



A reference sample dataset of stable, disturbed and forestated forests over China from 1986 to near present

Rong Shang^{1,8}, Shengwei Xu^{1,8}, Ziyi Yang^{1,*}, Xudong Lin¹, Lingyun Fan¹, Keyan Fang¹, Mingzhu Xu¹, Anxin Ding¹, Yulin Yan², Yunjian Liang¹, Chenningya Song¹, Wenjie Chen¹, Jiaorong Qian¹, Annan Zhang¹, Jing Heng¹, Siyu Chen¹, Xiao Zhang³, Liangyun Liu⁴, Wang Li⁵, Licong Dai⁶, Zhongmin Hu⁶, Jing M. Chen^{1,7,*}

¹Key Laboratory of Humid Subtropical Eco-Geographical Process of Ministry of Education, School of Geographical Sciences, Fujian Normal University, Fuzhou, 350117, China

²Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul, 08826, Republic of Korea

³International Research Center of Big Data for Sustainable Development Goals, Beijing, 100094, China

⁴School of Remote Sensing, Beijing Normal University, Beijing, 100875, China

⁵State Key Laboratory of Remote Sensing and Digital Earth, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, 100101, China

⁶Hainan Baoting Tropical Forest Ecosystem Observation and Research Station, School of Ecology, Hainan University, Haikou 570228, China

⁷Department of Geography and Planning, University of Toronto, Ontario, ON M5S 3G3, Canada

⁸These authors contributed equally: Rong Shang and Shengwei Xu

Correspondence to: Ziyi Yang (zyzy372200@gmail.com) and Jing M. Chen (jing.chen@utoronto.ca)

Abstract. Forest disturbances and forestation profoundly affect terrestrial ecosystem dynamics and carbon cycling. However, the lack of long-term national reference samples has greatly restricted the development and application of forest disturbance and forestation monitoring across China. In this study, we constructed a 30 m resolution Reference sample dataset of Stable, Disturbed and Forestated forests (RSDF), covering China from 1986 to 2025 with 12,991 geographically representative samples obtained via stratified random sampling. Across the full dataset, 79.61% of the samples represent stable forest during 1986–2025, 15.99% record a single disturbance or forestation event, and 4.40% undergo two or more events. All samples were annotated through hierarchical visual interpretation by integrating Landsat time-series data, multi-source high-resolution satellite imagery, and available forest inventory records. A strict multi-expert blind review framework combined with a sequential consensus mechanism was further implemented for multi-stage quality control. All samples with forest disturbance or forestation were cross-validated using standardized visual diagnostic evidence, including time-series trajectories of six spectral bands, two vegetation indices, and matched pre- and post-change RGB composite imagery. Spatial-temporal pattern analysis revealed distinct regional disparities in forest change activities: forest disturbances were most frequent in East and South China and relatively weak in Northwest, North and Northeast China; harvest was the dominant forest disturbance type nationwide, while fire, other disturbance and forestation events accounted for a small proportion. Quality assessments showed over 95% high-confidence samples, with a mean inter-expert pairwise agreement of 0.83 ± 0.07 and an average Cohen's Kappa coefficient of 0.65 ± 0.13 , demonstrating high reliability and interpretative consistency of the RSDF dataset. This 40-year nationwide sample dataset can be widely applied to algorithm calibration, model training, and product validation. It also



provides a credible benchmark for large-scale forest dynamic monitoring and carbon budget assessment in China. The dataset is publicly available at <https://doi.org/10.6084/m9.figshare.33313359> (Shang et al., 2026).

1 Introduction

Forest disturbances, defined as any discrete events that cause tree cover loss, including harvest, fire, windthrow, insects, and diseases, substantially reshape terrestrial ecosystem structures and regulate forest dynamics (Qiu et al., 2025; Zhang et al., 2024b). These events alter forest composition and vertical structure, which regulate global carbon cycles, biodiversity conservation, and surface energy balance (Liu et al., 2024; Venter et al., 2016). In contrast, forestation encompasses afforestation (conversion of long-term non-forest land to forest) and reforestation (forest recovery after temporary degradation or disturbance) that can increase tree cover and restore forest structure, enhancing carbon sequestration, habitat connectivity, soil conservation and ecosystem resilience to climate change (Liao et al., 2024; Yu et al., 2022). Rising extreme events and enhanced human activities under climate change have complicated forest disturbance patterns and created major uncertainties in carbon budget estimations (Fan et al., 2022; Seidl et al., 2017). As a country undergoing dramatic forest changes, China has experienced extensive forestation and intensive forest disturbances over the past four decades. Such forest dynamics exert non-negligible impacts on global carbon cycling (Cheng et al., 2024; Zhu et al., 2025). Thus, distinguishing and monitoring forest disturbances and forestation is essential for understanding terrestrial ecosystem responses to global change.

Remote sensing, particularly long-term Landsat observations at a 30 m resolution, has become the mainstream technique for large-scale forest monitoring (Liang et al., 2025; Zhu et al., 2020). A variety of time-series analysis algorithms, including LandTrendr (Kennedy et al., 2010), mLandTrendr (Qiu et al., 2023), VCT (Huang et al., 2010), CCDC (Zhu and Woodcock, 2014), BFAST (Verbesselt et al., 2010), COLD (Zhu et al., 2020), NRT-MONITOR (Shang et al., 2022), and mCOLD (Shang et al., 2025a), have been developed and widely applied to detect diverse forest disturbance events. Despite substantial advances in monitoring techniques, the lack of high-quality, long-term reference sample datasets remains a major bottleneck limiting the development and practical application of forest remote sensing techniques. Currently, spatially explicit and temporally continuous labelled samples of forest disturbance and forestation remain scarce across China. Most existing studies rely on aggregated administrative records or limited field surveys, which cannot provide standardized benchmarks for the consistent validation of forest disturbance and forestation events.

The absence of standardized reference samples brings three major challenges for national forest monitoring. First, algorithm optimization is severely restricted. Traditional time-series algorithms such as LandTrendr, BFAST, CCDC, and COLD rely on predefined thresholds to separate real disturbances from spectral noise (Liang et al., 2025; Zhu et al., 2020). Without sufficient accurately annotated reference samples, threshold determination lacks objective criteria, resulting in inconsistent detection performance across varying disturbance magnitudes and types. Second, data-driven models suffer from insufficient training data. Machine learning approaches, such as random forest, require large quantities of high-confidence labelled samples to ensure stable model performance (Zhang et al., 2024a). Currently, publicly available samples with explicit change types and



precise occurrence timestamps remain extremely limited for China, which restricts the optimization and practical applicability of data-driven algorithms in this region. Third, product validation and uncertainty quantification are impeded. Both traditional and machine-learning-based forest monitoring products require independent high-quality reference datasets for accuracy assessment and uncertainty analysis (Qiu et al., 2025). The lack of benchmark samples hinders cross-comparison among different forest monitoring products and blocks quantitative evaluation of area estimation biases, further reducing the reliability of large-scale forest carbon estimation. To date, no nationwide continuous sample dataset spanning from 1986 to 2025 (40 years) is available to support long-term forest dynamic monitoring and forest carbon modelling across China.

To fill this data gap, this study constructs a nationwide Reference sample dataset of Stable, Disturbed and Forestated forests (RSDF) for China spanning 1986 to 2025. The dataset includes 12,991 samples at a 30 m spatial resolution, covering major forest types, typical disturbance types and forestation patterns across the country. All samples are annotated by integrating Landsat time-series spectral analysis, multi-stage visual interpretation of high-resolution imagery (Google Earth, PlanetScope and Sentinel-2), and available forest inventory or survey records. This dataset features four distinct strengths: extended temporal coverage that capture four decades of intense forest transitions, 30-m spatial resolution capable of detecting small and fragmented forest changes, stratified random sampling that guarantees robust spatial representativeness, and reliable annotation quality supported by multi-source data fusion and cross-analyst verification. Serving as a standardized benchmark, the RSDF dataset can support the training, validation and performance evaluation of diverse forest monitoring algorithms, enable localized optimization of remote sensing models, and enhance the accuracy of regional forest dynamic products. Furthermore, this open-access dataset provides solid data support for nationwide ecological research involving carbon budget estimation, ecological restoration assessment and biodiversity conservation.

2 Data and methods

2.1 Study Area

The study area covers all forest regions of China (Fig. 1), geographically ranging from 18°N to 53°N in latitude and 73°E to 135°E in longitude. Benefiting from large territorial extent and varied climatic conditions, China contains abundant forest vegetation types, spanning from boreal coniferous forests in high-latitude regions to tropical rainforests and monsoon forests in southern areas (Yao et al., 2024). The major forest types include boreal coniferous forests, temperate mixed coniferous and broad-leaved forests, warm-temperate deciduous broad-leaved forests, and subtropical evergreen broad-leaved forests (Shang et al., 2023). Driven by diverse natural conditions and intensive human activities, forest disturbance regimes exhibit substantial spatial heterogeneity across China. Forest disturbances in China consist of anthropogenic disturbances (e.g., harvest) and natural disturbance events (e.g., fire, insects, and diseases). Furthermore, China has launched large-scale national ecological restoration projects over recent decades, leading to extensive afforestation and reforestation across varied topographic zones (Liao et al., 2024; Yu et al., 2022). The coexistence of heterogeneous forest types, spatially differentiated disturbance drivers,



and widespread forestation makes China an ideal region for developing a national benchmark dataset for long-term forest
 100 dynamic monitoring.

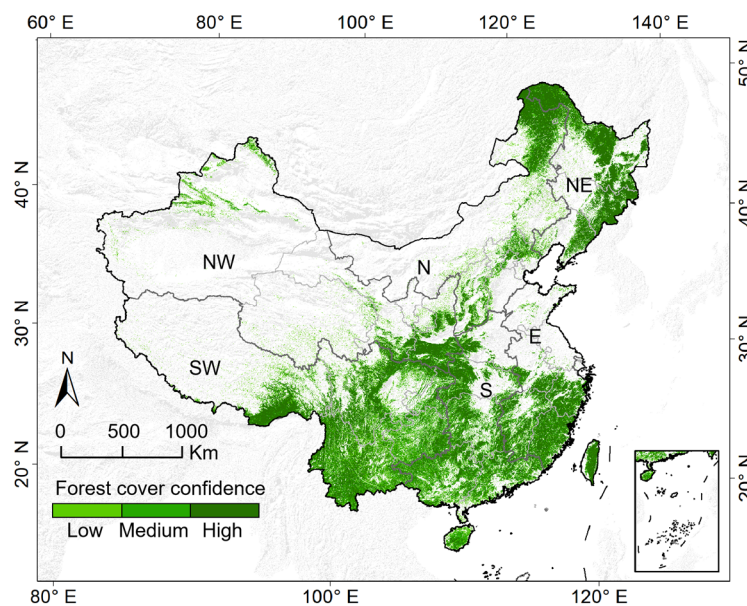


Figure 1: Spatial distribution of China’s forests with different confidence levels. Provincial boundaries are in light gray, and its six major regions are in dark gray with abbreviated labels. NE: Northeast China; E: East China; N: North China; NW: Northwest China; S: South China; SW: Southwest China. The map lines may contain disputed territories.

2.2 Data

This study utilizes Landsat 5/7/8/9 Collection 2 Level-2 surface reflectance data as the primary input to construct the 30 m resolution RSDF dataset spanning 1986 to 2025. Pre-processing was implemented to generate standardized and high-quality Landsat time-series data spanning 1986 to 2025. The quality assessment (QA) band was first used to mask clouds, shadows, and snow/ice contaminants, and all valid observations were retained to build dense time series. To mitigate residual noises, an
 110 adaptive time-series filtering method (Shang et al., 2022) was applied. This method effectively removes random residual noise while preserving genuine abrupt and gradual spectral signals associated with forest disturbances and forestation activities.

Multiple types of high-resolution satellite imagery were also integrated to identify forest disturbance and forestation types and improve interpretation reliability. First, high-resolution Google Earth imagery provides sub-meter true-colour details for visual discrimination of forest changes. Second, PlanetScope imagery (3 m resolution, available since 2016) captures fine-
 115 scale forest changes due to its high spatial and temporal observation frequency. Third, Sentinel-2 imagery (10–20 m resolution) compensates for persistent cloud cover in cloudy regions. All imagery was selected following a resolution-based priority principle: Google Earth > PlanetScope > Sentinel-2 > Landsat. In addition to satellite imagery, available forest inventory, field survey records, statistical yearbooks, and public reports were also used to confirm disturbance dates and types.



