented development has led to extensive land cover changes, making it one of the most intensively managed forest regions globally (Tong et al., 2020). This has resulted in significant forest fragmentation, posing challenges in using NBR and other indices for change detection (Li et al 2024), and creating uncertainty in forest age identification. Furthermore, while the LandTrendr algorithm effectively captures sharp disturbances like fires, clearcutting, and reforestation, it falls short in detecting subtle changes such as silviculture and thinning (Huang et al. 2023; Zhu 2017), This limitation may lead to the omission of young trees and an overestimation of forest age.

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 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.

Banskota, A., Kayastha, N., Falkowski, M. J., Wulder, M. A., Froese, R. E., and White, J. C. Forest Monitoring Using Landsat Time Series Data: A Review. Can. J. Remote. Sens., 40(5), 362–384. https://doi.org/10.1080/07038992.2014.987376, 2014.

Belgiu, M., and Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens., 114, 24-31. https://doi.org/10.1016/j.isprsjprs.2016.01.011, 2016.

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, 2021.

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

449   length for regional mapping of lidar-derived forest structure. Remote Sens Environ., 239:111645,
450   https://doi.org/10.1016/j.rse.2020.111645, 2020.

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

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

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

456 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
457   China with 30 m resolution (1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.8354262, 2023b.

458 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
459   resolution imagery and massive amounts of crowdsourced samples. ISPRS J. Photogramm. Remote Sens., 196:356-371,
460   https://doi.org/10.1016/j.isprsjprs.2023.01.005, 2023a.

461 Chen, Dong, Tatiana V. Loboda, Alexander Krylov, and Peter V. Potapov. Mapping stand age dynamics of the Siberian larch
462   forests from recent Landsat observations, Remote Sensing of Environment, 187: 320-31.
463   https://doi.org/10.1016/j.rse.2016.10.033, 2016.

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

467 de Jong, S.M., Shen, Y., de Vries, J., Bijnaar, G., van Maanen, B., Augustinus, P., and Verweij, P. Mapping mangrove
468   dynamics and colonization patterns at the Suriname coast using historic satellite data and the LandTrendr algorithm. Int.
469   J. Appl. Earth Obs. Geoinformation., 97, 102293. https://doi.org/10.1016/j.jag.2020.102293, 2021.

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

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

474 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
475   plantations. Scientific Data., 9:141, https://doi.org/10.1038/s41597-022-01260-2, 2022.

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

479 Fernández-Martínez, M., S. Vicca, I. A. Janssens, S. Luyssaert, M. Campioli, J. Sardans, M. Estiarte, and J. Peñuelas. Spatial
480   variability and controls over biomass stocks, carbon fluxes, and resource-use efficiencies across forest ecosystems. Trees.
481   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.

Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., and Hobart, G. W. Regional detection, characterization, and attribution of annual forest change from 1984 to 2012 using Landsat-derived time-series metrics. Remote Sens Environ., 170, 121–132. https://doi.org/10.1016/j.rse.2015.09.004, 2015.

Hua, J., Chen, G., Yu, L., Ye, Q., Jiao, H., and Luo, X. Improved Mapping of Long-Term Forest Disturbance and Recovery Dynamics in the Subtropical China Using All Available Landsat Time-Series Imagery on Google Earth Engine Platform. IEEE J-STARS., 14, 2754–2768. https://doi.org/10.1109/JSTARS.2021.3058421, 2021.

Huang, C., Goward, S.N., Masek, J.G., Thomas, N., Zhu, Z., and Vogelmann, J.E. An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. Remote Sens Environ., 114, 183-198. https://doi.org/10.1016/j.rse.2009.08.017, 2010.

Huang, Z., Li, X., Du, H., Zou, W., Zhou, G., Mao, F., Fan, W., Xu, Y., Ni, C., Zhang, B., Chen, Q., Chen, J., and Hu, M. An Algorithm of Forest Age Estimation Based on the Forest Disturbance and Recovery Detection. IEEE Trans Geosci Remote Sens., 61, 1-18. https://doi.org/10.1109/TGRS.2023.3322163, 2023.

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.

Kennedy, R.E., Yang, Z., and Cohen, W.B. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr — Temporal segmentation algorithms. Remote Sens Environ., 114, 2897-2910. https://doi.org/10.1016/j.rse.2010.07.010, 2010.

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.

Li, P., Li, H., Si, B., Zhou, T., Zhang, C., and Li, M. Mapping planted forest age using LandTrendr algorithm and Landsat 5–8 on the Loess Plateau, China. Agric For Meteorol, 344, 109795. https://doi.org/10.1016/j.agrformet.2023.109795, 2024.

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, 2023.

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, 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.

