Annual vegetation maps in Qinghai-Tibet Plateau (QTP) from 2000 to 2022 based on MODIS series satellite imagery

Guangsheng Zhou[1†], Hongrui Ren[2†], Lei Zhang[2], Xiaomin Lv[1], and Mengzi Zhou[1]

[1]State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081, China
[2]Department of Geomatics, Taiyuan University of Technology, Taiyuan 030024, China

Correspondence: Guangsheng Zhou (zhougs@cma.gov.cn)

† Co-first authors

Abstract. The Qinghai Tibet Plateau (QTP), known as the "Third Pole of the Earth" and the "Water Tower of Asia", plays a crucial role in global climate regulation, biodiversity conservation, and regional socio-economic development. Continuous annual vegetation types and their geographical distribution data are essential for studying the response and adaptation of vegetation to climate change. However, there is very limited data on vegetation types and their geographical distributions on the QTP due to harsh natural environment. Currently, land cover/surface vegetation (LCSV) data are typically obtained using independent classification methods for each period's product, based on remote sensing information. These approaches do not consider the temporal continuity of vegetation presence, leading to a gradual increase in misclassified pixels and uncertainty in their locations, consequently decreasing the interpretability of the long-time series remote sensing products. To address this issue, this study developed a new approach to long-time continuous annual vegetation mapping from remote sensing imagery, and mapped the vegetation of the QTP from 2000 to 2022 at a 500 m spatial resolution through the MOD09A1 product. The overall accuracy of continuous annual QTP vegetation mapping from 2000 to 2022 reached 83.27%, with the reference annual 2020 reaching an accuracy of 83.32% and a Kappa coefficient of 0.82. This study supports the use of remote sensing data for long-term continuous annual vegetation mapping.

1 Introduction

Vegetation, an integral component of Earth's ecosystems, plays an irreplaceable role in maintaining climate stability, preserving biodiversity, and supplying vital resources for humans. Vegetation maps not only facilitate a visual comprehension of vegetation types and their geographic distribution, but also provide essential data for natural resource management and environmental protection (Immerzeel et al., 2010). Particularly, long-time series of geographical distribution data on vegetation types are crucial for revealing the impacts of climate change and human activities on vegetation, elucidating the succession processes of vegetation, and scientifically managing vegetation ecosystems.

The availability of large-scale, long-term, and free remote sensing imagery has significantly advanced the development of land cover and surface vegetation (LCSV) data. Notable international products include the European Space Agency (ESA)'s

30   GlobCover_2005 and GlobCover_2009 (300 m) (Bontemps et al., 2011), and National Aeronautics and Space Administration (NASA)'s annual MCD12Q1 product since 2001 (500 m) (Friedl et al., 2010; Sulla-Menashe et al., 2019). In China, representative products include GlobeLand30 by Chen et al. (2015), Finer Resolution Observation and Monitoring of Global Land Cover (FROM_GLC)10 and FROM_GLC30 by Tsinghua University (Gong et al., 2013; Gong et al., 2019), Global 30m Land-cover Classification with a Fine Classification System (GLC_FCS30) and Global 30m Land-cover

35   Dynamics Monitoring Dataset (GLC_FCS30D) by the Chinese Academy of Sciences (Zhang et al., 2019b, Zhang et al., 2023), and China Land Cover Dataset (CLCD) by Wuhan University (Yang and Huang, 2021). Leveraging the intelligent remote sensing mapping (iMap) concept and framework, there has been rapid progress in developing global-scale seamless daily data cubes and in creating annual and seasonal land surface maps (Feng and Li, 2020; Liu et al., 2021).

    In traditional long-time series LCSV datasets, it is common to classify each period's product independently. For example,

40   ESA produced the global 300 m land cover products GlobCover_2005 and GlobCover_2009 using a combination of supervised and unsupervised classification methods (Bontemps et al., 2011). However, there are significant spatial distribution differences between GlobCover_2009 and GlobCover_2005, making these two land cover maps not directly comparable and unsuitable for change detection studies (Bontemps et al., 2011). NASA, using MODIS data, produced the annual MCD12Q1.v5 product using the decision tree classifier (Friedl et al., 2010). A significant drawback of MCD12Q1.v5

45   is the instability of land cover labels. The product shows an annual change of approximately 11.4% in land cover labels between 2001 and 2013. The MCD12Q1.v6 product, which uses the Hidden Markov Model (HMM) for post-processing, reduced this instability to 1.6% (Sulla-Menashe et al., 2019). The strategy of independently classifying each period often leads to significant spurious changes between products from different periods, further undermining their reliability (Liu et al., 2021). The temporal stability of land cover products generated using such methods largely depends on post-processing

50   techniques, such as the HMM (Sulla-Menashe et al., 2019) and spatial-temporal consistency methods (Yang and Huang, 2021). Therefore, there is an urgent need to develop new methods that can effectively integrate spatial-temporal information during the classification process, to overcome the limitations of existing independent classification and post-processing approaches. This would make better use of publicly available, free, long-time series data, and improve the temporal consistency and reliability of land cover products.