A merged maximum forest extent dataset with different confidence levels was adopted for initial forest screening (Fig. 1). This dataset was produced by integrating seven publicly available land cover and forest extent datasets with spatial resolutions ranging from 10 m to 30 m. The 30 m datasets include GLC_FCS30D (1985–2022; Zhang et al., 2024a), Global Forest Cover Change (GFCC) (2000, 2005, 2010, 2015; Sexton et al., 2013), GLASS-GLC (1985–2020; Liu et al., 2021), and CLCD (1985–2019; Yang and Huang, 2021). The 10 m datasets consist of ESA WorldCover 2021 (Zanaga et al., 2022), JRC_GFC2020 (Bourgoin et al., 2024), and Dynamic World (2015–2024; Brown et al., 2022), which were resampled to 30 m using the nearest-neighbour method. Among the seven input datasets, five contained multi-year records, and a pixel's forest label in any year defined its maximum potential forest extent for that dataset. A voting strategy was then applied to quantify forest confidence levels: ≥ 5 datasets for high confidence, 3–4 datasets for median confidence, and 1–2 datasets for low confidence. All candidate forest pixels were further verified through manual visual interpretation during sample annotation to confirm final forest status.

2.3 Sample selection strategy

A spatially-constrained stratified random sampling framework was applied to guarantee the spatial uniformity of samples. This sampling workflow was implemented in three steps. First, a nationwide regular grid system was established based on the Asia North Albers Equal Area Conic projection, dividing China into continuous $150 \text{ km} \times 150 \text{ km}$ analysis-ready data (ARD) tiles (Fig. 2a). Each tile contains 5000×5000 pixels at a 30 m resolution and serves as an independent stratum for stratified sampling. Second, adaptive random sampling was conducted within each stratum. The initial sample size of each ARD tile was determined by its forest coverage (Fig. 2a), calculated as follows: $\text{sample size} = \text{round}(100 \times \text{forest coverage})$, yielding 14,848 initial samples across China (Fig. 2b). This rule allocates more samples to densely forested regions while preserving a reasonable number of samples in areas with sparse forest cover. Last, all candidate samples were visually interpreted using multi-source time-series imagery, including high-resolution Google Earth imagery, PlanetScope, Sentinel-2, and Landsat observations. Samples identified as non-forest land throughout the 1986–2025 period were excluded. This sampling framework satisfies the core statistical criteria for stratified sampling by producing low intra-stratum heterogeneity and high inter-stratum variability.

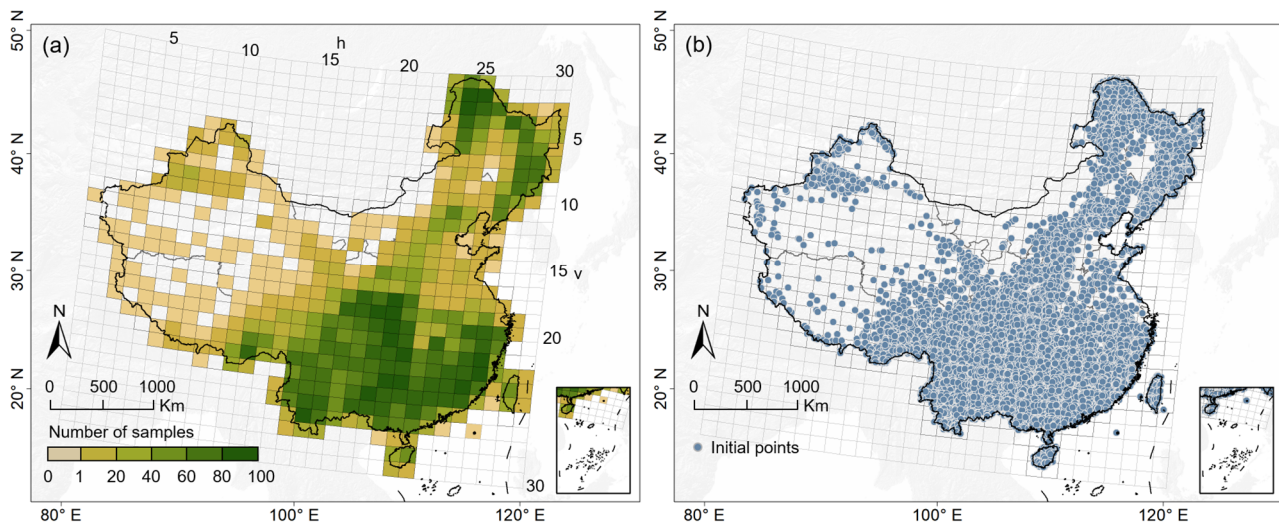


Figure 2: National sampling framework and sample spatial distribution. (a) The 150 km × 150 km ARD tile system for stratified sample assignment; (b) distribution of initial candidate samples. The map lines may contain disputed territories.

2.4 Visual interpretation

A hierarchical visual interpretation framework was implemented to identify forest disturbances and forestations. For each sample, the full annotation workflow consists of three sequential stages (Fig. 3). First, potential spectral breakpoints were extracted from the pre-processed Landsat time series data. Surface reflectance at six spectral bands and two vegetation indices including Normalized Difference Vegetation Index (NDVI) and Normalized Burn Ratio (NBR) were used for breakpoint identification. A candidate breakpoint was flagged when abrupt shifts occurred in the time-series surface reflectance or vegetation indices, and these spectral mutations persisted for a certain period. Second, multi-source remote sensing imagery acquired before and after each candidate breakpoint was applied for visual verification to confirm whether the detected spectral mutation corresponded to actual forest dynamics. The judgement criterion relies on clear pre- and post-event RGB imagery to visually identify explicit forest alteration traces. Third, all verified pixels were uniformly annotated. Samples with confirmed forest changes were recorded with precise occurrence dates and detailed disturbance or forestation types (Table 1), while samples without changes were labelled as stable forest.

To ensure interpretation consistency and accuracy, a strict priority principle following the highest spatial resolution rule was applied: Google Earth > PlanetScope > Sentinel-2 > Landsat 5/7/8/9. The optimal source was determined considering temporal coverage and data availability. The hierarchical verification rules were defined as follows. First, candidate pixels were imported into Google Earth Pro, and sub-meter high-resolution imagery spanning before and after breakpoint timing was retrieved. Confirmed forest changes were recorded with Landsat-derived breakpoint dates, with change types identified through visual interpretation. Second, for pixels lacking available Google Earth archives, near daily 3 m resolution PlanetScope imagery was utilized. This dataset provides continuous observations since 2016, offering reliable support for validating recent forest change



events. Third, for historical changes before 2016, composite RGB imagery was automatically extracted from Google Earth
 Engine (GEE) platform. Sentinel-2 imagery was prioritized, while Landsat sensors were selected in the order of Landsat 9,
 Landsat 8, Landsat 7, and Landsat 5. If no visible forest alteration was identified from multi-temporal imagery, the pixel was
 ultimately categorized as stable forest. Figure 4 shows a typical visual interpretation example of forest harvest based on Landsat
 time series and multi-temporal PlanetScope imagery. Both the six spectral bands of surface reflectance and two vegetation
 indices successfully detected an obvious spectral breakpoint in the time series. Correspondingly, the PlanetScope imagery
 covering the breakpoint period clearly recorded the complete disturbance trajectory, encompassing forest cover before harvest,
 sudden canopy loss during harvest, and altered land surface conditions after the disturbance.

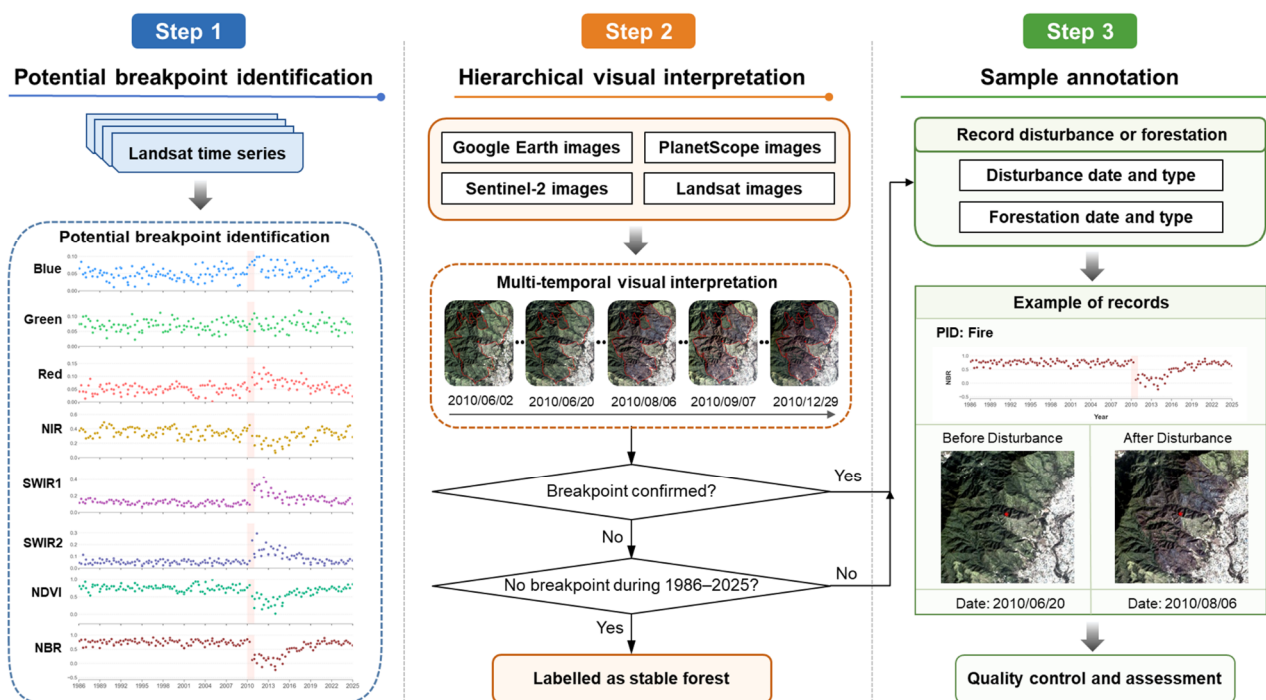


Figure 3: Flowchart of the visual interpretation for forest disturbance and forestation samples.

Table 1: Definitions of forest disturbance and forestation types.

Category	Type	Definition
Forest disturbance	Harvest	Removal of forest through clear-cutting, selective logging, or other timber-harvest activities, resulting in sudden tree cover loss.
	Fire	Natural or human-caused fire events that burn forest, leading to partial or complete canopy damage or removal.
	Other disturbances	All other forest disturbances not classified as harvest or fire, including windthrow, landslides, drought, insects, diseases and so on.