Rodman, K.C., Andrus, R.A., Veblen, T.T., and Hart, S.J. Disturbance detection in landsat time series is influenced by tree mortality agent and severity, not by prior disturbance. Remote Sens Environ., 254, 112244. https://doi.org/10.1016/j.rse.2020.112244, 2021.

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, 2020.

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.

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.

Verbesselt, J., Hyndman, R., Newnham, G., and Culvenor, D. Detecting trend and seasonal changes in satellite image time series. Remote Sens Environ., 114, 106-115. https://doi.org/10.1016/j.rse.2009.08.014, 2010a.

Verbesselt, J., Hyndman, R., Zeileis, A., and Culvenor, D. Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. Remote Sens Environ., 114, 2970-2980. https://doi.org/10.1016/j.rse.2010.08.003, 2010b.

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.

Yang, Y., Erskine, P. D., Lechner, A. M., Mulligan, D., Zhang, S., and Wang, Z. Detecting the dynamics of vegetation disturbance and recovery in surface mining area via Landsat imagery and LandTrendr algorithm. J. Clean. Prod., 178, 353–362. https://doi.org/10.1016/j.jclepro.2018.01.050, 2018.

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, 2020.

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.

617   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

618   sensed forest height and observation data. J. Geophys. Res. Biogeosci., 119:1163-1179,

619   https://doi.org/10.1002/2013JG002515, 2014.

620   Zhang, Y., Y. Yao, X. Wang, Y. Liu, and S. Piao. Mapping spatial distribution of forest age in China. Earth Space Sci., 4:108-

621   116, https://doi.org/10.1002/2016EA000177, 2017.

622   Zhang, Z., F. Zhang, L. Wang, A. Lin, and L. Zhao. Biophysical climate impact of forests with different age classes in mid-

623   and high-latitude North America. For. Ecol. Manag., 494:119327. https://doi.org/10.1016/j.foreco.2021.119327, 2021.

624   Zhu, Z. Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications.

625   ISPRS J. Photogramm. Remote Sens., 130, 370-384. https://doi.org/10.1016/j.isprsjprs.2017.06.013, 2017.

626   Zhu, Z., and Woodcock, C.E. Continuous change detection and classification of land cover using all available Landsat data.