55   The QTP, known as the "Roof of the World" and the "Water Tower of Asia," is critical for global climate regulation and regional socio-economic development. In the 1970s, China carried out the first extensive scientific survey of the QTP. Currently, the second QTP scientific expedition is underway, with vegetation survey being one of the major components. Understanding the evolution of vegetation types on the QTP is important for revealing the effects of climate change on vegetation structure and function. Additionally, such insights are essential for elucidating the carbon and water cycles of the

60   QTP and for formulating high-quality, sustainable development strategies for the region amidst global warming.

    As the "Third Pole" of the Earth, the QTP has very limited data on vegetation types and geographical distribution due to its harsh natural environment. Currently, the primary vegetation data for the QTP include the "Vegetation map of Qinghai Tibet Plateau in 2020 with 10 m spatial resolution" (Zhou et al., 2022a), the "A new vegetation map for Qinghai-Tibet

Plateau by integrated classification from multi-source data products (2020)" (Zhang et al., 2022), and the "Vegetation map of Qinghai Tibet Plateau in 1980s" (Zhou et al., 2022b). These vegetation distribution products are single-period maps, which are insufficient to depict dynamic changes in vegetation, thus limiting the understanding of vegetation evolution trends and mechanisms on the QTP. Therefore, many scholars have conducted thematic mapping studies on long-time series of LCSV data of the QTP, such as wetland, water, and glacier (Li et al., 2023; Zhang et al., 2019a; Hu et al., 2022). However, these long-time series products, created with independent classification methods, are primarily focused on specific LCSV types, overlooking the need to identify temporal changes across different types, which limits the accuracy of the products. Therefore, in order to provide continuous data support for research on the interaction between vegetation and climate change, there is an urgent need for long-time series vegetation mapping.

This study aims to develop a new approach to long-time series vegetation mapping using remote sensing imagery, and to map vegetation of the entire QTP at 500 m from 2000 to 2022, using the MOD09A1 remote sensing data.

2 Materials and Methods

2.1 Study Area

The QTP, situated between 25°59′30″ N and 40°1′0″ N, 67°40′37″ E and 104°40′57″ E, has an average elevation of about 4320 m and a total area of 3.08 million km² (Zhang et al., 2021b). This study covers the entire QTP, spanning six provinces and regions within China - Tibet, Qinghai, Gansu, Sichuan, Yunnan, and Xinjiang - as well as areas in India, Pakistan, Tajikistan, Afghanistan, Nepal, Bhutan, Myanmar, and Kyrgyzstan. In particular, the area of the QTP within China is about 2.58 million km² (approximately 83.7% of the QTP) with an average elevation of about 4400 m (Yu et al., 2014; Zhang, 2019; Zhang et al., 2021c; Zhang et al., 2021d). The QTP slopes downward from the high northwest to the lower southeast, with a humid and rainy climate in the southeast and arid conditions in the northwest. The vegetation distribution across the QTP is influenced by topography and climatic conditions, revealing a clear horizontal zonation ranging from forests, shrubs, and meadows in the southeast to grasslands and deserts in the northwest. Additionally, as temperatures decrease with increasing altitude, there is a distinct vertical stratification in vegetation, ranging from forests at lower elevations to permanent glaciers, snow cover, and permafrost at higher altitudes.

2.2 Data Sources

2.2.1 Training Sample Data for Vegetation Type

The vegetation of the QTP is primarily categorized into 15 types (Zhou et al., 2023). Given the significance of glaciers and snow cover, this study expanded the vegetation classification to 16 types for the 500 m spatial resolution mapping, including evergreen broad-leaved forest (EBF), evergreen coniferous forest (ECF), coniferous and broad-leaved mixed forest (CBMF), deciduous broad-leaved forest (DBF), deciduous coniferous forest (DCF), scrub (SC), alpine scrub meadow (ASM), alpine

meadow (AM), alpine grassland (AG), alpine vegetation (AV), alpine desert (AD), cultivated vegetation (CV), wetland (WE), water (WA), non-vegetated area (NVA), and glacier and snow (GS).

The training vegetation samples for the QTP within China were obtained from the "Vegetation map of Qinghai Tibet Plateau in 2020 with 10 m spatial resolution" (Zhou et al., 2022a). This product was created using a regional vegetation mapping method based on terrain-climate-remote sensing information, with a spatial resolution of 10 m and an overall accuracy of 89.5%. Its classification system is consistent with that of this study, except for the exclusion of GS, and was therefore used to generate the vegetation type training sample data for this research (Zhou et al., 2023). The vegetation map was resampled to 500 m. Subsequently, the dominant vegetation type and its proportional area, derived from the corresponding 50×50 pixels at 10 m spatial resolution, were determined as the vegetation type categorization and pixel purity (reflected as the percentage of the dominant vegetation type within these 50×50 pixels) for the 500 m spatial resolution map. Concurrently, different purity levels and their area proportions were calculated (Table 1). Random sampling was then conducted in areas with purity levels exceeding 70%, representing approximately 62.34% of the QTP. For the QTP regions outside China, vegetation type sample data were obtained from high-resolution Google Earth imagery from 2020. By visually interpreting these images, samples were selected from areas with consistent vegetation types within a 500 m range. Using these methods, a total of 8,937 training samples were collected to train the 2020 vegetation classification model (Fig. 1). The sample distribution included: EBF (936 points), ECF (768 points), CBMF (475 points), DBF (570 points), DCF (370 points), SC (397 points), ASM (213 points), AM (1,375 points), AG (872 points), AV (503 points), AD (499 points), CV (469 points), WE (246 points), WA (377 points), NVA (204 points), and GS (663 points).

Table 1. Purity and area proportion of 500 m resolution pixel samples of vegetation types on the QTP within China.