175

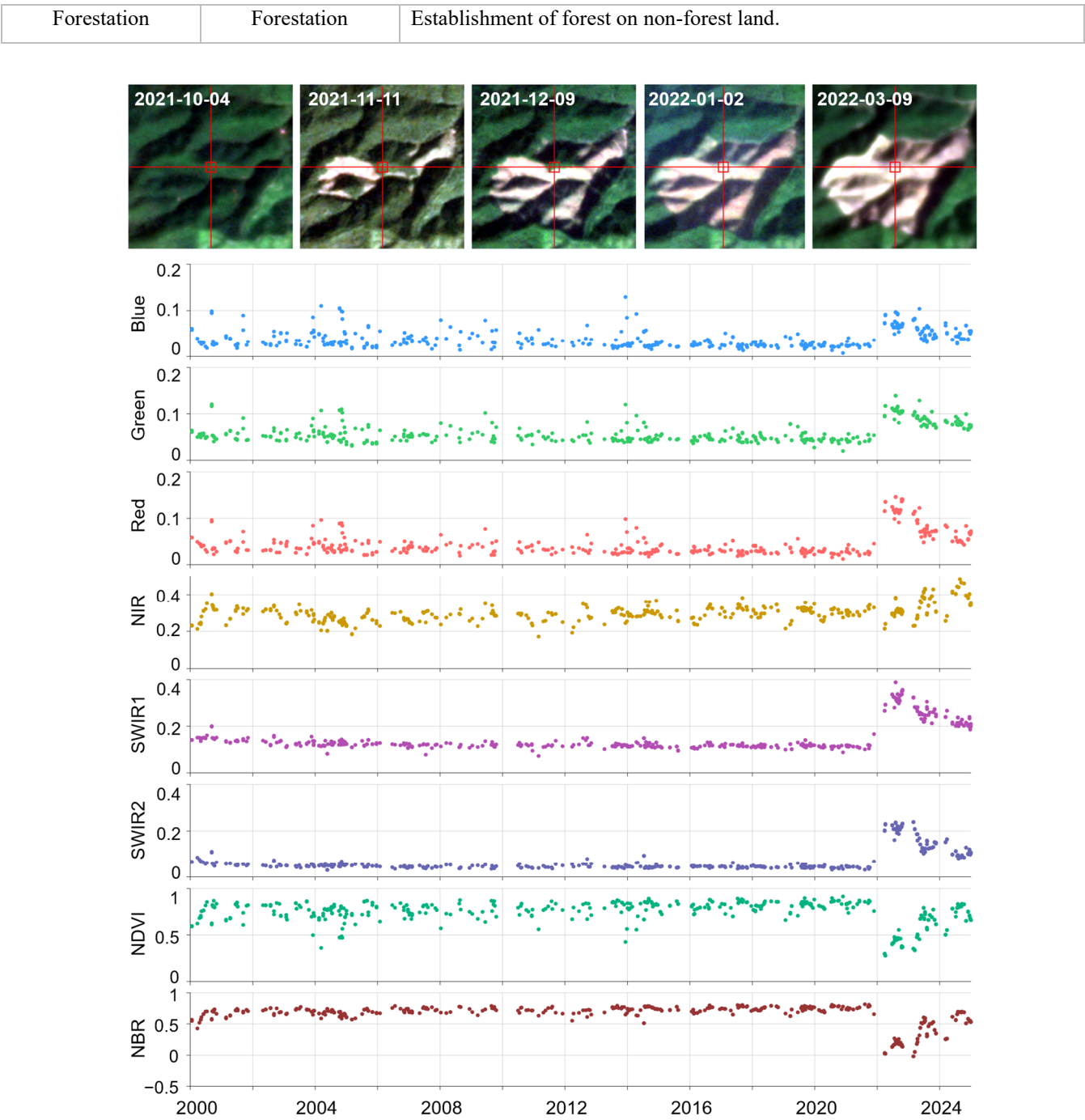


Figure 4: Visual interpretation example of a forest harvest event, showing multi-temporal PlanetScope imagery at the sample location and corresponding Landsat time series of six spectral bands and two vegetation indices.



2.5 Quality control and assessment

180 To guarantee the temporal accuracy and label reliability of forest reference samples, this study developed a multi-stage quality control framework following widely recognized best practices for reference dataset construction (Olofsson et al., 2014; Stehman and Foody, 2019). Adopting balanced incomplete block designs (BIBDs) (Yasmin et al., 2015), all 14,848 candidate samples were randomly split into 15 independent subsets, each containing approximately 1,000 samples. Fifteen professional interpreters each annotated three subsets, ensuring every sample received an initial blind evaluation by three experts. To reduce
 185 labor costs for disputed samples without inconsistent initial judgements, a sequential consensus dynamic termination framework was implemented. The total number of experts involved in the k^{th} round of interpretation was calculated as Eq. (1).

$$E_k = E_1 + S \times (k - 1) = 2k + 1 \quad (1)$$

where, E_1 is the fixed number of initial experts ($E_1 = 3$); S denotes the number of newly added experts per additional round ($S = 2$); and E_k represents the cumulative number of experts participating in the k^{th} interpretation round. A quantitative threshold
 190 governing the stopping rule for iterative cross-validation is defined by Eq. (2).

$$V_{max} - (E_k - V_{max}) \geq M \quad (2)$$

where, V_{max} is the maximum number of experts; and M is the predefined consensus margin threshold. Setting $M = 3$, the termination criterion is simplified to Eq. (3).

$$V_{max} \geq (E_k + 3)/2 \quad (3)$$

195 The iterative interpretation continues until the vote margin satisfies the upper expert limit of 15 is reached. This mechanism accelerates processing for unambiguous high-confidence samples and concentrates expert resources on confusing boundary samples with indistinct visual features. Once the interpretation of a sample is completed, its confidence level is quantified via Eq. (4).

$$CL = V_{max}/E_k \times 100\% \quad (4)$$

200 where, CL refers to the final interpreted confidence of the sample. All reviewer IDs, annotation timestamps and individual confidences were recorded throughout the workflow. These traceable audit records enable quantitative uncertainty assessment and support sample weighting in subsequent machine learning and remote sensing inversion models.

Two statistical metrics were further calculated to quantify pairwise interpretation consistency among the 15 interpreters: simple pairwise agreement rate and Cohen's Kappa coefficient (Cohen, 1960). The simple pairwise agreement rate is defined
 205 by Eq. (5).

$$P_a = N_{agree}/N \quad (5)$$

where, P_a is the simple agreement rate; N is the number of samples jointly assessed by a pair of interpreters; and N_{agree} is the count of samples receiving identical labels from both interpreters. This metric ranges from 0 to 1 and directly reflects raw



annotation overlap, yet it fails to eliminate spurious agreement caused by random label matching. The Cohen's Kappa
 210 coefficient removes such random bias using the Eqs. (6–9).

$$\kappa = (P_a - P_e) / (1 - P_e) \quad (6)$$

$$P_e = \sum_{i=1}^c P_{A,i} P_{B,i} \quad (7)$$

$$P_{A,i} = a_i / N \quad (8)$$

$$P_{B,i} = b_i / N \quad (9)$$

215 where, κ is the Cohen's Kappa coefficient; P_e denotes expected chance agreement, the theoretical probability of identical labels
 under fully random annotation; c is the total number of forest change categories; $P_{A,i}$ and $P_{B,i}$ represent the marginal
 probability of interpreters A and B assigning a sample to category i ; a_i and b_i are the counts of samples labelled category i by
 interpreter A and B . The Kappa coefficient ranges from -1 to 1 : a value of 1 represents perfect pairwise interpretative
 consistency, a value of 0 signifies agreement equivalent to random labelling, and negative values indicate lower annotation
 220 overlap than random chance.

For dataset validation and applications, overall accuracy (OA), Producer's Accuracy (PA), User's Accuracy (UA), and F1
 score (Liang et al., 2025) calculated against the RSDF reference were used to evaluate the performance of the forest disturbance
 or forestation classification models, as well as existing forest disturbance products. Higher OA and F1 scores indicate superior
 model and product performance. UA corresponds to 100% minus commission error, while PA corresponds to 100% minus
 225 omission error. Balanced UA and PA indicates that a model or product achieves stable and reliable performance in detecting
 both changed and unchanged forest samples, avoiding systematic overestimation or underestimation biases in forest change
 detection (Zhu et al., 2020; Shang et al., 2022).

3 Results

3.1 Dataset description

230 The RSDF dataset contains three core archived files. The first file, titled "*RSDF.csv*", stores basic annotation information of
 all forest dynamic samples. This dataset comprises seven attributes: PID, Latitude, Longitude, Year, DOY, Type and
 Confidence, with detailed field specifications provided in Table 2. Sample PID follows the *yhhvvxxxx* coding format. The *y* is
 the version of the dataset, *hh* and *vv* represent the identification number of China's ARD tiles, and *xxxx* denotes the unique
 number of each individual sample within the ARD tile. Sample IDs are non-consecutive, as non-forest pixels and samples with
 235 interpretation confidence below 50% have been excluded. Null values for Year, DOY, and Type indicate stable forest with no
 forest disturbance or forestation occurred between 1986 and 2025. Non-null values represent valid forest disturbance or
 forestation records. Specifically, Year and DOY document the occurrence date, and Type categorizes forest change events.
 Disturbance types include harvest, fire and other disturbance, as summarized in Table 1. The Confidence field quantifies the



consensus level of judgements from multiple interpreters. A single data entry corresponds to one individual forest change
 240 event, while multiple entries for the same pixel indicate repeated forest dynamics over the study period.

The second file, titled “*RSDF_SurfaceReflectance.zip*”, is a compressed ZIP archive that stores Landsat time-series surface
 reflectance data for all samples. Each sample has a standalone CSV file named “*Pyhhvvxxx.csv*”, where the prefix *P* denotes
 sample point and *yhhvvxxx* matches the sample’s PID. Each CSV file records Year, Month, Day, six spectral bands (Blue,
 Green, Red, NIR, SWIR1 and SWIR2), and the QA quality flag. Relevant band parameters are applicable to Landsat 5, 7, 8
 245 and 9 sensors. The QA value serves as a validity indicator: a value of 1 indicates a valid observation, whereas a value of 0
 indicates an invalid observation.

The third file, titled “*RSDF_Figures.zip*”, is another ZIP archive organized according to the ARD tile IDs (H). Each tile is
 represented by an integrated PDF file named “*Hxx.pdf*”, where *Hxx* denotes the corresponding ARD tile ID. As tiles H01–H05
 contain relatively few samples, the materials for these five tiles were combined into a single file named “*H01toH05.pdf*”. These
 250 files provide visual evidence of forest disturbance and forestation samples, including time-series profiles of the six spectral
 bands and two vegetation indices, as well as pre-event and post-event RGB composite imagery for visual validation.

Table 2: Field description of the RSDF dataset.

Fields	Descriptions
PID	Unique identification number assigned to each interpreted sample unit.
Latitude	Central latitude of the sample, decimal degree format, WGS-84 coordinate system.
Longitude	Central longitude of the sample, decimal degree format, WGS-84 coordinate system.
Year	The year in which the forest disturbance or forestation was interpreted and validated.
DOY	The days within the year of the interpreted forest disturbance or forestation, ranging from 1 to 365 or 366.
Type	The type of the interpreted forest disturbance or forestation.
Confidence	The confidence of the interpreted stable forest, forest disturbance or forestation between different interpreters.

3.2 Spatial coverage and patterns of the dataset

Samples within the RSDF dataset exhibit notable spatial heterogeneity. Figure 5a shows the spatial distribution of forest
 255 disturbance or forestation frequencies across China. Across the 12,991 samples, 79.61% represent stable forest with no
 detectable change from 1986 to 2025, 15.99% undergo a single disturbance or forestation event, and 4.40% experience two or
 more events. Forest change frequencies vary substantially across different geographical regions (Fig. 6a). Northwest, North
 and Northeast China are dominated by stable forest samples, exceeding 87%, while samples with two or more changes account
 for less than 1%, indicating an extremely low occurrence probability of forest disturbance or forestation. In Southwest China,
 260 stable forest samples constitute 88.97%, with a moderately elevated share of change samples. East and South China stand out
 as the regions with the most intensive forest change activity; stable forest samples only account for 50.90% and 70.99%,
 respectively, and these two regions boast the highest proportions of samples affected by single and multiple changes nationwide.



Overall, forest disturbance frequency follows a clear spatial gradient, with substantially higher disturbance activity in eastern and southern forested areas compared to North, Northeast, and Northwest China. Figure 5b further illustrates the spatial distribution patterns of all forest change classes, encompassing forest disturbances and forestation. In total, the dataset documents 2,649 harvests, 547 fires, 126 other disturbances, and 54 forestation events. Each forest change type displays distinct regional spatial signatures (Fig. 6b). As the dominant disturbance category, harvest occurs across all forested regions nationwide, with the densest concentrations concentrated in South, East, and Southwest China. Fires and other disturbances are largely confined to East, South, and Southwest China; fire frequency drops drastically across North and Northeast China, and no fire samples are present in Northwest China. Overall, disturbance event is intense in eastern and southern regions but sparse across northern and northwestern territories. Meanwhile, forestation events remain rare in every region and constitute a negligible fraction of total national forest dynamics.