627   Remote Sens Environ., 144, 152-171. https://doi.org/10.1016/j.rse.2014.01.011, 2014.

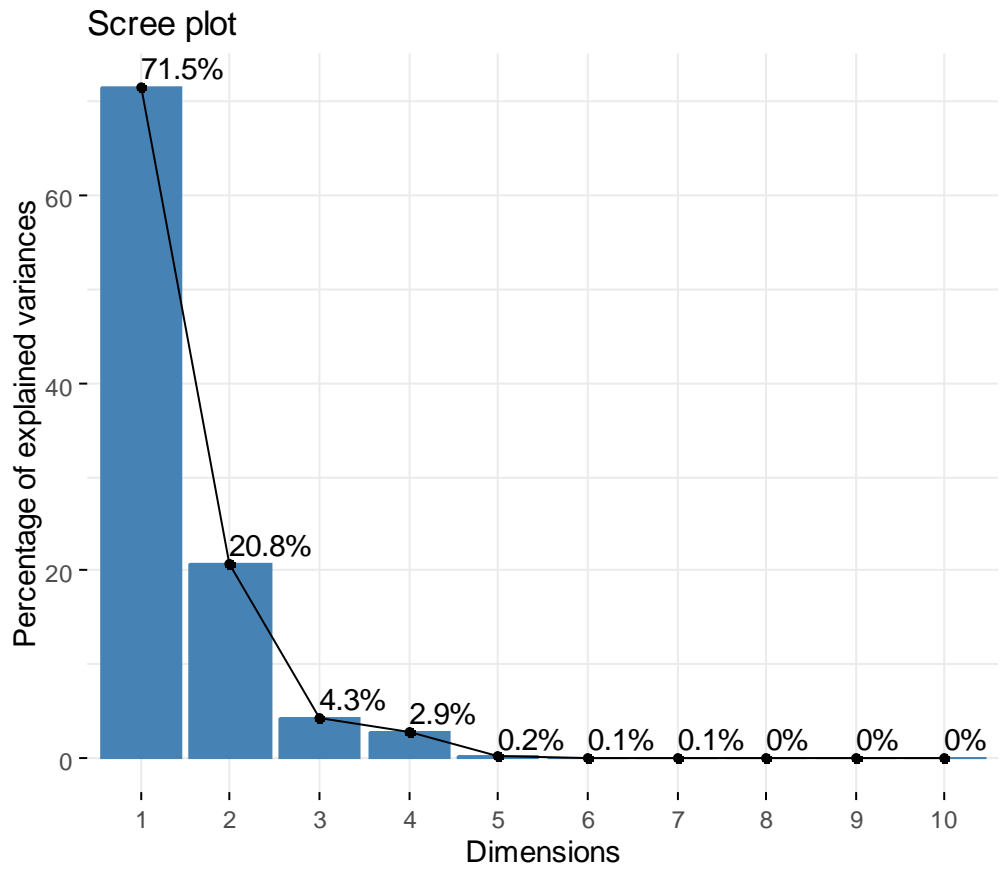
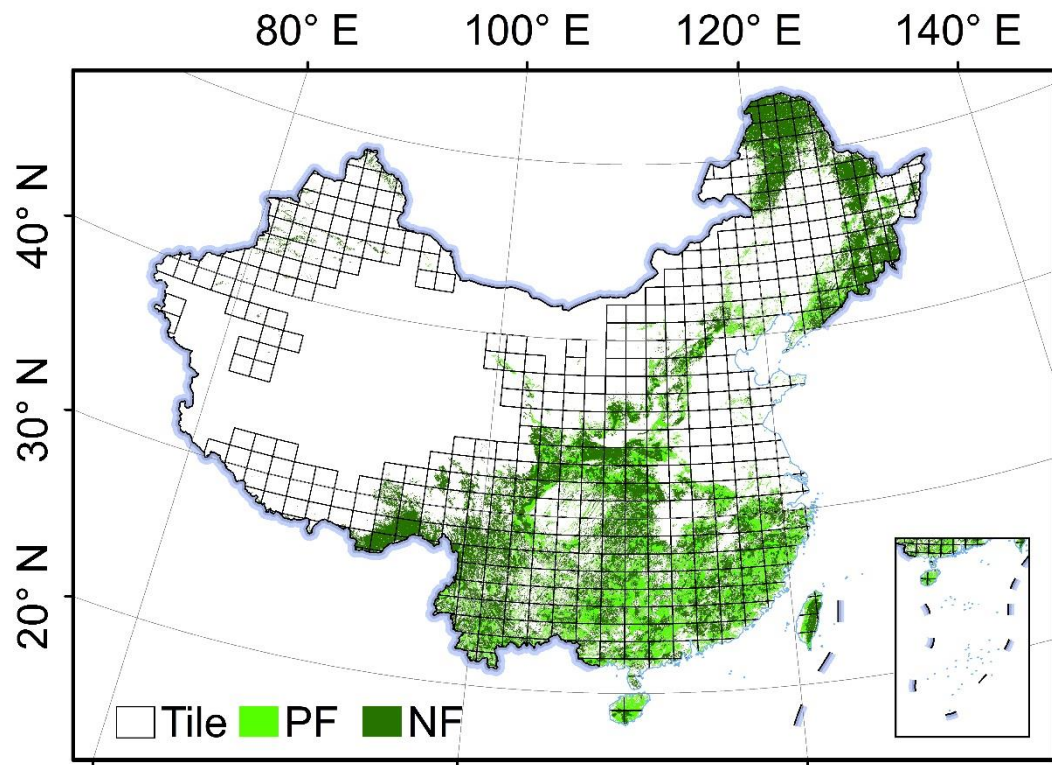
Supplementary

Supplementary Table 1 Principal component analysis of 19 bioclimatic variables

Variables	Describe	Principal Component		
		PC1	**PC2**	**PC3**
BIO1	Annual Mean Temperature	**0.272**	0.112	0.224
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))	**-0.205**	0.177	-0.145
BIO3	Isothermality (BIO2/BIO7) ($\times$100)	-0.129	**0.403**	-0.123
BIO4	Temperature Seasonality (standard deviation $\times$100)	0.039	**-0.476**	0.128
BIO5	Max Temperature of Warmest Month	**0.266**	-0.093	0.261
BIO6	Min Temperature of Coldest Month	**0.262**	0.181	0.200
BIO7	Temperature Annual Range (BIO5-BIO6)	-0.059	**-0.447**	0.041
BIO8	Mean Temperature of Wettest Quarter	0.221	0.029	**0.447**
BIO9	Mean Temperature of Driest Quarter	**0.253**	0.220	0.097
BIO10	Mean Temperature of Warmest Quarter	**0.272**	-0.069	0.264
BIO11	Mean Temperature of Coldest Quarter	0.243	**0.253**	0.168
BIO12	Annual Precipitation	**0.272**	-0.013	-0.271
BIO13	Precipitation of Wettest Month	0.256	0.077	**-0.266**
BIO14	Precipitation of Driest Month	0.255	-0.147	**-0.257**
BIO15	Precipitation Seasonality (Coefficient of Variation)	-0.198	**0.286**	0.066
BIO16	Precipitation of Wettest Quarter	0.255	0.110	**-0.269**
BIO17	Precipitation of Driest Quarter	0.248	-0.165	**-0.275**
BIO18	Precipitation of Warmest Quarter	**0.211**	0.210	-0.162
BIO19	Precipitation of Coldest Quarter	0.250	-0.124	**-0.303**

PC1 is related to BIO1, BIO2, BIO5, BIO6, BIO9, BIO10, BIO12 and BIO18

PC2 is related to BIO3, BIO4, BIO7, BIO11 and BIO15

PC3 is related to BIO8, BIO13, BIO14, BIO16, BIO17 and BIO19.