Purity (%)	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Area Proportion (%)	0.02	0.04	1.01	4.20	8.28	12.55	11.56	11.40	12.55	38.39

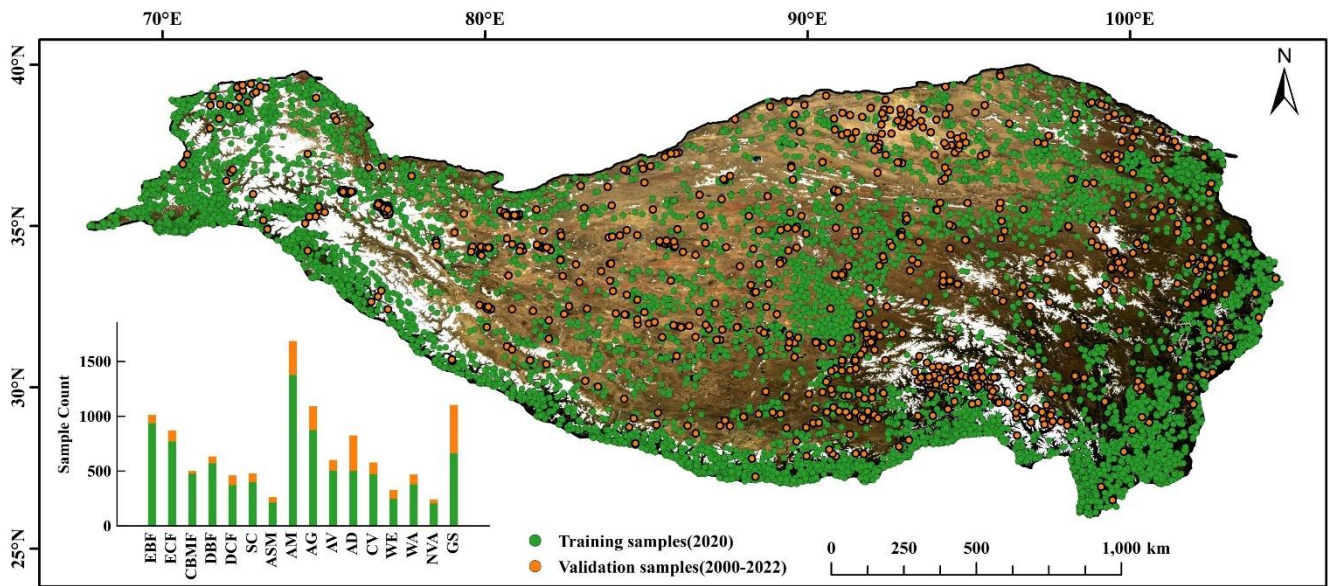



Figure 1. Spatial distribution of training (2020) and validation (2000-2022) samples for vegetation mapping at 500 m resolution on the QTP.

2.2.2 Validation Sample Data for Vegetation Type

As the "Third Pole of the Earth ", the QTP has high altitudes and sparse human activity, making it difficult to obtain validation samples. The validation samples in this study consist of three parts: field measurements, visual interpretation, and third-party samples.

The field measurement samples include 173 samples collected during field expeditions on the QTP in 2019, 2020, and 2021. These samples were collected during the vegetation growing season on the QTP, with neighboring samples spaced approximately 50 km apart. Each sample is ensured to have uniform and consistent vegetation distribution within a 1 km×1 km area. The visual interpretation samples (1002 in total) were obtained from previous work by Wu et al. (2024). This sample set was obtained through visual interpretation on Google Earth and Google Earth Engine (GEE). The selected samples are all located in areas where the vegetation distribution types remained consistent from 1990 to 2020, and each sample covers a spatial extent greater than 500 m. Each sample was independently interpreted regarding its vegetation type by three interpreters, using long-time series temperature and precipitation data as well as Landsat remote sensing images. For samples with conflicting results, the second interpretation was conducted; if inconsistencies remained, the sample was abandoned.

The third-party samples include the First All-season Sample Set produced by Li et al. (2017), the Global Land Cover Validation Samples collected by Zhang et al. (2021a), and validation samples gathered through literature search. The First All-season Sample Set was obtained through visual interpretation with Global Mapper software using Landsat 8 data from 2013, 2014, and 2015, including approximately 140,000 validation samples worldwide. This sample set uses the FROM-

GLC classification system, which contains 12 primary classes and 30 secondary classes. The second sample set is derived

135 from various auxiliary datasets, including Global Cropland Reference Data, Global Observation for Forest Cover and Land Dynamics reference data, the FROM_GLC global validation sample set, and the Global Lakes and Wetlands Database. This set includes 44,043 validation samples covering the globe. It is based on detailed land cover types, including 24 types, primarily from around 2015, and was published as a validation dataset for the GLC-FCS30-2015 product. The literature search samples were obtained from the China National Knowledge Infrastructure (https://www.cnki.net/) and Web of

140 Science (https://www.webofscience.com/) databases using keywords such as "Qinghai-Tibet Plateau," "vegetation", and "vegetation cover" to collect published validation samples from 2000 to 2022.

Unlike the validation samples obtained from field expeditions and visual interpretation, the quality of third-party validation sample data carries uncertainties, such as the spatial range and spatial homogeneity. Li et al. (2017) provided the unit sizes for each sample in the First All-season Sample Set, including 30 m, 100 m, 250 m, 500 m, and 1000 m,

145 representing 1×1, 3×3, 9×9, 17×17, and 33×33 Landsat pixel ranges, respectively. Among them, samples with unit sizes greater than 500 m are rare, accounting for only 10.6%. Samples with unit sizes less than 500 m are difficult to apply directly for validating the 500 m vegetation maps in this study. Additionally, the Global Land Cover Validation Samples from Zhang et al. (2021a) and the samples collected through literature searches can only ensure the accuracy of sampling time and type, without confirm their spatial extent. To address this issue, this study employed the method proposed by Feng et al. (2012) to

150 extract the patially homogeneous regions of MOD09A1 data. Specifically, at the spatial scale, this study statistically analyzed the maximum and minimum values of nine 500 m resolution pixels in a 3×3 grid across various bands and calculated the range as an indicator of homogeneity. For the blue, green, red, near-infrared, shortwave infrared 1, and shortwave infrared 2 bands, thresholds of 0.03, 0.03, 0.03, 0.06, 0.03, and 0.03 were used, respectively. If the range for all bands fell within the threshold, the corresponding area was deemed homogeneous. This method ensures that there is at most

155 one representative sample within a 500m spatial extent for third-party validation samples, and that it lies within a 1.5 km×1.5 km spatially homogeneous region. GS on the QTP are mostly classified as heterogeneous areas. Therefore, the samples labeled as GS in the third-party samples are visually interpreted to remove unreasonable ones. After the above quality control, a total of 1,014 third-party validation samples were obtained, including 327 from the global all-season sample library, 499 from the GLC-FCS30-2015 global validation sample set, and 188 from the literature search samples.