Provincial-level spatial disparities are also evident in forest change frequency and change types. Figure 7 shows provincial statistics including forest coverage and forest areas from the 9th National Forest Inventory, total sample counts, and sample counts for all forest disturbance and forestation categories. Yunnan (1,334), Sichuan (1,230), and Heilongjiang (1,030) have the largest samples, reflecting their extensive forest coverage and diverse native forest vegetation. By contrast, Ningxia, Qinghai and other provincial-level administrative divisions host far fewer samples. When comparing total disturbance events, Fujian ranks first nationwide with 535 disturbance events, followed closely by Guangdong with 488 events. Southern provinces generally experience frequent change activity: Guangxi, Jiangxi, Yunnan, Hunan each record more than 150 disturbance events. Most provinces in North and Northwest China, however, register fewer than 30 disturbance events, with Qinghai, Shanxi, and Shandong showing minimal disturbance activity. Forest coverage rate exhibits a strong positive correlation with disturbance frequency: southeast provinces including Fujian, Jiangxi, Guangxi, and Yunnan all maintain forest coverage rates above 55% alongside frequent disturbance events, while arid and semi-arid provinces in North and Northwest China typically have forest coverage below 20% and drastically fewer disturbance records. When disaggregated by disturbance types, harvest represents the most pervasive disturbance type across all provinces. Fujian records the highest volume of harvest events at 427, with Guangdong, Guangxi, and Jiangxi following closely behind. Fires demonstrate strong spatial clustering: Fujian contains 104 fire samples, vastly exceeding all other provinces, while Guangdong, Inner Mongolia, and Jiangxi report the second-highest fire frequencies. Plain provinces in North and East China and all northwest provinces hold almost no wildfire samples. Other disturbances are scattered geographically with no clear concentrated hotspots, and provincial totals for this category mostly sit below 30 samples. Notably, forestation events remain limited across the entire dataset.

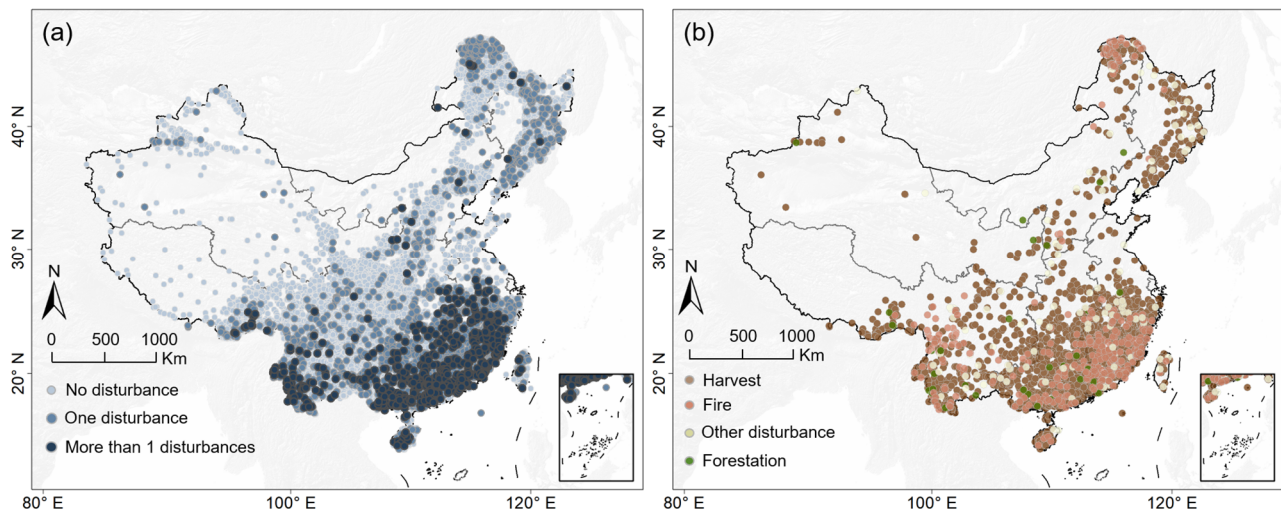


Figure 5: Spatial distribution of forest change frequencies (a) and forest disturbance and forestation types (b) in the RSDF dataset.

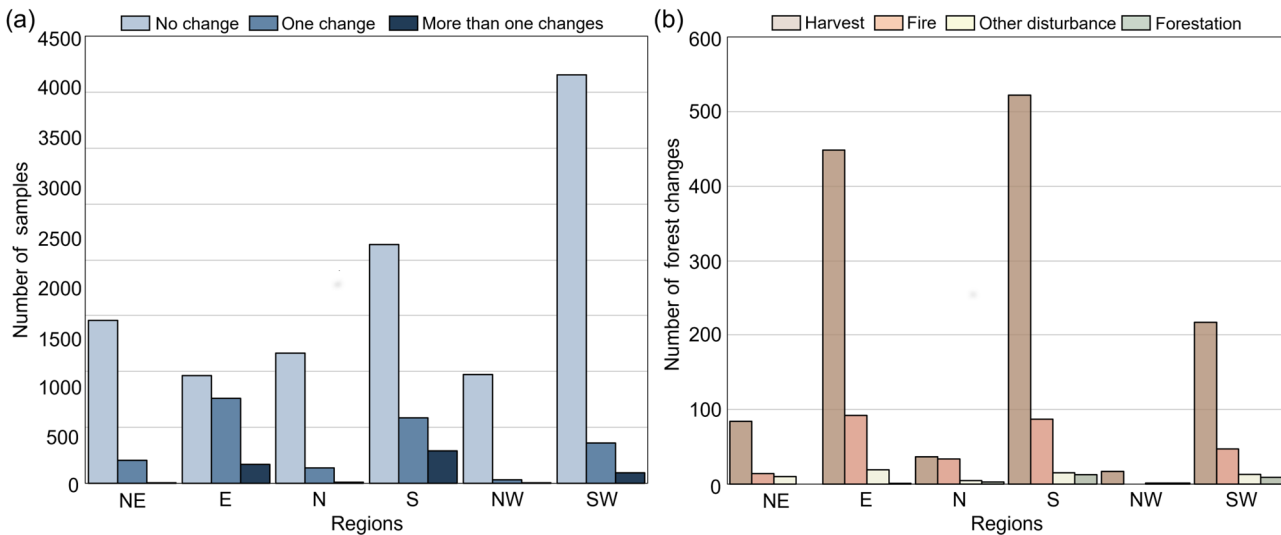


Figure 6: Histogram of forest change frequencies (a) and forest disturbance and forestation types (b) in the RSDF dataset.

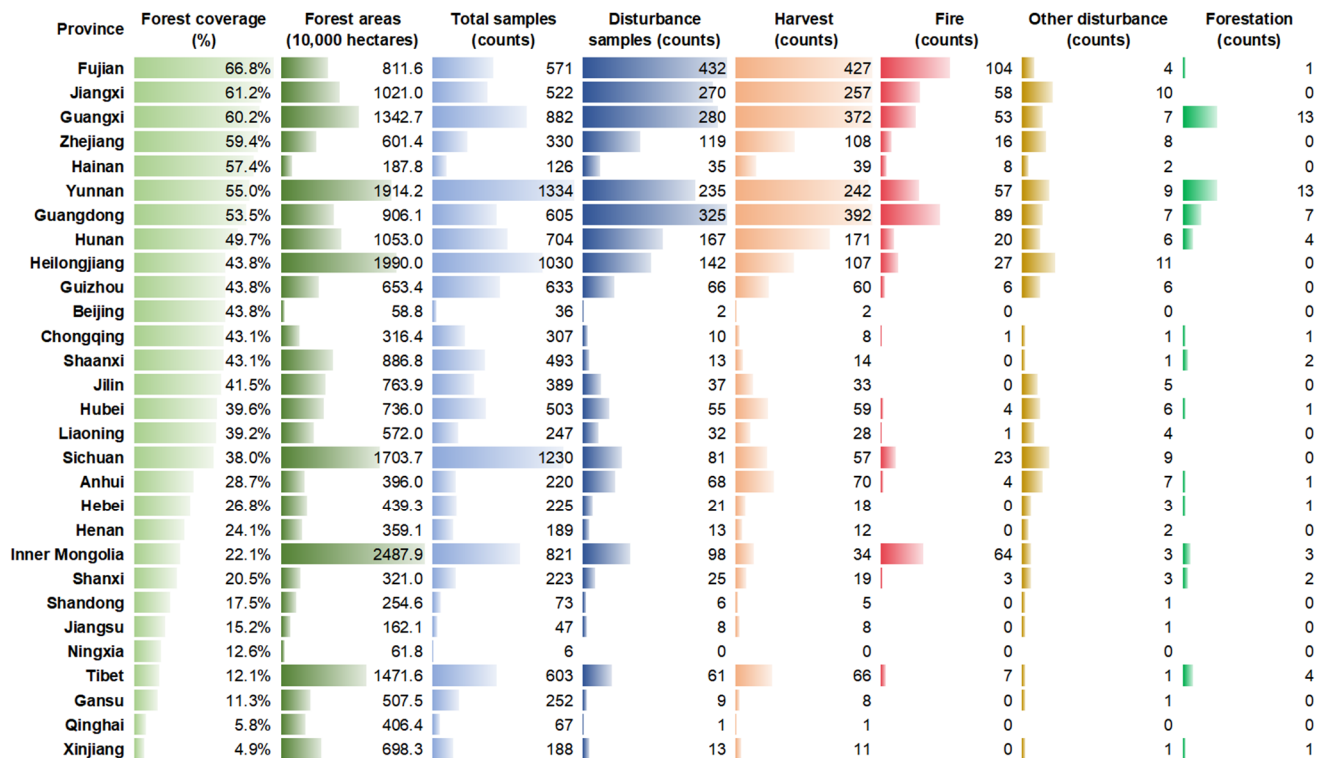


Figure 7: Statistics of the reference samples in provinces of China. Forest coverage and forest areas are derived from China's 9th national forest inventory data during 2014–2018. Provinces with less than 10 samples are excluded.

3.3 Temporal coverage and patterns of the dataset

Samples within the RSDF dataset exhibit divergent temporal trends. Figure 8 shows the time-series changes in the number of harvest, fire, other disturbance, and forestation events, alongside the total number of forest change events. Among all forest change categories, harvest acts as the dominant driver of forest dynamics and accounts for the largest share of events across every year. It trends upward with persistent fluctuations starting in 1986, peaking at 121 events in 2021. While harvest levels remained relatively high after this peak, the metric has shown a short-term declining trend in recent years. As the second prevalent disturbance type, fire features distinct episodic outbreaks: fire counts surged dramatically in 2003 (72 events) and 2000 (42 events), substantially increasing the total annual disturbance volume in those years. In contrast, other disturbance and forestation activities show mild interannual variability with no marked upward or downward trends; their annual event counts remain consistently low, with no years featuring large-scale clustered occurrences. The 40-year period of forest disturbance and forestation samples can be split into two distinct phases based on their temporal trajectories. From the late 1980s to the early 2010s, the fluctuating upward trend in harvest activity, combined with recurrent fire outbreaks, collectively drove a nationwide increase in total forest disturbance frequency. From the early 2010s through 2025, the upward trend in harvest remains evident, while the severity of episodic fires weakens correspondingly, leading to a continued increase in overall



national forest disturbance magnitudes primarily driven by harvesting. Collectively, the four forest change types exhibit highly divergent temporal patterns. The long-term rise and subsequent decline in anthropogenic harvest governs the overall temporal variability of forest disturbances across the four decades, whereas fire, other disturbance, and forestation exert only secondary modulating effects.