Supplementary Figure1. Scree plot of 19 bioclimatic variables

Supplementary Figure 2. 1°×1° grids for China's forest region. PF: planted forest, NF: Natural forest.

Supplementary Table 2. The thirteen alternative machine learning algorithms and their accuracies(R^2)

Model	Validation								Test							
	CT	WT	QT	TM	TS	TD	TN	SE	CT	WT	QT	TM	TS	TD	TN	SE
DecisionTree	-0.1443	0.2008	0.2540	0.6540	0.5230	0.4843	0.2743	0.3521	-0.0116	0.2232	0.0815	0.7207	0.5480	0.5409	0.2882	0.3586
ExtraTree	-0.1450	0.1879	0.1295	0.6354	0.4200	0.4582	0.2306	0.3115	-0.2244	0.2527	0.1297	0.6657	0.4828	0.5660	0.2398	0.3526
GaussianProcess	-2.9984	-1.5756	-2.2477	-13.0362	-1.2202	-2.3692	-2.7937	-1.2905	-2.8423	-1.5995	-2.5428	-11.2968	-1.1853	-2.1186	-2.8901	-1.2705
XGB	0.3311	0.5718	0.5109	0.7942	0.7106	0.7077	0.5845	0.6573	0.3402	0.5857	0.3609	0.8241	0.7272	0.7586	0.6031	0.6690
RandomForest	0.4124	0.6017	0.5546	0.8113	0.7343	0.7067	0.6236	0.6609	0.4262	0.6121	0.5363	0.8414	0.7641	0.7724	0.6272	0.6749
Bagging	0.3646	0.5661	0.5268	0.7961	0.7043	0.6699	0.5859	0.6309	0.3674	0.5791	0.4387	0.8287	0.7315	0.7487	0.6055	0.6457
HistGradientBoost	0.3935	0.6039	0.5093	0.8051	0.7325	0.6931	0.6234	0.6685	0.4092	0.6113	0.4526	0.8341	0.7572	0.7825	0.6303	0.6885
LGBM	0.3986	0.6052	0.5202	0.8083	0.7234	0.6962	0.6262	0.6726	0.4203	0.6112	0.4552	0.8362	0.7513	0.7845	0.6341	0.6891
GradientBoosting	0.4208	0.5950	0.5484	0.8095	0.7260	0.7007	0.6223	0.6605	0.4543	0.5923	0.4930	0.8433	0.7435	0.7705	0.6228	0.6796
MLP	0.3990	0.5198	0.1415	0.5954	0.7076	0.6016	0.5953	0.6659	0.4311	0.5120	0.2354	0.7695	0.7170	0.6685	0.5900	0.6685
AdaBoost	0.1650	0.4622	0.4775	0.6436	0.5348	0.5867	0.5443	0.3152	0.1700	0.4808	0.4589	0.7705	0.6095	0.5980	0.5434	0.3559
KNeighbors	-0.0304	0.3386	0.3247	0.5851	0.4612	0.5574	0.0958	0.3883	0.0013	0.3779	0.1892	0.6067	0.5253	0.6302	0.1312	0.4284
CatBoost	0.4263	0.6109	0.5814	0.8162	0.7409	0.7322	0.6271	0.6749	0.4501	0.6203	0.5022	0.8444	0.7602	0.7922	0.6275	0.6924

Supplementary Table 3. The hyperparameters and their range for different models.