160 The validation sample dataset for this product contains 2,189 samples from field measurements, visual interpretation, and third-party samples. Among them, 173 field measurement samples are available for validating the 2020 product There are 1,002 samples from visual interpretation, characterized by stable attributes over multiple years, making them suitable for validating annual products from 2000 to 2020. Furthermore, 1,014 third-party validation samples can be used to validate annual products from 2000 to 2019.

2.2.3 Remote Sensing Data

Remote sensing data in this study were sourced from the Moderate Resolution Imaging Spectroradiometer (MODIS) series satellite imagery provided by NASA. As one of the longest-operating Earth observation instruments currently in orbit, MODIS is a multi-spectral observation sensor launched in May 1999, enabling efficient and comprehensive observation of the Earth. The MOD09A1 surface reflectance dataset was obtained through the GEE platform and had been processed with atmospheric and topographic corrections (Gorelick et al., 2017). The MOD09A1 dataset provides surface reflectance in seven spectral bands (Red, Blue, Green, NIR, MIR, SWIR 1, and SWIR 2) with 500 m spatial resolution, and all cloud-contaminated pixels are removed. There were 1051 MOD09A1 remote sensing images from January 1, 2000, to December 31, 2022, selected for this study, which were converted from the original sinusoidal projection to the WGS84 geographic coordinate grid for the preparation of annual vegetation maps at 500 m of the QTP from 2000 to 2022.

2.2.4 Climate and Terrain Data

The Digital Elevation Model (DEM) and derived features are crucial for vegetation mapping. Additionally, the distribution of vegetation on the QTP demonstrates distinct vertical zonation, greatly influenced by climate and topography (Zhou et al., 2023). This study utilized climate data which included annual precipitation (AP) and annual average temperature (AT) across the entire QTP from 2000 to 2022. For the QTP within China, climate data at 1,000 m were obtained from the National Tibetan Plateau Data Center, specifically the "1-km monthly precipitation dataset for China (1901-2023)" and the "1-km monthly mean temperature dataset for China (1901-2023)" (Ding and Peng, 2020; Peng et al., 2017a; Peng et al., 2017b; Peng, 2019; Peng et al., 2019; Peng, 2020). To derive the precipitation and temperature features required for this study, the monthly precipitation data were summed to calculate annual totals, while the monthly mean temperature data were averaged over the year. In contrast, climate data for areas of the QTP outside China were derived from the CRU high-resolution gridded dataset, featuring a spatial resolution of approximately 50,000 m. The terrain data was from the SRTM by the USGS (Farr et al., 2007), with a spatial resolution of 30 m. At last, this study applied the mean sampling method in GEE to resample data on AP, AT, elevation, slope, and aspect variables derived from SRTM data to a 500 m spatial resolution for integration into vegetation mapping.

2.3 Vegetation Mapping of the QTP at 500 m Resolution in 2020

2.3.1 Classification Platform and Algorithm

The GEE cloud platform (Gorelick et al., 2017) offers a variety of machine learning models, such as SVM and RF, the latter of which was utilized in this study for vegetation mapping. The RF model (Breiman, 2001) is an ensemble learning-based machine learning algorithm that mitigates the overfitting risk inherent in single decision trees by constructing and integrating multiple decision trees. This approach not only enhances the accuracy and stability of predictions but has also been widely applied in the vegetation mapping.

2.3.2 Construction of Vegetation Mapping Features

The features used in vegetation mapping are divided into four categories (Table 2): terrain (elevation, slope, aspect), climate (AT and AP), surface reflectance (R, N, B, G, M, S1, S2), and 14 index features, which are constructed from the single-band surface reflectance. These features were derived from the MOD09A1 remote sensing imagery data spanning from January 1, 2020, to December 31, 2020. Additionally, six percentiles—15%, 30%, 45%, 60%, 75%, and 90%—were calculated for the 7 reflectance bands and the 14 indices, representing the time series characteristics of each pixel. The 15% and 90% percentiles were used as substitutes for the minimum and maximum values of the time series observations, effectively mitigating the influence of extreme values in time-series data (Zhang et al., 2021a). The 30%, 45%, 60%, and 75% percentile features were selected to capture temporal trends uniformly while avoiding feature redundancy (Sulla-Menashe et al., 2019; Zhang et al., 2024). A total of 131 features were formulated from these 4 categories for vegetation mapping on the QTP.

Table 2. Vegetation mapping features at 500 m resolution on the QTP.