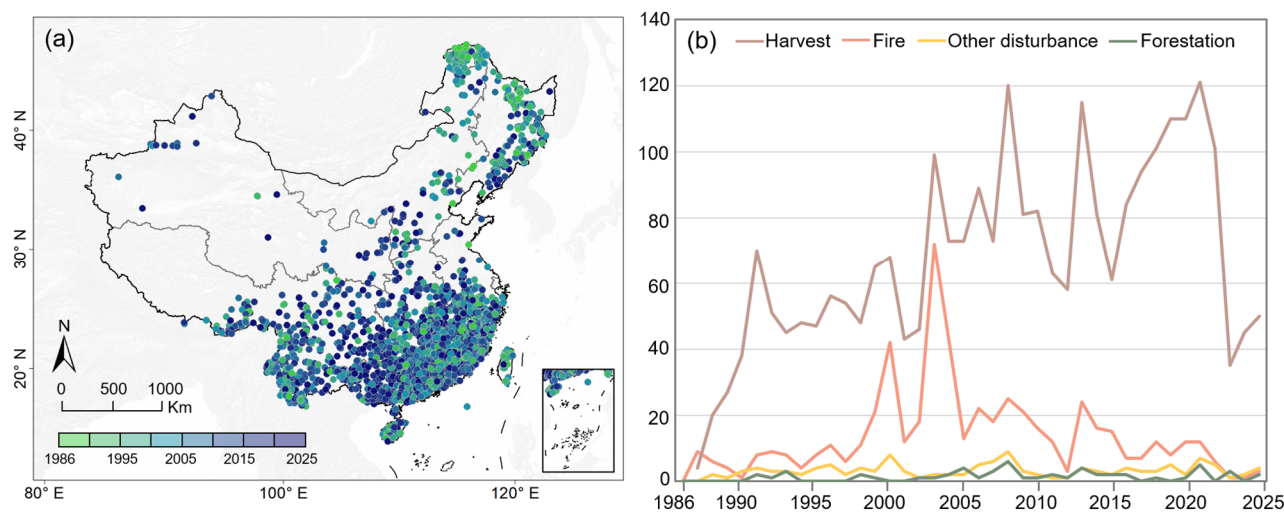


Figure 8: Spatial distribution of the latest forest disturbance or forestation year (a) and temporal variations of different forest change types (b) in the RSDF dataset.

3.4 Quality assessment

The annotation reliability of the RSDF dataset is assessed from two dimensions: sample-level interpretative confidence and pairwise consistency across the 15 interpreters. Figure 9 shows the national spatial pattern and statistical histogram of sample confidence levels generated via sequential consensus cross-validation. Samples with confidence $\geq 80\%$ account for more than 98% of all records; low-confidence samples are sparsely distributed across western areas and fragmented forest transition zones, where mixed pixels and ambiguous vegetation signals easily lead to interpretive disputes. Statistically, confidence values follow a strongly positive skewed distribution: 13,186 records achieve a perfect confidence of 100% (record count exceeds sample count because individual samples can generate multiple records following repeated forest changes), and 358 records meet the 80% confidence threshold. Figure 10 shows pairwise consistency heatmaps of the 15 interpreters derived from raw agreement rates and chance-adjusted Cohen's Kappa coefficient. Interpreters exhibited strong baseline concordance during the cross-validation process, achieving an overall agreement of 0.83 ± 0.07 . The Cohen's Kappa coefficient yielded an average value of 0.65 ± 0.13 , indicating substantial and reliable inter-expert annotation consistency. Specifically, the highest agreement among well-aligned interpreter pairs reached 0.98, with a corresponding Kappa value of 0.95, while pairs evaluating ambiguous sample subsets obtained moderate Kappa values averaging 0.41. The raw agreement matrix shows uniformly high values (mostly > 0.85), whereas the Kappa coefficient matrix exhibits lower values with greater variance. These inconsistent



340

annotations stem primarily from subjective interpretation discrepancies over fragmented mixed forest edge pixels concentrated in central-western mountainous regions. Collectively, the dual validation of sample confidence metrics and cross-interpreter pairwise consistency demonstrates that the RSDF dataset delivers stable and high-quality reference labels applicable to forest remote sensing research.

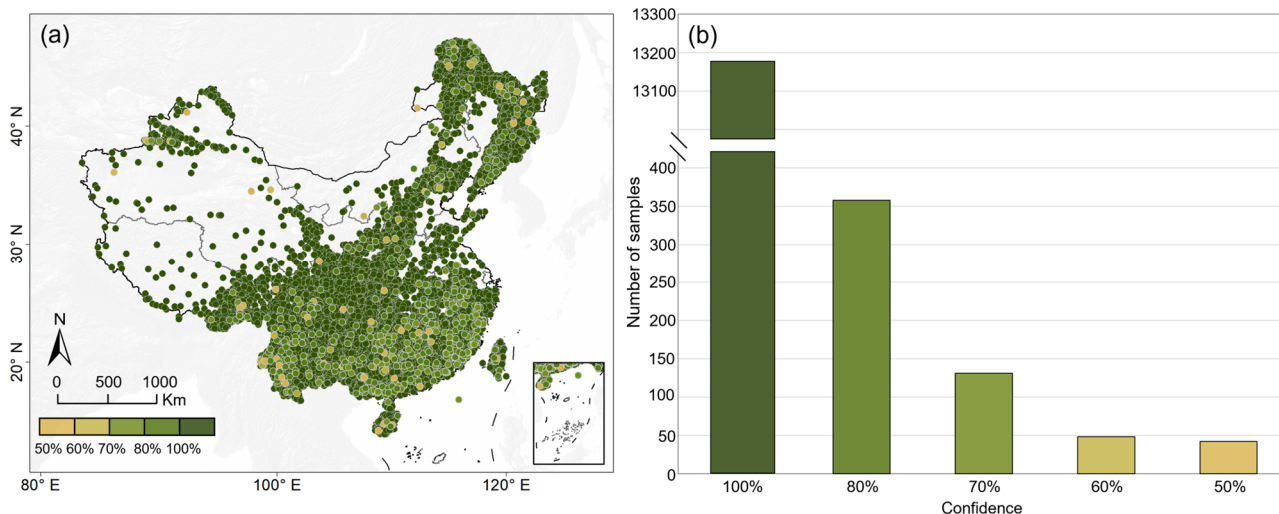
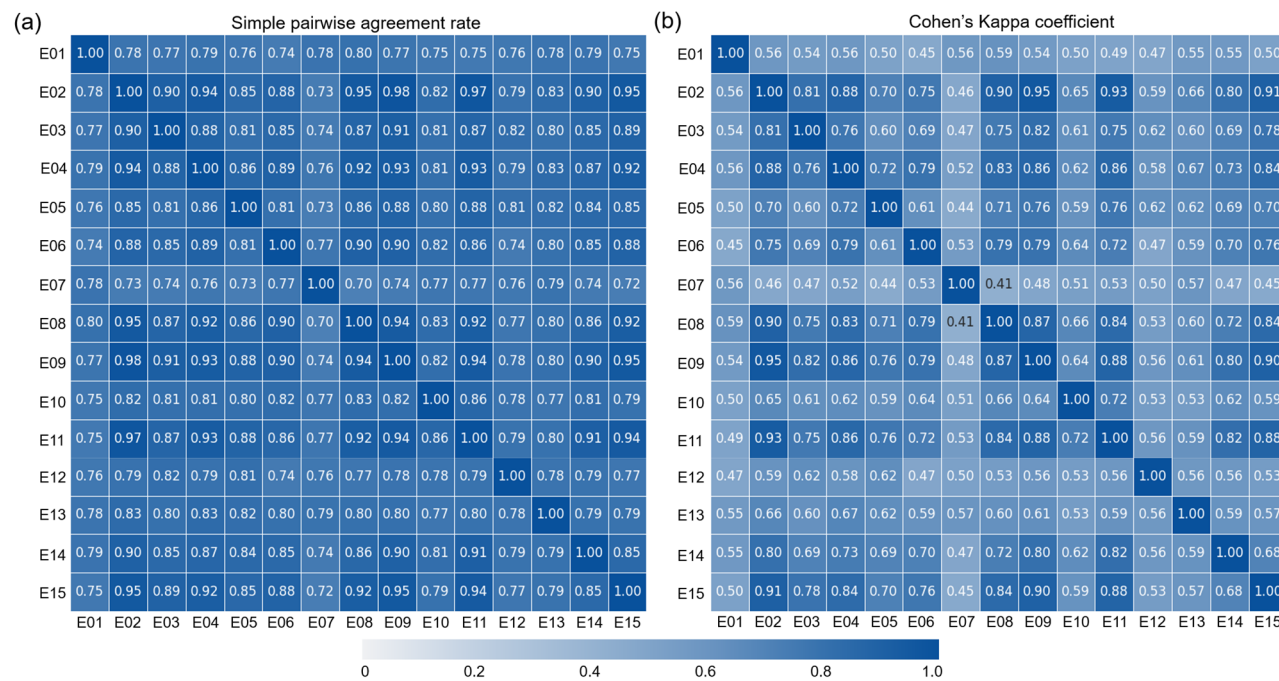


Figure 9: Spatial distribution (a) and histogram (b) of the confidence levels for the reference samples in the RSDF dataset.





345 **Figure 10: Pairwise inter-expert agreement heatmaps for quality control of the RSDF dataset: (a) Simple pairwise agreement rate and (b) Cohen’s Kappa coefficient.** Colour intensity corresponds to the magnitude of agreement values.

4 Dataset applications

4.1 Optimal parameterization and validation of forest disturbance monitoring algorithms

The RSDF dataset serves as a robust validated benchmark for calibrating optimal thresholds and independently evaluating forest disturbance monitoring algorithms. Using a balanced error minimization framework, Liang et al., (2025) used a previous version of this dataset to compare six representative algorithms (VCT, LandTrendr, mLandTrendr, BFAST, CCDC, and COLD) at national and regional scales across China. Specifically, 50% of the reference samples were used for parameter calibration, and the other 50% for independent validation. National validation showed that COLD achieved the highest accuracy, with an F1 score of 81.81%. mLandTrendr performed best among annual-scale algorithms with an F1 score of 73.04%, even outperforming the daily-scale BFAST algorithm. Across China’s six regions, COLD maintained consistently high accuracy (F1: 79.62%–82.45%), peaking in North China and declining in Southwest China due to persistent cloud contamination and complex topography.

The dataset also supports quantitative assessments of topographic impacts and the effectiveness of topographic correction (TC) for forest disturbance monitoring. Based on a previous version of the RSDF dataset, Yang et al. (2026) evaluated the six disturbance algorithms across gradients of slope steepness and sunlit–shaded conditions. Four commonly used TC methods were tested: Cosine Correction (CC) (Teillet et al., 1982), Sun-Canopy-Sensor + C (SCS+C) (Soenen et al., 2005), Illumination Correction (IC) (Tan et al., 2013), and Path Length Correction (PLC) (Yin et al., 2018). The results indicate that all disturbance algorithms exhibit substantial accuracy degradation on shaded slopes and steep slopes $\geq 30^\circ$, with F1 reductions ranging from 6.35% to 14.01%. IC yielded the most stable and consistent accuracy improvements across all algorithms. Reflectance-based CCDC and COLD algorithms benefited more substantially from TC, with a maximum F1 score improvement of 5.50%, whereas other algorithms showed moderate improvements of 1.40%–4.67%. These findings confirm that topography strongly distorts Landsat-based forest disturbance monitoring in mountainous areas, and appropriate TC strategies can effectively enhance monitoring reliability. Collectively, the RSDF dataset enables standardized algorithm comparison, robust parameter optimization, and terrain-adaptive monitoring design, offering a reliable benchmark for operational large-scale forest disturbance and forestation mapping.

4.2 Training and validating the forest disturbance or forestation classification models

The RSDF dataset enables robust training and independent validation of supervised classification models for mapping forest disturbance types across heterogeneous landscapes. A representative sub-region (tile ID: h from 20 to 22 and v from 24 to 26) in Southern China was selected as the test area. This region features complex terrain, frequent human disturbances, and diverse disturbance types, making it ideal for evaluating model generalization. A total of 310 labelled samples were extracted from



the RSDF dataset and randomly split into 50% for model training and 50% for independent validation. Based on the critical predictors identified via feature importance evaluation, including spectral change magnitude (e.g., ΔNBR , $\text{SWIR1}/\text{SWIR2}$ differences), temporal dynamics (e.g., disturbance duration), patch spatial geometry (e.g., patch area, shape complexity index), and topographic constraints (e.g., elevation, slope) derived from Landsat time series (Qiu et al., 2025), a random forest classifier was trained to identify forest disturbance or forestation types. Figure 11a shows the spatial distribution of the classified forest disturbance or forestation, and the corresponding confusion matrix are provided in Fig. 11b. The four-category classification yielded an overall accuracy of 89.68%, with solid F1 scores across all reference types. Harvest achieved the highest F1 score of 0.94, demonstrating its most robust classification performance. Fire (0.76), other disturbances (0.75) and forestation (0.73) produced lower but still credible F1 scores in descending order. The uniformly competitive F1 metrics verify the classifier's strong capacity to differentiate diverse forest change signals. These results confirm that the RSDF dataset can provide reliable training and validation data for operational forest disturbance and forestation type classification.