Algorithm	Python Package	Hyperparameter Range
RF	sklearn.ensemble.RandomForestRegressor	max_depth:[3,18] n_estimators:[5000, 8000] max_features:['auto', 'sqrt', 'log2'] min_samples_split:[2, 10] min_samples_leaf:[2, 10] random_state: 2023
GBDT	sklearn.ensemble.GradientBoostingRegressor	max_depth:[2, 10] learning_rate: [0.001,0.005,0.01,0.05,0.1] n_estimators:[4000, 5000] subsample:[0.7, 0.9] max_features: ['auto', 'sqrt', 'log2'] min_samples_split:[2, 10] random_state:2023
HistGradientBoost	sklearn.ensemble.HistGradientBoostingRegressor	max_depth:[2, 10] learning_rate: [0.001,0.02,0.03,0.005,0.01,0.05,0.1] max_leaf_nodes:[30, 40] min_samples_leaf:[15, 25] random_state:2023
LightGBM	lightgbm	reg_alpha: [0.001, 10.0] reg_lambda: [0.001, 10.0] num_leaves: [11, 333] min_child_samples: [5, 100] max_depth: [3, 20] learning_rate: [0.001,0.005,0.01,0.05,0.1] colsample_bytree: [0.1, 0.5] n_estimators: [7000, 8000] cat_smooth: [10, 100] cat_l2: [1, 20] min_data_per_group: [50, 200] cat_feature: [10, 60] n_jobs: -1 random_state: 2023
CatBoost	catboost	depth: [3, 10] learning_rate: [0.001,0.005,0.01,0.05,0.1] iterations: [5000, 9000] max_bin: [200, 400] min_data_in_leaf: [1, 30] l2_leaf_reg: [0.0001, 1.0, log=True] subsample: [0.6, 0.9] random_state: 2023

Supplementary Table 4. The optimal hyperparameter parameter values for different MLAs.

Vegetation division	Algorithm	Hyperparameter values
CT	RF	(max_depth =8,n_estimators=6589,max_features= 'auto',min_samples_leaf=9,min_samples_split =8,random_state=2023)
	GBDT	{'max_depth': 4, 'learning_rate': 0.001, 'n_estimators': 4909, 'subsample': 0.7234085712326702, 'max_features': 'auto', 'min_samples_split':10 ,'random_state': 2023}
	CatBoost	(depth=10,learning_rate=0.1,iterations=86,max_bin=320,min_data_in_leaf=27,l2_leaf_reg=0.17934206956587195,subsample=0.6773452775007673,random_seed=2023)
WT	RF	(max_depth =19,n_estimators=348,max_features= 'sqrt',min_samples_leaf=1,min_samples_split =3,random_state=2023)
	LGBM	{'reg_alpha': 4.188760632650688, 'reg_lambda': 4.255499587500175, 'num_leaves': 75, 'min_child_samples': 7, 'max_depth': 19, 'learning_rate': 0.001, 'colsample_bytree': 0.4928730464443524, 'n_estimators': 7117, 'cat_smooth': 84, 'cat_l2': 15, 'min_data_per_group': 193, 'cat_feature': 28, 'random_state': 2023}
	CatBoost	(depth=12,learning_rate=0.05,iterations=133,max_bin=314,min_data_in_leaf=8,l2_leaf_reg=0.0021616691540516635,subsample=0.827218563526197,random_seed=2023)
QT	RF	(max_depth =9,n_estimators=100,max_features= 'auto',min_samples_leaf=1,min_samples_split =2,random_state=2023)
	GBDT	{'max_depth': 5, 'learning_rate': 0.001, 'n_estimators': 4873, 'subsample': 0.6338013854778914, 'max_features': 'sqrt', 'min_samples_split':8 ,'random_state': 2023}
	CatBoost	(depth=7,learning_rate=0.1,iterations=90,max_bin=337,min_data_in_leaf=4,l2_leaf_reg=0.0008155227484111563,subsample=0.7808941610379249,random_seed=2023)
TM	HistGradientBoost	(learning_rate=0.05, max_leaf_nodes=33, max_depth=5, min_samples_leaf=21, l2_regularization=0.0001, max_bins=200, early_stopping='auto', random_state=2023)
	GBDT	{'max_depth': 5, 'learning_rate': 0.001, 'n_estimators': 4970, 'subsample': 0.7323582473497865, 'max_features': 'auto', 'min_samples_split':9 ,'random_state': 2023}
	CatBoost	(depth=5,learning_rate=0.1,iterations=228,max_bin=298,min_data_in_leaf=14,l2_leaf_reg=0.912715671115768,subsample=0.7691332886798857,random_seed=2023)
TS	RF	(max_depth =12,n_estimators=105,max_features= 'auto',min_samples_leaf=2,min_samples_split =8,random_state=2023)
	LightGBM	{'reg_alpha': 3.0022329902119083, 'reg_lambda': 6.129604703602383, 'num_leaves': 69, 'min_child_samples': 38, 'max_depth': 15, 'learning_rate': 0.001, 'colsample_bytree': 0.48372749547013316, 'n_estimators': 7571, 'cat_smooth': 82, 'cat_l2': 5, 'min_data_per_group': 128, 'cat_feature': 41, 'random_state': 2023}
	CatBoost	(depth=7,learning_rate=0.005,iterations=4164,max_bin=215,min_data_in_leaf=24,l2_leaf_reg=0.00017571003237103587,subsample=0.8817081947911567,random_seed=2023)
TD	HistGradientBoost	(learning_rate=0.1, max_leaf_nodes=39, max_depth=4, min_samples_leaf=22, l2_regularization=0.0001, max_bins=200,