Category	Features	Description
Terrain	Elevation	
	Slope	
	Aspect	
Climate	AT	Annual average temperature
	AP	Annual precipitation
Surface reflectance	R	Red
	N	NIR
	B	Blue
	G	Green
	M	MIR
	S1	Swir1
	S2	Swir2
Vegetation Index	NDVI	$\dfrac{N - R}{N + R}$
	EVI	$2.5\dfrac{N - R}{N + 6R - 7.5B + 1}$
	RVI	$\dfrac{N}{R}$
	DVI	$N - R$
	SAVI	$\dfrac{(N - R) * 1.5}{N + R + 0.5}$
	GCVI	$\dfrac{N}{G} - 1$

	NIRV	$\dfrac{(N - R) * N}{N + R}$
Urban Index	NDBI	$\dfrac{S1 - N}{S1 + N}$
	IBI	$\dfrac{NDBI - (SAVI + (G - S1)/(G + S1))/2}{NDBI + (SAVI + (G - S1)/(G + S1))/2}$
Water Index	NDWI	$\dfrac{G - N}{G + N}$
	LSWI	$\dfrac{N - S1}{N + S1}$
Snow Index	NDSI	$\dfrac{G - S1}{G + S1}$
	NDGlaI	$\dfrac{G - R}{G + R}$
Soil Index	BI	$\dfrac{(S1 + R) - (N + B)}{(S1 + R) + (N + B)}$

2.3.3 Feature Importance Evaluation and Feature Selection

Among the 131 features used for vegetation mapping, 126 optical remote sensing features (excluding topography and climate) were constructed based on surface reflectance. These features are prone to severe collinearity issues, which can lead to model overfitting, increased computational costs, and diminished interpretability. To mitigate the issue of high collinearity among these features, the Variance Inflation Factor (VIF) (James et al., 2013) was employed, providing the ratio of variance in a model with multicollinearity among features to the variance in a model where multicollinearity is absent:

$$VIF_j = \frac{1}{1 - R_j^2} \tag{1}$$

where VIF_j is the VIF for feature j, and R_j^2 is the squared multiple correlation coefficient obtained from the regression of feature j with all other features. A higher VIF value indicates more severe collinearity, with a VIF_j greater than 30 suggesting significant collinearity in feature j. This study utilized the RF classifier to determine the optimal feature combination for vegetation classification. The importance of terrain, climate, and optical remote sensing features was assessed through RF classifier and filtered using the VIF method (Ramosaj and Pauly, 2019). The features were then ranked according to the calculated importance, and the top-ranking feature, along with combinations such as the top two, top three, etc., were used to construct various RF models for all possible feature combinations. The feature set with the smallest out-of-bag error was chosen as the best feature set for vegetation mapping on the QTP.

2.3.4 Evaluation of Mapping Accuracy

An optimal combination of terrain, climate, and optical remote sensing data in 2020 was integrated to achieve vegetation mapping of the QTP based on RF model in this study. The mapping accuracy was evaluated using the confusion matrix method, which involved calculating the OA (Eq. 2), Kappa (Eq. 3), PA (Eq. 4), and UA (Eq. 5).

$$OA = \frac{\sum_{i=1}^{n} m_i}{N} \tag{2}$$

$$Kappa = \frac{N \times \sum_{i=1}^{n} m_i - \sum_{i=1}^{n} (G_i \times C_i)}{N^2 - \sum_{i=1}^{n} (G_i \times C_i)} \tag{3}$$

$$PA = \frac{m_i}{G_i} \tag{4}$$

$$UA = \frac{m_i}{C_i} \tag{5}$$

225   where m_i is the count of correctly classified pixels for category i; n is the count of categories; N is the overall quantity of classified pixels; C_i and G_i are the total counts of pixels classified as and actually in category i, respectively; OA is the overall accuracy; Kappa is the Kappa coefficient; PA is the mapping accuracy; and UA is the user accuracy.

2.4 Continuous Annual Vegetation Mapping at 500 m Resolution from 2000 to 2022

Utilizing the 2020 vegetation map at a 500 m spatial resolution as the reference, this study dynamically updated the 2019