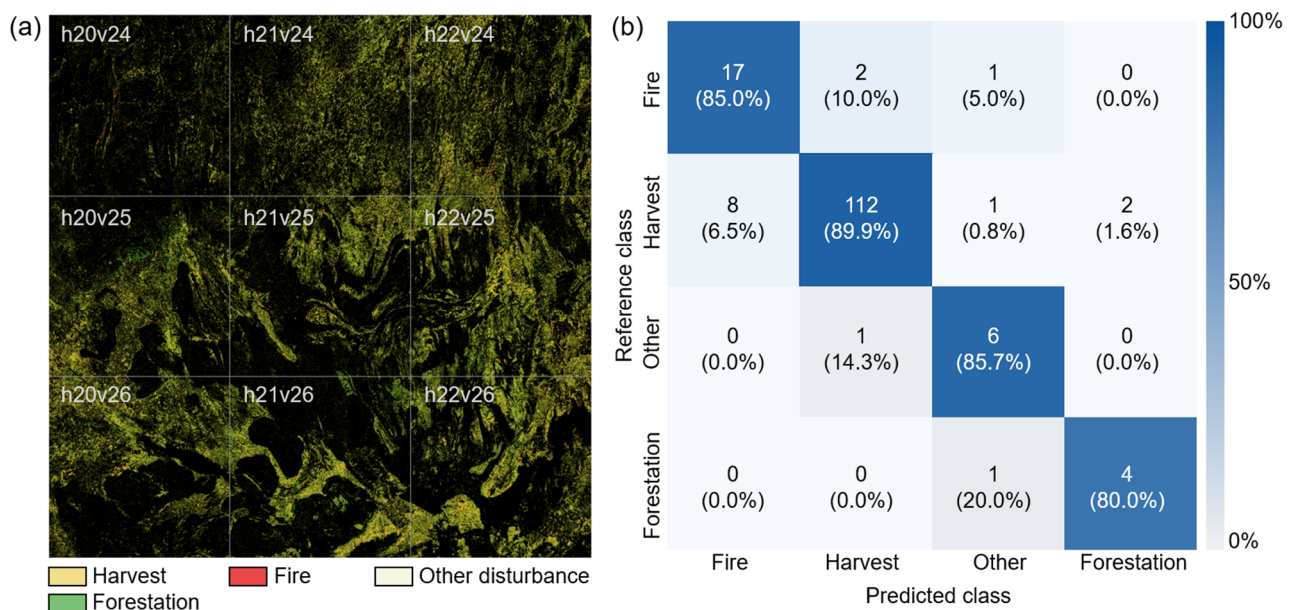


Figure 11: Example of training model for the classification of forest disturbance or forestation types and validation using the RSDF dataset. (a) Map of latest forest disturbance or forestation at a 30 m resolution from 1986 to 2025; (b) Confusion matrix of the classification.

4.3 Training and validating young forest age mapping models

The RSDF dataset provides high-quality forest disturbance and forestation labels to support the development and validation of young forest age mapping across China. Forest disturbance monitoring and random forest models were tailored for forest age estimation in China's annual forest age (CAFA) product spanning 1986 to 2022 (Shang et al., 2025b). A stratified estimation strategy across six geographical regions and five forest cover types was adopted to mitigate environmental and vegetation-



related heterogeneity. By leveraging accurately identified disturbance years from an early version of the RSDF dataset, the proposed methods achieve high accuracy for young forest age mapping, with a root mean square error (RMSE) of ± 2.48 years. These results confirm the practical value of the RSDF dataset for developing and validating robust young forest age mapping frameworks across China.

400 4.4 Validation of existing forest disturbance products

The RSDF dataset provides reliable and high-resolution reference data to validate and intercompare mainstream forest disturbance products. Five representative ARD tiles (h25v02 in Northeast China, h24v03 in Northern China, h26v24 in Eastern China, h25v25 in Southern China, and h16v23 in Southwest China) were selected to benchmark three widely used global forest disturbance datasets, including the 500 m MODIS Version 6.1 global burned area product (MCD64A1) (Giglio et al., 2021),
 405 the 250 m ESA Climate Change Initiative burned area dataset (FireCCI51) (Lizundia-Loiola et al., 2020), and the 30 m Hansen global forest loss product (Hansen et al., 2013). Figure 12 shows spatial patterns of mapped forest disturbance years across the five tiles for all three products from 2001 to 2020. Quantitative accuracy metrics including OA, PA, UA, and F1 score derived from cross-comparison against the RSDF reference across all tiles are shown in Fig. 13. The OA of the three products relative to RSDF ranges from 80.67% to 88.36%, with PA values of 51.52%–67.65%, UA values of 74.19%–86.08%, and F1 scores
 410 (scaled $\times 100$ for visualization) of 0.64–0.73. Among the three products, MCD64A1 yielded the highest overall agreement with RSDF (OA = 88.36%, F1 = 0.73), whereas the Hansen forest loss product showed the lowest performance (OA = 80.67%, F1 = 0.64). The two burned-area products exhibited relatively high overall accuracy, but their error characteristics varied substantially among regions. FireCCI51 showed relatively strong performance in Northern China but pronounced commission errors in Northeast China, where PA remained relatively high while UA decreased to 41.18%. This regional commission bias
 415 may be associated with differences in burned-patch characteristics, active-fire density, and heterogeneous land-cover backgrounds (Franquesa et al., 2022). In contrast, the Hansen product exhibited a pronounced omission tendency, with an overall PA of only 51.52% despite a high UA of 86.08%, which may be attributed to its tree-cover-loss-based mapping framework that does not explicitly represent partial or repeated disturbance events within the same pixel (Hansen et al., 2013). Overall, these results highlight the utility of RSDF as a high-resolution benchmark for evaluating regional uncertainties and
 420 diagnosing product-specific omission and commission errors.

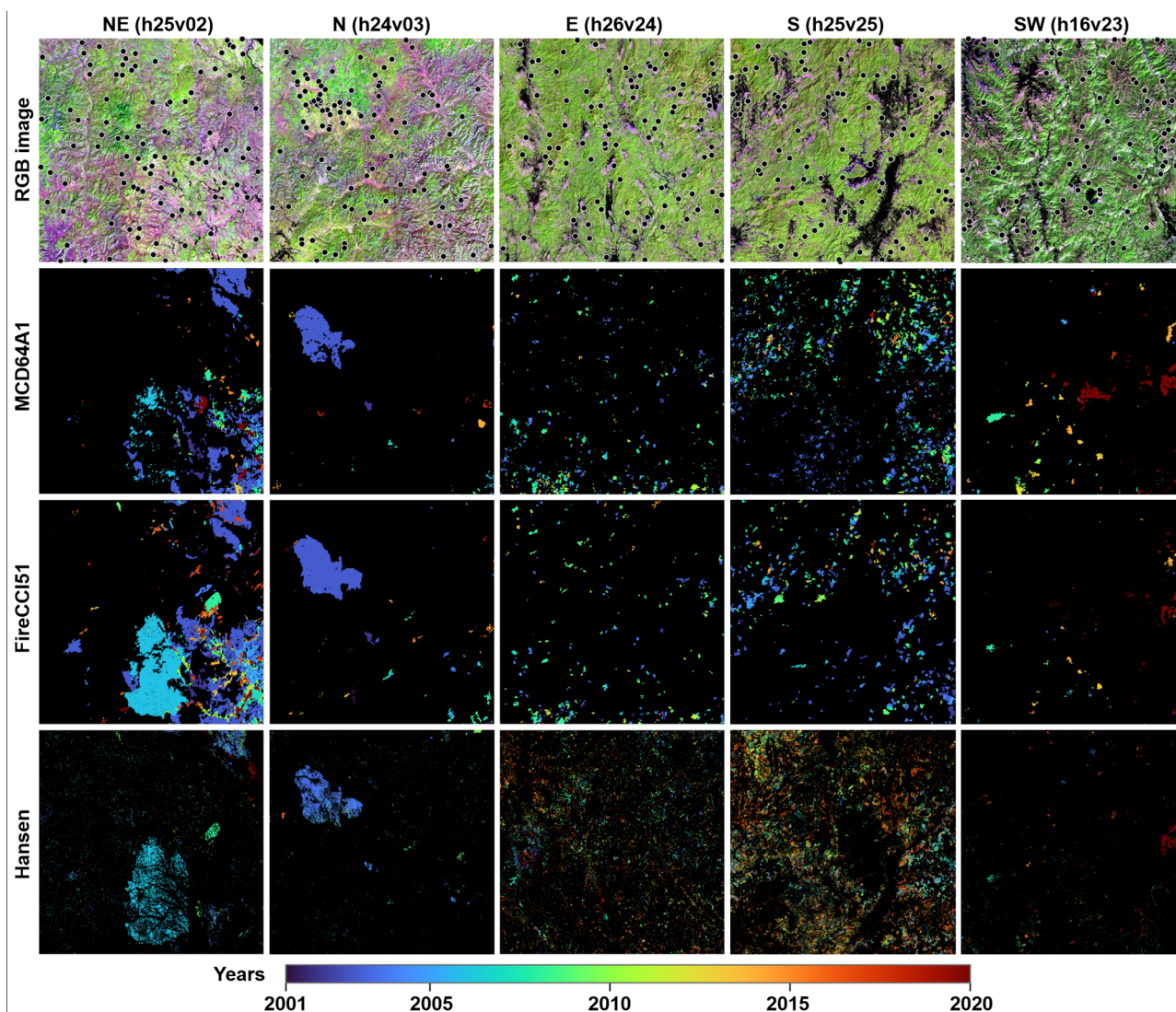


Figure 12: Maps of annual median composited Landsat RGB images (R: SWIR1, G: NIR, B: R) for the year 2021 and three widely used global forest disturbance products from 2001 to 2020 across five ARD tiles in China. Colour gradients represent the most recent disturbance year, and black dots overlaid on RGB composites denote RSDF samples. NE: Northeast China; E: East China; N: North China; S: South China; SW: Southwest China.

425

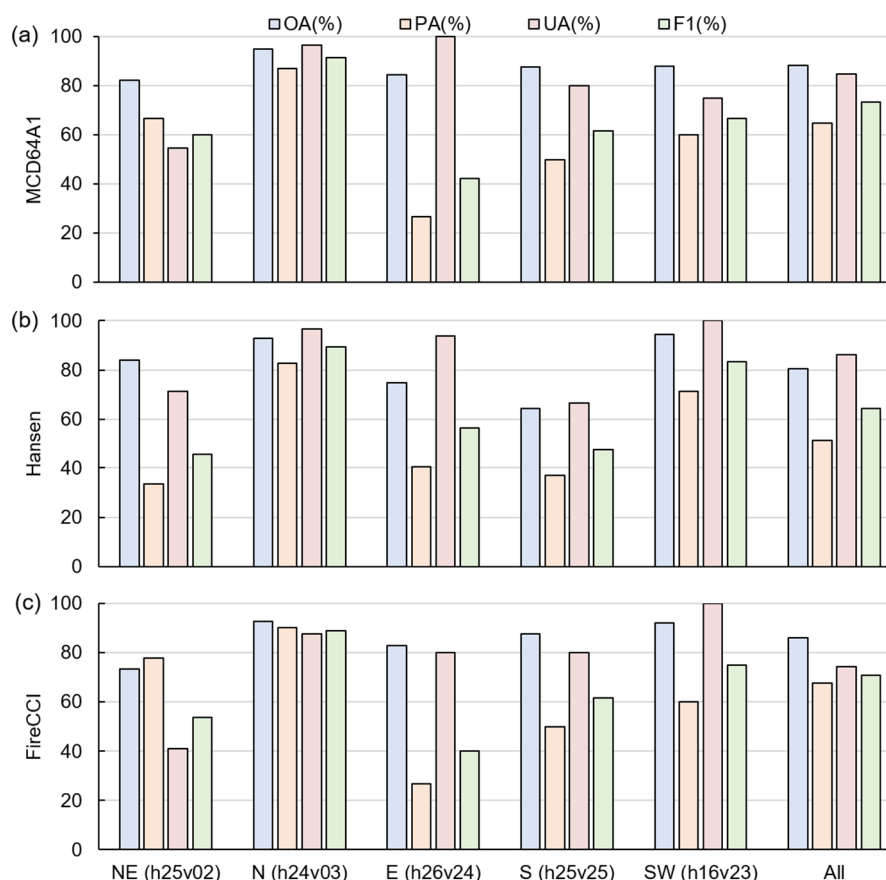


Figure 13: Validation of the MCD64A1 (a), Hansen (b), and FireCCI (b) products at five ARD tiles across China.