		early_stopping='auto', random_state=2023)
	LightGBM	{'reg_alpha': 2.1333544399270994, 'reg_lambda': 7.980678166407649, 'num_leaves': 217, 'min_child_samples': 5, 'max_depth': 5, 'learning_rate': 0.003, 'colsample_bytree': 0.43984300935044063, 'n_estimators': 7992, 'cat_smooth': 76, 'cat_l2': 7, 'min_data_per_group': 187, 'cat_feature': 47, 'random_state': 2023}
	CatBoostRegressor	(depth=5,learning_rate=0.01,iterations=5203,max_bin=246,min_data_in_leaf=6,l2_leaf_reg=0.00045191356462636874,subsample =0.6391293474573634,random_seed=2023)
TN	HistGradientBoost	(learning_rate=0.05, max_leaf_nodes=38, max_depth=6, min_samples_leaf=20, l2_regularization=0.0001, max_bins=200, early_stopping='auto', random_state=2023)
	LightGBM	{'reg_alpha': 3.0022329902119083, 'reg_lambda': 6.129604703602383, 'num_leaves': 69, 'min_child_samples': 38, 'max_depth': 15, 'learning_rate': 0.001, 'colsample_bytree': 0.48372749547013316, 'n_estimators': 7571, 'cat_smooth': 82, 'cat_l2': 5, 'min_data_per_group': 128, 'cat_feature': 41, 'random_state': 2023}
	CatBoost	(depth=10,learning_rate=0.1,iterations=155,max_bin=319,min_data_in_leaf=1,l2_leaf_reg=0.33907394509650335,subsample=0.76 82844712570389,random_seed=2023)
SE	LightGBM	{'reg_alpha': 6.116715128459515, 'reg_lambda': 5.231647634428009, 'num_leaves': 18, 'min_child_samples': 73, 'max_depth': 8, 'learning_rate': 0.002, 'colsample_bytree': 0.4726653436124117, 'n_estimators': 7288, 'cat_smooth': 67, 'cat_l2': 6, 'min_data_per_group': 79, 'cat_feature': 36, 'random_state': 2023}
	MLP	{'hidden_layer_sizes': (200, 200, 200), 'activation': 'relu', 'solver': 'adam', 'alpha': 0.0001, 'batch_size': 'auto', 'learning_rate': 'constant', 'learning_rate_init': 0.001, 'max_iter': 155, 'random_state': 2023}
	CatBoost	(depth=7,learning_rate=0.005,iterations=6248,max_bin=383,min_data_in_leaf=20,l2_leaf_reg=0.6887500276693759,subsample=0. 7127716543175433,random_seed=2023)

Supplementary Table 5. The training results of different MLAs

Vegetation division	Algorithm	Train		Validation	
		R2	RMSE	R2	RMSE
CT	RF	0.6102	4.6402	0.4394	25.6178
	GBDT	0.6463	4.5287	0.4398	25.6177
	CatBoost	0.7573	4.1219	0.4307	25.8266
WT	RF	0.9286	2.2631	0.6108	11.9231
	LGBM	0.8303	2.8096	0.6112	11.9211
	CatBoost	0.8309	2.8073	0.6085	11.9593
QT	RF	0.9326	4.7895	0.5135	56.7438
	GBDT	0.8846	5.4791	0.5430	55.6378
	CatBoost	0.9637	4.1029	0.5648	54.1309
TM	HistGradientBoost	0.8729	1.9054	0.8129	4.3884
	GBDT	0.9059	1.7673	0.8139	4.3779
	CatBoost	0.8969	1.8083	0.8202	4.3006
TS	RF	0.9068	3.2872	0.7365	18.0381
	LightGBM	0.8685	3.5823	0.7456	17.7228
	CatBoost	0.9375	2.9747	0.7493	17.5945
TD	HistGradientBoost	0.8555	4.8085	0.7125	32.2791
	LightGBM	0.9485	3.7154	0.7065	32.6168
	CatBoostRegressor	0.9834	2.8009	0.7235	31.5870
TN	HistGradientBoost	0.6581	4.0517	0.6246	17.1609
	LightGBM	0.8080	3.5072	0.6292	17.0570
	CatBoost	0.7633	3.6956	0.6198	17.2713
SE	LightGBM	0.7273	3.3467	0.6722	12.2440
	MLP	0.7013	3.4237	0.6571	12.5179
	CatBoost	0.7899	3.1354	0.6774	12.1481