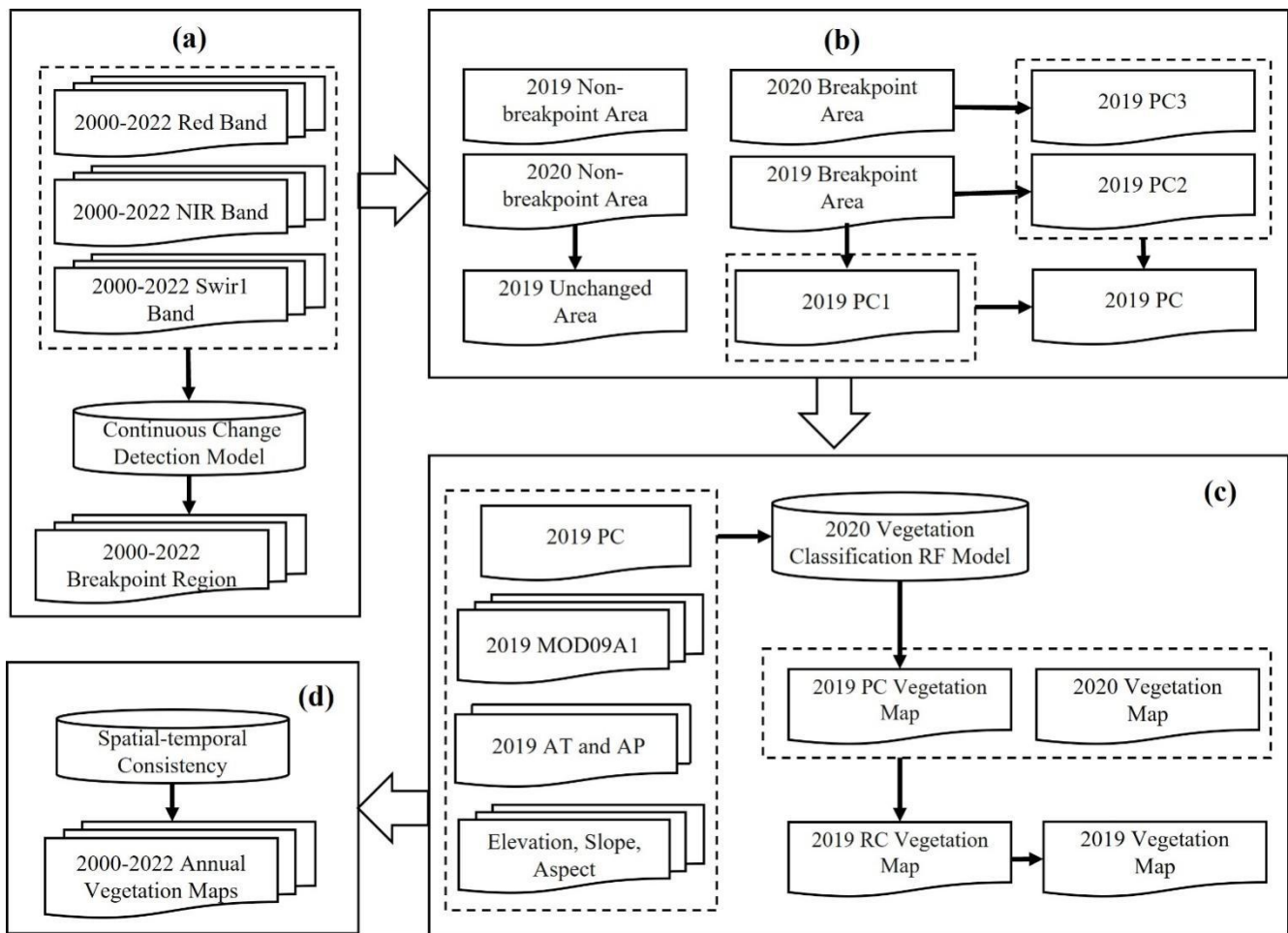
230   vegetation map of the QTP with various algorithms: continuous change detection (CCD) (Eq. 6), potential vegetation change area identification (Eq. 7), and actual vegetation type identification (Eq. 8-9). The process was then repeated, taking the newly updated 2019 map as the reference to dynamically update the 2018 vegetation map. This method was consistently employed to enable the dynamic updating annual vegetation maps from 2000 to 2019, as well as for 2021 and 2022.

235

Figure 2. Vegetation mapping technology route at 500 m resolution on the QTP from 2000 to 2022 (dynamically updating the 2019 vegetation map with the 2020 vegetation map as a reference).

2.4.1 Continuous Change Detection Algorithm

The Continuous Change Detection and Classification (CCDC) is an "online" algorithm that utilizes all available Landsat
240   observations to detect "breaks" in the time series, allowing for the generation of land cover maps for any given time (Zhu and Woodcock, 2014; Zhu et al., 2019). While initially developed for Landsat data, this algorithm has also been used for MODIS and Sentinel datasets (Shimizu et al., 2019; Tang et al., 2019; Xian et al., 2022). The CCDC algorithm comprises two components: "Continuous Change Detection (CCD)" and "Classification (C)".

The CCD part uses a time series fitting model (Eq. 6) to analyze observational data and employs the root mean square
245   error (RMSE) of the fitting model and the residuals of new observations to detect changes. The model includes harmonic terms and a slope term, where the harmonic terms capture intra-annual seasonal variations, and the slope term estimates inter-annual changes. When the residuals of multiple consecutive observations exceed a specified threshold, the

corresponding time period is flagged as a "break." This process is then repeated to identify subsequent "breaks" until all observations have been evaluated. The CCD segments the entire observation sequence into several subsequences, each separated by "breaks." Each subsequence is associated with its own fitting model and specific fitting coefficients.

$$\hat{\rho}(i, t) = c_{0i} + \sum_{n=1}^{3} \left(a_{ni} \cos \frac{2\pi n}{T} t + b_{ni} \sin \frac{2\pi n}{T} t \right) + c_{1i} t \tag{6}$$

where $\hat{\rho}(i, t)$ is the predicted value of the ith band on Julian day t; T is the average number of days in a year; a_{ni} and b_{ni} are the harmonic coefficients of nth order for the ith band i; c_{0i} and c_{1i} are the intercept and slope coefficients, respectively.

The second part of the CCDC, the "C," uses the RF method to classify each subsequence based on its fitting coefficients (Zhu and Woodcock, 2014). Due to the unique geographical location and climatic conditions of the QTP, terrian and climate factors significantly influence vegetation distribution (Zhou et al., 2023). Therefore, instead of focusing solely on the differences in subsequence coefficients from the CCD results, all detected "breaks" were used to identify potential change areas. Subsequently, this study classified vegetation using an RF model based on terrain-climate-remote sensing data.

The CCDC algorithm was initially developed in MATLAB (Zhu and Woodcock, 2014) and later in Python (Brown et al., 2020). While its "online" advantage enables effective detection of changes and model updates as new observations are collected, this advantage comes with high computational resource demands, particularly for long-term monitoring over extensive areas. The GEE platform, with its robust cloud computing capabilities, resolves these challenges by providing the CCDC algorithm as a time segmentation tool labeled "ee.Algorithms.TemporalSegmentation.Ccdc" (Arévalo et al., 2020; Pasquarella et al., 2022). On the GEE platform, the CCDC algorithm includes parameters such as "breakpointBands," "minObservations," "chiSquareProbability," "minNumOfYearsScaler," "dateFormat," "lambda," and "maxIterations." While some studies have fine-tuned these parameters for better results (Awty-Carroll et al., 2019; Brown et al., 2020; Cohen et al., 2020), this study primarily used the default parameters provided by GEE, modifying only "dateFormat" and "breakpointBands." The "breakpointBands" parameter specifies the bands for breakpoint detection, including Red, NIR, and SWIR 1, which correlate with chlorophyll content, leaf structure, and water content, respectively. The "dateFormat" parameter is set to 2, indicating that all breakpoint times are represented as Unix timestamps.