4.5 Labelling stable forest samples for diverse forest applications

Stable forest samples in the RSDF dataset provide ideal benchmarks for evaluating time-series reconstruction and forecasting, as they remain unaffected by land cover change and forest disturbances. Zhang et al. (2021) evaluated six widely used time-series models for reconstructing and forecasting dense Landsat time series using stable forest samples: harmonic (HR) (Brooks et al., 2012; Zhu et al., 2020), autoregressive (AR) (Poggi et al., 2003), linear Gaussian state-space (LGSS) (Tandeo et al., 2011), cubic spline (SP) (Chen et al., 2006; Xu et al., 2023), double logistic (DL) (Zhang et al., 2003; Shang et al., 2017), and asymmetric Gaussian (AG) (Che et al., 2014; Shao et al., 2016). Results shows that the HR model outperforms other models in time-series forecasting but exhibits limited performance for reconstruction. The DL and AG models achieve the highest reconstruction accuracy yet lack the capacity for time-series forecasting. Model performance is further modulated by data density and temporal irregularity: model accuracy improves substantially when valid clear-sky observations exceed seven per year, and data irregularity values below 1.0 produce more stable and reliable results (Zhang et al., 2021).



Beyond algorithm evaluation, these high-quality stable forest samples provide a unique basis for detecting long-term, gradual changes in forest ecosystems by eliminating confounding effects from forest disturbances or land cover conversions. Future studies can leverage these samples to identify and quantify persistent forest greening trends (Chen et al., 2019a; Yan et al., 2024), enhanced canopy vigour (Freeborn et al., 2022), improved carbon sequestration capacity (Chen et al., 2019b; Shang et al., 2023), and climate-driven shifts in forest productivity patterns (Yan et al., 2025), thereby supporting robust assessments of forest resilience and long-term carbon sink dynamics.

5 Discussions

This study developed a high-quality RSDF dataset for China spanning 1986 to 2025 via systematic visual interpretation, multi-expert validation, and standardized quality control protocols. The dataset provides precise records for forest disturbances and forestation activities, supported by complete Landsat time-series surface reflectance, spectral indices, and interpretable evidences. It has three distinct advantages: first, it supports inter-comparison, optimal parameterization, and accuracy validation of forest disturbance monitoring algorithms at national and regional scales; second, it enables robust training and validation of disturbance type classification models and young forest age mapping models; and third, it provides a large set of stable, undisturbed forest samples for evaluating time-series reconstruction and forecasting models and for detecting gradual long-term trends in forest ecosystems. Collectively, this dataset establishes a reliable benchmark for advancing high-precision forest disturbance monitoring, forest age mapping, and long-term forest carbon cycle research across China.

Despite rigorous interpretation and quality control procedures, the RSDF dataset has some limitations that deserve further improvement in future updates. First, forest disturbance classification (Qiu et al., 2025) remains coarse with unbalanced sample distributions across categories. The current version only distinguishes three generalized disturbance classes: harvest, fire, and other disturbances. The other disturbance class aggregates multiple discrete disturbance types, including wind damage, ice and snow breakage, landslides, drought, and insect infestations, without individual subtype differentiation. Furthermore, the sample sizes for rare disturbance types remain relatively limited. Future efforts should adopt a detailed hierarchical classification system to subdivide low-frequency disturbance types and conduct targeted supplementary sampling and interpretation to balance class distributions and improve subtype representativeness (Olofsson et al., 2014; Stehman and Foody, 2019).

Second, the dataset may contain subtle temporal sampling biases despite robust spatial representativeness. This study employed spatially-constrained stratified random sampling to proportionally distribute sample points across geographic strata, ensuring reliable spatial coverage of national forest disturbance patterns. However, forest disturbance events exhibit substantial interannual variability in frequency and spatial distribution over the 40-year time series (Liu et al., 2023; Shang et al., 2023), leading to uneven temporal sample coverage and mild representativeness gaps for specific periods. Future dataset iterations will integrate a spatiotemporally balanced sampling strategy to supplement samples for underrepresented time intervals, further enhancing the reliability of temporal trend extrapolation across decades.



470 Third, as a point-based sample dataset, RSDF may not be directly applicable for wall-to-wall estimation of national forest disturbance area. While point samples accurately record disturbance occurrence and attribution attributes, they cannot delineate the continuous spatial extent of individual disturbance patches. To mitigate this limitation, RSDF can be coupled with gridded forest disturbance products within a probabilistic sampling-inference framework (Qiu et al., 2025; Olofsson et al., 2014; Stehman and Foody, 2019). In this framework, RSDF samples serve as independent references to validate and calibrate wall-to-wall disturbance maps, enabling statistically rigorous national-scale disturbance area estimation with explicit confidence intervals.

Overall, the RSDF dataset provides a highly consistent, reliable, and practically valuable benchmark for forest disturbance monitoring and forest dynamics research in China. Future updates will further enhance its application potential by refining disturbance classification schemes, optimizing spatiotemporal sampling representativeness, and strengthening integration with gridded remote sensing products. Combined with wall-to-wall monitoring datasets, time-series decomposition algorithms, and carbon cycle simulation models, the RSDF dataset will support robust and innovative investigations of forest disturbance regimes, long-term vegetation greening trends, terrestrial carbon source–sink dynamics, and forest ecological responses to climate change across China.

6 Data availability

485 The Reference sample dataset of Stable, Disturbed and Forestated forests (RSDF) over China from 1986 to near present is publicly available at <https://doi.org/10.6084/m9.figshare.33313359> (Shang et al., 2026).

7 Conclusions

This study developed the RSDF, a nationwide reference sample dataset of stable, disturbed and forestated forests covering China from 1986 to 2025, to address the lack of consistent long-term ground truth samples for forest monitoring. We generated 12,991 representative 30 m sample points via spatially constrained stratified random sampling and implemented a multi-source hierarchical visual interpretation workflow combining Landsat time series, multi-platform high-resolution satellite imagery, and available forest inventory records. A novel sequential consensus mechanism integrated with multi-expert blind review was established for rigorous multi-stage quality control. Quality assessment confirms that over 95% of samples achieve high annotation confidence, with a mean inter-expert pairwise agreement of 0.83 ± 0.07 and an average Cohen's Kappa coefficient of 0.65 ± 0.13 . RSDF-based spatio-temporal analysis reveals pronounced nationwide forest change heterogeneity. Spatially, forest change intensity follows a southeast–northwest gradient, tightly coupled with provincial forest coverage. East and South China are the most dynamic zones, where stable forest drops below 51% and 71%, respectively, and harvests concentrate heavily. By contrast, North, Northeast, and Northwest China are dominated by stable forests ($>87\%$), with minimal disturbance activity. This pattern reflects a strong positive correlation between forest coverage and disturbance frequency. While harvests



500 dominate all regions, fires and other disturbances are geographically clustered in the east and south, and forestation events are sparse across all regions. Temporally, total disturbance intensity climbed from the late 1980s to the early 2010s then moderately declined. Long-term dynamics are primarily governed by harvests, whereas fires only trigger periodic peaks, and forestation maintains low frequencies. These quantified patterns deliver empirical evidence for understanding drivers of forest dynamics across China's diverse zones.

505 As an open-access benchmark, RSDF supports forest disturbance algorithm calibration, comparison and validation, forest age model training and validation, and assessment of global forest dynamic products. Its abundant stable forest samples further enable detection of long-term gradual changes by eliminating confounding effects from forest disturbances or land cover changes, thereby supporting robust assessments of forest resilience and carbon sink dynamics. This 40-year national dataset fills China's long-term forest reference data gap and provides core support for forest carbon modelling, ecological restoration
510 evaluation and sustainable forest management amid global change.

Author contributions

Conceptualization, R.S.; Methodology, R.S., S.X., Z.Y. and J.C.; Validation, R.S., S.X. and Z.Y.; Formal analysis, R.S., S.X., Z.Y. X.L., L.F., K.F., M.X., A.D., Y.Y., Y.L., C.S., W.C., J.Q., A.Z., J.H., S.C., X.Z., L.L., W.L., L.D., Z.H., and J.C.; Writing—original draft, R.S.; Writing—review & editing, R.S., J.C. and M.X.; Funding acquisition, R.S., W.L.
515 and Z.H..

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

We gratefully acknowledge all individuals who contributed data and provided valuable suggestions that helped improve this
520 study. During the preparation of this manuscript, the authors used ChatGPT 5.6 to improve the clarity and readability. After using this tool, the authors carefully reviewed and edited the manuscript as necessary and take full responsibility for its content.

Financial supports

This research was funded by the National Natural Science Foundation of China (42471356, 42522106, U23A2002, and 42101367) and Natural Science Foundation of Fujian Province (2025J09033).



525 References

- Brooks, E. B., Thomas, V. A., Wynne, R. H., and Coulston, J. W.: Fitting the multitemporal curve: A fourier series approach to the missing data problem in remote sensing analysis, *IEEE Transactions on Geoscience and Remote Sensing*, 50, 3340-3353, <https://doi.org/10.1109/TGRS.2012.2183137>, 2012.
- Che, M., Chen, B., Zhang, H., Fang, S., Xu, G., Lin, X., and Wang, Y.: A new equation for deriving vegetation phenophase
 530 from time series of Leaf Area Index (LAI) data, *Remote Sensing*, 6, 5650-5670, <https://doi.org/10.3390/rs6065650>, 2014.
- Chen, C., Park, T., Wang, X., Piao, S., Xu, B., Chaturvedi, R. K., Fuchs, R., Brovkin, V., Ciais, P., Fensholt, R., Tømmervik, H., Bala, G., Zhu, Z., Nemani, R. R., and Myneni, R. B.: China and India lead in greening of the world through land-use management, *Nature Sustainability*, 2, 122-129, <https://doi.org/10.1038/s41893-019-0220-7>, 2019a.
- Chen, J. M., Deng, F., and Chen, M.: Locally adjusted cubic-spline capping for reconstructing seasonal trajectories of a
 535 satellite-derived surface parameter, *IEEE Transactions on Geoscience and Remote Sensing*, 44, 2230-2237, <https://doi.org/10.1109/TGRS.2006.872089>, 2006.
- Chen, J. M., Ju, W., Ciais, P., Viovy, N., Liu, R., Liu, Y., and Lu, X.: Vegetation structural change since 1981 significantly enhanced the terrestrial carbon sink, *Nature Communications*, 10, 1-7, <https://doi.org/10.1038/s41467-019-12257-8>, 2019b.
- Cheng, K., Yang, H., Tao, S., Su, Y., Guan, H., Ren, Y., Hu, T., Li, W., Xu, G., Chen, M., Lu, X., Yang, Z., Tang, Y., Ma, K.,
 540 Fang, J., and Guo, Q.: Carbon storage through China's planted forest expansion, *Nature Communications*, 15, 4106, <https://doi.org/10.1038/s41467-024-48546-0>, 2024.
- Cohen, J.: A coefficient of agreement for nominal scales, *Educational and psychological measurement*, 20, 37-46, 1960.
- Fan, L., Wigneron, J.-P., Ciais, P., Chave, J., Brandt, M., Sitch, S., Yue, C., Bastos, A., Li, X., Qin, Y., Yuan, W., Schepaschenko, D., Mukhortova, L., Li, X., Liu, X., Wang, M., Frappart, F., Xiao, X., Chen, J., Ma, M., Wen, J., Chen,
 545 X., Yang, H., van Wees, D., and Fensholt, R.: Siberian carbon sink reduced by forest disturbances, *Nature Geoscience*, 1-7, <https://doi.org/10.1038/s41561-022-01087-x>, 2022.
- Franquesa, M., Stehman, S. V., and Chuvieco, E.: Assessment and characterization of sources of error impacting the accuracy of global burned area products, *Remote Sensing of Environment*, 280, 113214, <https://doi.org/10.1016/j.rse.2022.113214>, 2022.
- 550 Freeborn, P. H., Jolly, W. M., Cochrane, M. A., and Roberts, G.: Large wildfire driven increases in nighttime fire activity observed across CONUS from 2003–2020, *Remote Sensing of Environment*, 268, 112777, <https://doi.org/10.1016/j.rse.2021.112777>, 2022.
- Giglio, L., Justice, C., Boschetti, L., and Roy, D.: MODIS/terra+ aqua burned area monthly L3 global 500m SIN grid V061, NASA EOSDIS Land Processes Distributed Active Archive Center (DAAC) data set, MCD64A61. 061, 2021.
- 555 Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., and Townshend, J. R. G.: High-resolution global maps of 21st-century forest cover change, *Science*, 342, 850-853, <https://doi.org/10.1126/science.1244693>, 2013.