Supplementary Table 6. Mean and standard deviation of forest age in provinces

Province	Mean	S.D.
Anhui	30.0438	13.18046
Beijing	23.72932	6.941676
Chongqing	44.21707	14.81873
Fujian	34.71088	14.08378
Gansu	62.91986	28.90555
Guangdong	32.2954	14.03617
Guangxi	35.98391	14.40237
Guizhou	38.97128	14.09774
Hainan	31.67802	14.70032
Hebei	20.04016	9.452565
Heilongjiang	76.63183	26.88827
Henan	26.42557	13.65384
Hong Kong	2.905065	3.205983
Hubei	44.2699	17.3144
Hunan	30.98115	11.84794
Inner Mongolia	95.71522	35.21181

Jiangsu		22.35844	11.39468
Jiangxi		34.05567	13.07562
Jilin		75.20958	26.22582
Liaoning		29.84999	21.79875
Macao		3.805029	2.552609
Ningxia		44.01774	22.64126
Qinghai		115.7855	47.51708
Shaanxi		52.42966	20.90854
Shandong		15.81338	6.811547
Shanghai		5.695209	3.581797
Shanxi		30.59349	11.17819
Sichuan		75.09604	32.48283
Taiwan		53.18191	20.35996
Tianjin		15.6741	6.724648
Tibet		83.86414	30.52417
Xinjiang		103.28	50.83277
Yunnan		54.69701	21.25958
Zhejiang		35.3726	13.07106

Supplementary Table 7. Mean and standard deviation of forest age in eight vegetation zones

Vegetation zone	Mean	S.D.	
CT		106.6109	24.25592
WT		27.47919	13.87375
QT		136.9514	36.23872
TM		52.04598	26.27037
TS		106.0278	56.54841
TD		59.32036	33.32852
TN		67.25398	22.54142
SE		48.23603	26.65255

Supplementary Table 8. Field measurements of forest age collected from published papers

ID	Longitude	Latitude	Year	Mean age (2020)	Reference
1	23.2450	113.4210	2020	6	Chen et al., (2022)
2	23.2260	113.3930	2020	10	Chen et al., (2022)
3	23.2560	113.4190	2020	15	Chen et al., (2022)
4	23.2120	113.3940	2020	20	Chen et al., (2022)
5	23.2550	113.3810	2020	30	Chen et al., (2022)
6	24.5200	114.4300	2010-2011	40	Di Y et al. (2012)
7	26.8814	117.9353	2017	10	Feng et al., (2021)
8	30.0800	110.5600	2010	40	Hu et al., (2012)
9	31.4300	110.3500	2010	45	Hu et al., (2012)
10	23.4800	100.5300	2013	67	Li et al. (2015)
11	23.0502	109.3289	2021	3	Li et al., (2021)
12	23.0535	109.3329	2021	8	Li et al., (2021)
13	23.1118	109.2420	2021	18	Li et al., (2021)

14	23.0533	109.1602	2021	21	Li et al., (2021)
15	25.3300	114.5700	2012	53	Qiu et al., (2020)
16	25.3300	114.5700	2012	53	Qiu et al., (2020)
17	25.3300	114.5700	2012	36	Qiu et al., (2020)
18	22.0483	110.4658	2020	5	Song et al., (2021)
19	21.9192	110.5008	2020	15	Song et al., (2021)
20	22.0222	110.5003	2020	5	Song et al., (2021)
21	36.4200	109.5300	2013	79	Sun et al., (2020)
22	23.7700	101.2700	2013	22	Tong et al. (2013)
23	23.7700	101.2700	2013	37	Tong et al. (2013)
24	23.9000	101.2700	2013	52	Tong et al. (2013)
25	28.6017	104.5672	2011	26	Wu et al. (2023)
26	28.6093	104.5769	2015	13	Wu et al. (2023)
27	22.0263	106.9073	2013	30	Wu et al. (2023)
28	22.0243	106.9102	2013	30	Wu et al. (2023)
29	22.0264	106.9132	2013	30	Wu et al. (2023)
30	22.8667	108.1667	2012	25	Wu et al. (2023)
31	21.9667	109.2833	2012	30	Wu et al. (2023)
32	26.6993	109.6076	2010	23	Wu et al. (2023)
33	26.7003	109.6077	2010	23	Wu et al. (2023)
34	24.7633	109.8933	2012	21	Wu et al. (2023)
35	34.0909	110.4029	2012	25	Wu et al. (2023)
36	30.9189	110.6969	2015	30	Wu et al. (2023)
37	27.2938	112.8481	2013	18	Wu et al. (2023)
38	27.2943	112.8486	2013	17	Wu et al. (2023)
39	27.3545	113.3865	2013	21	Wu et al. (2023)
40	26.8139	117.5247	2014	27	Wu et al. (2023)
41	26.8072	117.5408	2014	20	Wu et al. (2023)
42	52.9783	122.5456	2010	36	Wu et al. (2023)
43	25.9500	108.3700	2017	36	Zhou et al., 2018
44	18.7400	108.8500	2011-2016	59	Zhu et al. (2017)
45	24.4500	113.6800	2011-2016	52	Zhu et al. (2017)
46	24.8900	113.0300	2011-2016	49	Zhu et al. (2017)
47	24.9600	112.9600	2011-2016	98	Zhu et al. (2017)
48	25.3200	114.1500	2011-2016	71	Zhu et al. (2017)
49	25.7100	100.0400	2011-2016	56	Zhu et al. (2017)
50	25.7200	100.0500	2011-2016	85	Zhu et al. (2017)
51	25.7200	100.0500	2011-2016	108	Zhu et al. (2017)
52	25.8300	111.6500	2011-2016	31	Zhu et al. (2017)
53	25.8000	112.8700	2011-2016	26	Zhu et al. (2017)
54	25.6800	118.1900	2011-2016	40	Zhu et al. (2017)
55	26.1700	106.6500	2011-2016	45	Zhu et al. (2017)
56	26.5800	116.3400	2011-2016	61	Zhu et al. (2017)
57	27.0200	111.3400	2011-2016	40	Zhu et al. (2017)
58	27.0100	117.0700	2011-2016	64	Zhu et al. (2017)
59	27.8400	98.6800	2011-2016	41	Zhu et al. (2017)
60	27.9000	117.3600	2011-2016	84	Zhu et al. (2017)
61	27.9200	117.3500	2011-2016	93	Zhu et al. (2017)
62	27.7300	117.6300	2011-2016	74	Zhu et al. (2017)