2.4.2 Potential Vegetation Change Area Identification Algorithm

Identifying potential change areas annually is essential for achieving vegetation mapping each year. For example, in dynamically updating the 2019 vegetation map based on the 2020 vegetation map (Fig. 2b), the study area was divided into 4 types (Eq. 7): areas without breakpoints in both 2019 and 2020 (unchanged areas), areas with breakpoints in both 2019 and 2020 (PC 1), areas with breakpoints in 2019 but not in 2020 (PC 2), and areas without breakpoints in 2019 but with breakpoints in 2020 (PC 3). The three types of areas with breakpoints were combined as potential areas for vegetation type changes in 2019, while other areas remained the same as the vegetation types in 2020.

$$S_{T-1} = (B_{T-1} \cap B_T) \cup (B_{T-1} \cap \overline{B_T}) \cup (\overline{B_{T-1}} \cap B_T) \tag{7}$$

where S_{T-1} is the potential vegetation change area for year T-1, B_{T-1} is the breakpoint area for year T-1, B_T is the breakpoint area for year T, $\overline{B_{T-1}}$ is the non-breakpoint area for year T-1, and $\overline{B_T}$ is the non-breakpoint area for year T.

2.4.3 Actual Vegetation Type Identification Algorithm

280 The potential vegetation change areas merely indicate regions where vegetation types might change, thus necessitating further identification of the actual vegetation change areas. Based on the RF model constructed in 2020, and combining the terrain, climate, remote sensing data, and potential vegetation change areas of 2019, the actual vegetation types in the potential change areas for 2019 can be determined. Subsequently, the 2019 vegetation map was obtained by overlaying and analyzing this data with the 2020 vegetation map (Fig. 2c).

$$R'_{T-1} = M_{2020}(F_{T-1}, S_{T-1}) \tag{8}$$

$$R_{T-1} = \begin{cases} R'_{T-1}, & S_{T-1} \text{ and } R'_{T-1} \neq R_T \\ R_T, & \text{otherwise} \end{cases} \tag{9}$$

285 where M_{2020} is the RF model for the year 2020, F_{T-1} is the vegetation mapping features for the year T-1, S_{T-1} is the potential vegetation change area for the year T-1, R'_{T-1} is the vegetation classification result for the potential change area of the year T-1, R_T is the vegetation classification result for the year T, and R_{T-1} is the vegetation classification result for the year T-1.

2.4.4 Spatial-temporal Consistency

290 In this study, RF model is used to generate annual vegetation maps. Although RF handles complex data structures efficiently, its pixel-by-pixel classification method can result in salt-and-pepper noise. To mitigate this issue, this study employs a spatial-temporal constraint method, which assesses the consistency of each pixel's label within a 3×3×3 cube across both spatial and temporal dimensions (Fig. 2d). The consistency, $C_{x,y,t}$, is calculated by averaging the agreement of the central pixel's label with the labels of 27 surrounding pixels (Xu et al., 2021; Li et al., 2015):

$$C_{x,y,t} = \frac{1}{27} \sum_{i=x-1}^{x+1} \sum_{j=y-1}^{y+1} \sum_{k=t-1}^{t+1} I(Label_{x,y,t} = Label_{i,j,k}) \tag{10}$$

295 where I is an indicator function returning 1 if the labels match, otherwise 0.

If $C_{x,y,t} < 0.5$ and $t > 2000$, the pixel is deemed misclassified, and its label is corrected to match the previous year's central label. For the year 2000, where previous year data is unavailable, the label is adjusted to the most frequent label in the spatial 3×3 area. This approach assumes that significant, inconsistent changes are unlikely both spatially and temporally, thereby enhancing the accuracy of the classification.

3 Results

3.1 Vegetation Mapping of the QTP at 500 m Resolution in 2020

Due to the significant collinearity among the 126 features in MODIS data, this study focused on 13 features with VIF below 30, including various band reflectance features and vegetation indices such as NDGlaI 15%, NDGlaI 90%, IBI 30%, NDBI 90%, IBI 15%, IBI 90%, EVI 90%, NDVI 90%, IBI 75%, IBI 60%, LSWI 90%, M 90%, and IBI 45%. Combined with two climate factors (AT and AP) and three terrain features (Elevation, Slope, and Aspect), an 18-feature set was created for OOB error analysis. The results indicated that the OOB error decreased as the number of features increased, reaching a minimum of 0.135 with 11 features, beyond which there was a slight increase (Fig. 3a). Subsequently, the top 11 features were selected to construct the RF model: AP, Elevation, NDVI 90%, AT, Slope, M 90%, EVI 90%, NDBI 90%, NDGlaI 90%, NDGlaI 15%, and IBI 75%. Among them, AP, Elevation, AT, and Slope ranked 1st, 2nd, 4th, and 5th in importance score, respectively. Therefore, climate and terrain are significant factors influencing vegetation distribution on the QTP (Fig. 3b). Finally, the 2020 vegetation map of the QTP was generated on the GEE platform at a 500 m spatial resolution (Fig. 4) based on the developed RF model.