- 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 Sensing of Environment*, 114, 183-198, <https://doi.org/10.1016/j.rse.2009.08.017>, 2010.
- 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 Sensing of Environment*, 114, 2897-2910, <https://doi.org/10.1016/j.rse.2010.07.008>, 2010.
- Liang, Y., Shang, R., Chen, J. M., Lin, X., Li, P., Yang, Z., Fan, L., Xu, S., Lin, Y., and Chen, Y.: Comprehensive Comparison and Validation of Forest Disturbance Monitoring Algorithms Based on Landsat Time Series in China, *Remote Sensing*, 17, 680, <https://doi.org/10.3390/rs17040680>, 2025.
- Liao, Z., Yue, C., He, B., Zhao, K., Ciais, P., Alkama, R., Grassi, G., Sitch, S., Chen, R., Quan, X., Xu, M., and Wang, M.: Growing biomass carbon stock in China driven by expansion and conservation of woody areas, *Nature Geoscience*, 17, 1127-1134, <https://doi.org/10.1038/s41561-024-01569-0>, 2024.
- Liu, M., Trugman, A. T., Peñuelas, J., and Anderegg, W. R. L.: Climate-driven disturbances amplify forest drought sensitivity, *Nature Climate Change*, 14, 746-752, <https://doi.org/10.1038/s41558-024-02022-1>, 2024.
- Liu, Z., Wang, W. J., Ballantyne, A., He, H. S., Wang, X., Liu, S., Ciais, P., Wimberly, M. C., Piao, S., Yu, K., Yao, Q., Liang, Y., Wu, Z., Fang, Y., Chen, A., Xu, W., and Zhu, J.: Forest disturbance decreased in China from 1986 to 2020 despite regional variations, *Communications Earth and Environment*, 4, 15, <https://doi.org/10.1038/s43247-023-00676-x>, 2023.
- Lizundia-Loiola, J., Otón, G., Ramo, R., and Chuvieco, E.: A spatio-temporal active-fire clustering approach for global burned area mapping at 250 m from MODIS data, *Remote Sensing of Environment*, 236, 111493, <https://doi.org/10.1016/j.rse.2019.111493>, 2020.
- Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A.: Good practices for estimating area and assessing accuracy of land change, <https://doi.org/10.1016/j.rse.2014.02.015>, 2014.
- Poggi, P., Muselli, M., Notton, G., Cristofari, C., and Louche, A.: Forecasting and simulating wind speed in Corsica by using an autoregressive model, *Energy Conversion and Management*, 44, 3177-3196, [https://doi.org/10.1016/S0196-8904\(03\)00108-0](https://doi.org/10.1016/S0196-8904(03)00108-0), 2003.
- Qiu, D., Liang, Y., Shang, R., and Chen, J. M.: Improving LandTrendr Forest Disturbance Mapping in China Using Multi-Season Observations and Multispectral Indices, *Remote Sensing*, 15, 2381, <https://doi.org/10.3390/rs15092381>, 2023.
- Qiu, S., Zhu, Z., Yang, X., Woodcock, C. E., Fahey, R. T., Stehman, S., Zhang, Y., Cullerton, M., Grinstead, A., Hong, F., Song, K., Suh, J. W., Li, T., Ren, W., and Nemani, R. R.: A shift from human-directed to undirected wild land disturbances in the USA, *Nature Geoscience*, 18, 989-996, <https://doi.org/10.1038/s41561-025-01792-3>, 2025.
- Seidl, R., Thom, D., Kautz, M., Martin-Benito, D., Peltoniemi, M., Vacchiano, G., Wild, J., Ascoli, D., Petr, M., Honkaniemi, J., Lexer, M. J., Trotsiuk, V., Mairota, P., Svoboda, M., Fabrika, M., Nagel, T. A., and Reyer, C. P. O.: Forest disturbances under climate change, *Nature Climate Change*, 7, 395-402, <https://doi.org/10.1038/nclimate3303>, 2017.



- Shang, R., Liu, R., Xu, M., Liu, Y., Zuo, L., and Ge, Q.: The relationship between threshold-based and inflexion-based approaches for extraction of land surface phenology, *Remote Sensing of Environment*, 199, 167-170, <https://doi.org/10.1016/j.rse.2017.07.020>, 2017.
- Shang, R., Zhu, Z., Zhang, J., Qiu, S., Yang, Z., Li, T., and Yang, X.: Near-real-time monitoring of land disturbance with harmonized Landsats 7–8 and Sentinel-2 data, *Remote Sensing of Environment*, 278, 113073, <https://doi.org/10.1016/j.rse.2022.113073>, 2022.
- Shang, R., Chen, J. M., Xu, M., Lin, X., Li, P., Yu, G., He, N., Xu, L., Gong, P., and Liu, L.: China's current forest age structure will lead to weakened carbon sinks in the near future, *The Innovation*, 4, 10515, <https://doi.org/10.1016/j.xinn.2023.100515>, 2023.
- Shang, R., Yang, Z., Liang, Y., Chen, J. M., Zhu, Z., Cao, G., Fang, K., Lin, X., Liu, L., and Li, J.: Mapping annual forest disturbance from 1986 to 2021 at 30-m resolution in China using the modified COLD algorithm, Available at SSRN 5207098, 2025a.
- Shang, R., Lin, X., Chen, J. M., Liang, Y., Fang, K., Xu, M., Yan, Y., Ju, W., Yu, G., He, N., Xu, L., Liu, L., Li, J., Li, W., Zhai, J., and Hu, Z.: China's annual forest age dataset at a 30 m spatial resolution from 1986 to 2022, *Earth Syst. Sci. Data*, 17, 3219-3241, <https://doi.org/10.5194/essd-17-3219-2025>, 2025b.
- Shang, R., Xu, S., Yang, Z., and Chen, J. M.: A reference sample dataset of stable, disturbed and forestated forests over China from 1986 to near present, figshare [data set], <https://doi.org/10.6084/m9.figshare.33313359>, 2026.
- Shao, Y., Lunetta, R. S., Wheeler, B., Iames, J. S., and Campbell, J. B.: An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data, *Remote Sensing of Environment*, 174, 258-265, <https://doi.org/10.1016/j.rse.2015.12.023>, 2016.
- Stehman, S. V. and Foody, G. M.: Key issues in rigorous accuracy assessment of land cover products, *Remote Sensing of Environment*, 231, 111199, <https://doi.org/10.1016/j.rse.2019.05.018>, 2019.
- Tandeo, P., Ailliot, P., and Autret, E.: Linear Gaussian state-space model with irregular sampling: Application to sea surface temperature, *Stochastic Environmental Research and Risk Assessment*, 25, 793-804, <https://doi.org/10.1007/s00477-010-0442-8>, 2011.
- Venter, O., Sanderson, E. W., Magrath, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., Laurance, W. F., Wood, P., Fekete, B. M., Levy, M. A., and Watson, J. E. M.: Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation, *Nature Communications*, 7, <https://doi.org/10.1038/ncomms12558>, 2016.
- Verbesselt, J., Hyndman, R., Newnham, G., and Culvenor, D.: Detecting trend and seasonal changes in satellite image time series, *Remote Sensing of Environment*, 114, 106-115, <https://doi.org/10.1016/j.rse.2009.08.014>, 2010.
- Xu, M., Shang, R., Chen, J. M., and Zeng, L.: LACC2. 0: Improving the LACC Algorithm for Reconstructing Satellite-Derived Time Series of Vegetation Biochemical Parameters, *Remote Sensing*, 15, 3277, <https://doi.org/10.3390/rs15133277>, 2023.



- Yan, Y., Piao, S., Hammond, W. M., Chen, A., Hong, S., Xu, H., Munson, S. M., Myneni, R. B., and Allen, C. D.: Climate-induced tree-mortality pulses are obscured by broad-scale and long-term greening, *Nature Ecology & Evolution*, 8, 912-923, <https://doi.org/10.1038/s41559-024-02372-1>, 2024.
- Yan, Y., Li, B., Dechant, B., Xu, M., Luo, X., Qu, S., Miao, G., Leng, J., Shang, R., Shu, L., Jiang, C., Wang, H., Jeong, S., Ryu, Y., and Chen, J. M.: Plant traits shape global spatiotemporal variations in photosynthetic efficiency, *Nature Plants*, <https://doi.org/10.1038/s41477-025-01958-2>, 2025.
- Yang, Z., Shang, R., Liang, Y., Xu, M., Hu, Z., and Chen, J. M.: Quantifying the efficacy of topographic correction for forest disturbance monitoring using Landsat time series, *International Journal of Applied Earth Observation and Geoinformation*, Under review, 2026.
- Yao, L., Liu, T., Qin, J., Jiang, H., Yang, L., Smith, P., Chen, X., Zhou, C., and Piao, S.: Carbon sequestration potential of tree planting in China, *Nature Communications*, 15, 8398, <https://doi.org/10.1038/s41467-024-52785-6>, 2024.
- Yasmin, F., Ahmed, R., and Akhtar, M.: Construction of Balanced Incomplete Block Designs Using Cyclic Shifts, *Communications in Statistics - Simulation and Computation*, 44, 525-532, <https://doi.org/10.1080/03610918.2013.784984>, 2015.
- Yu, Z., Ciais, P., Piao, S., Houghton, R. A., Lu, C., Tian, H., Agathokleous, E., Kattel, G. R., Sitch, S., Goll, D., Yue, X., Walker, A., Friedlingstein, P., Jain, A. K., Liu, S., and Zhou, G.: Forest expansion dominates China's land carbon sink since 1980, *Nature Communications*, 13, 5374, <https://doi.org/10.1038/S41467-022-32961-2>, 2022.
- Zhang, J., Shang, R., Rittenhouse, C., Witharana, C., and Zhu, Z.: Evaluating the impacts of models, data density and irregularity on reconstructing and forecasting dense Landsat time series, *Science of Remote Sensing*, 4, 100023, <https://doi.org/10.1016/j.srs.2021.100023>, 2021.
- Zhang, X., Zhao, T., Xu, H., Liu, W., Wang, J., Chen, X., and Liu, L.: GLC_FCS30D: the first global 30 m land-cover dynamics monitoring product with a fine classification system for the period from 1985 to 2022 generated using dense-time-series Landsat imagery and the continuous change-detection method, *Earth System Science Data*, 16, 1353-1381, 2024a.
- Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., Reed, B. C., and Huete, A.: Monitoring vegetation phenology using MODIS, *Remote Sensing of Environment*, 84, 471-475, [https://doi.org/10.1016/S0034-4257\(02\)00135-9](https://doi.org/10.1016/S0034-4257(02)00135-9), 2003.
- Zhang, Y., Wang, J. A., Berner, L. T., Goetz, S. J., Zhao, K., and Liu, Y.: Warming and disturbances affect Arctic-boreal vegetation resilience across northwestern North America, *Nature Ecology & Evolution*, <https://doi.org/10.1038/s41559-024-02551-0>, 2024b.
- Zhu, Y., Xia, X., Canadell, J. G., Piao, S., Lu, X., Mishra, U., Wang, X., Yuan, W., and Qin, Z.: China's carbon sinks from land-use change underestimated, *Nature Climate Change*, 15, 428-435, <https://doi.org/10.1038/s41558-025-02296-z>, 2025.
- Zhu, Z. and Woodcock, C. E.: Continuous change detection and classification of land cover using all available Landsat data, *Remote Sensing of Environment*, 144, 152-171, <https://doi.org/10.1016/j.rse.2014.01.011>, 2014.



Zhu, Z., Zhang, J., Yang, Z., Aljaddani, A. H., Cohen, W. B., Qiu, S., and Zhou, C.: Continuous monitoring of land disturbance based on Landsat time series, *Remote Sensing of Environment*, 238, 111116, <https://doi.org/10.1016/j.rse.2019.03.009>, 2020.