63	27.6000	117.4600	2011-2016	54	Zhu et al. (2017)
64	28.2900	99.1600	2011-2016	114	Zhu et al. (2017)
65	28.1100	117.0000	2011-2016	63	Zhu et al. (2017)
66	29.1000	115.5700	2011-2016	56	Zhu et al. (2017)
67	28.9300	118.0600	2011-2016	88	Zhu et al. (2017)
68	29.2400	118.1000	2011-2016	78	Zhu et al. (2017)
69	29.2200	119.5200	2011-2016	60	Zhu et al. (2017)
70	30.0300	102.8300	2011-2016	54	Zhu et al. (2017)
71	29.7700	110.0900	2011-2016	94	Zhu et al. (2017)
72	31.1700	102.9900	2011-2016	69	Zhu et al. (2017)
73	31.3500	102.8300	2011-2016	69	Zhu et al. (2017)
74	31.5100	110.4300	2011-2016	94	Zhu et al. (2017)
75	32.2200	102.6100	2011-2016	78	Zhu et al. (2017)
76	34.0500	107.7000	2011-2016	84	Zhu et al. (2017)
77	34.0700	107.6900	2011-2016	64	Zhu et al. (2017)
78	33.9300	112.1600	2011-2016	38	Zhu et al. (2017)
79	34.4800	110.5700	2011-2016	32	Zhu et al. (2017)
80	36.3100	118.0500	2011-2016	47	Zhu et al. (2017)
81	37.2600	122.4600	2011-2016	48	Zhu et al. (2017)
82	37.8600	111.4600	2011-2016	52	Zhu et al. (2017)
83	39.9900	115.0200	2011-2016	72	Zhu et al. (2017)
84	39.9500	115.4300	2011-2016	79	Zhu et al. (2017)
85	39.9600	115.4300	2011-2016	79	Zhu et al. (2017)
86	40.5700	115.7700	2011-2016	41	Zhu et al. (2017)
87	40.3900	117.4600	2011-2016	30	Zhu et al. (2017)
88	40.3100	117.5700	2011-2016	45	Zhu et al. (2017)
89	41.2700	125.4100	2011-2016	52	Zhu et al. (2017)
90	42.2000	127.5100	2011-2016	67	Zhu et al. (2017)
91	42.8100	127.9100	2011-2016	99	Zhu et al. (2017)
92	44.4900	81.2600	2011-2016	153	Zhu et al. (2017)
93	44.7800	129.2400	2011-2016	72	Zhu et al. (2017)
94	45.3800	127.6000	2011-2016	63	Zhu et al. (2017)
95	46.4700	131.1100	2011-2016	89	Zhu et al. (2017)
96	47.6700	128.0700	2011-2016	57	Zhu et al. (2017)
97	47.9900	88.2600	2011-2016	46	Zhu et al. (2017)
98	51.7800	123.0200	2011-2016	109	Zhu et al. (2017)
99	52.8200	123.2400	2011-2016	100	Zhu et al. (2017)