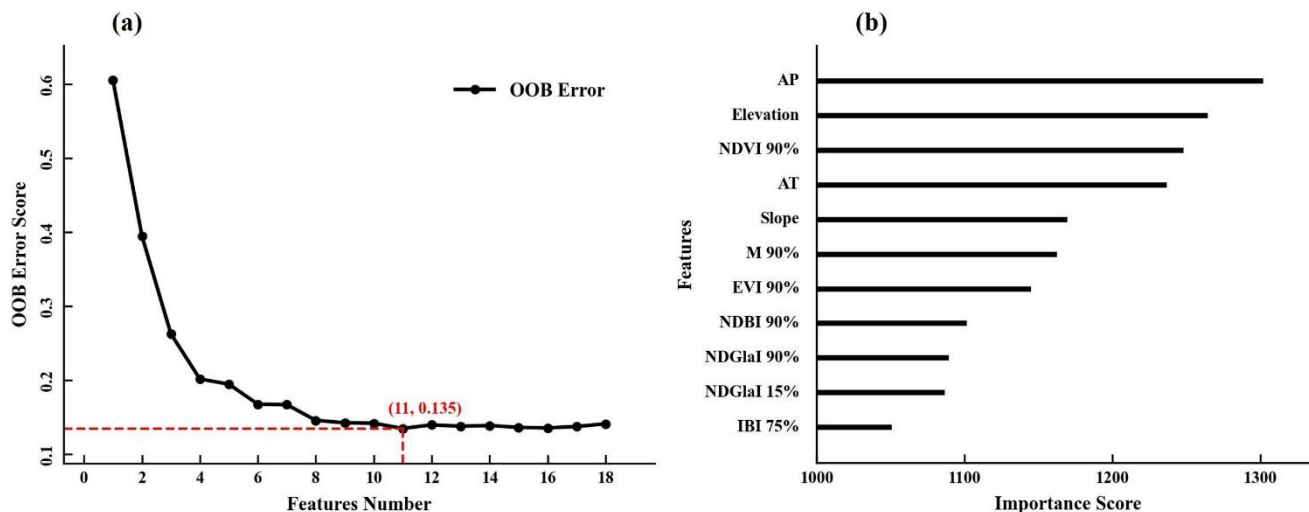



Figure 3. Evaluation of the 2020 vegetation classification model on the QTP.

The vegetation map of the QTP includes 16 types (Fig. 4). Among these, four types cover areas larger than 400,000 km²: AM, AG, AD, and AV. The AM covers the largest area, approximately 798,000 km², mainly in the eastern and southern regions of Gansu and Qinghai provinces. The AG covers around 578,000 km², predominantly in the central part of the plateau, with a distinct transitional zone adjoining the AM. The AD, covering about 562,000 km², is primarily located in the Xinjiang region. The AV covers approximately 474,000 km² and is widely distributed from above the treeline or shrub zone to the lower limit of the perennial snow line on the QTP. Among the five forest types in the QTP, ECF cover the largest area, about 154,000 km², mainly distributed in the Hengduan Mountains in the southeast of the QTP and near the Indus River in

the west. DCF cover the smallest area, approximately 12,000 km², primarily located along the northwestern border of the QTP.

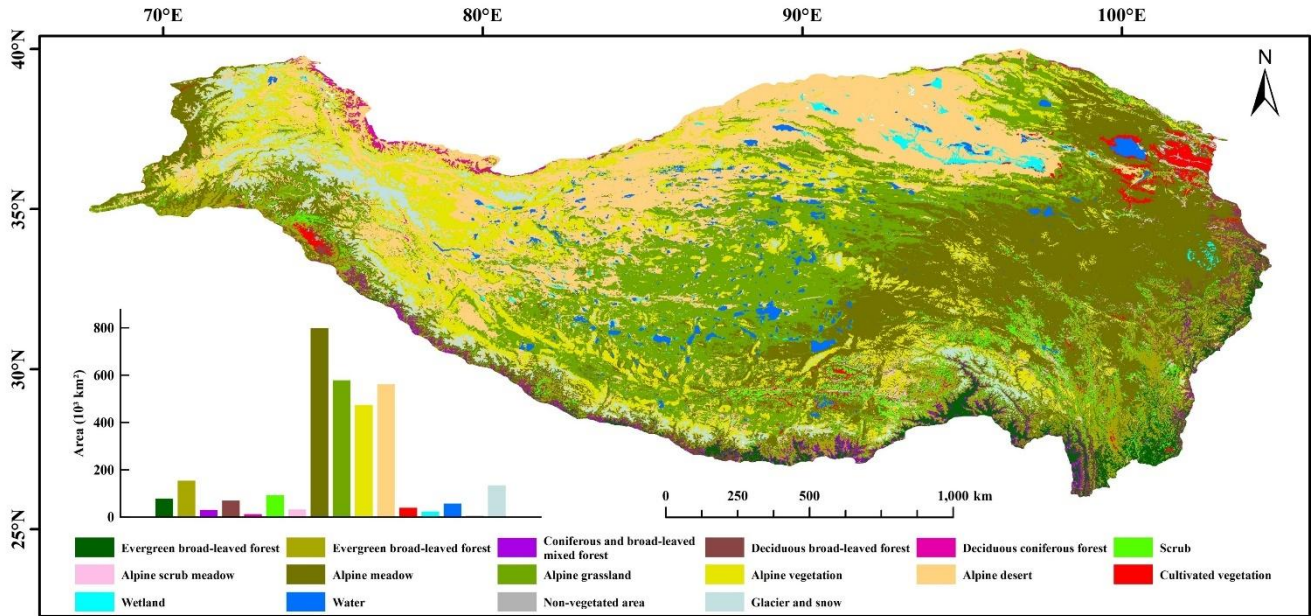



Figure 4. Vegetation types and spatial distribution at 500 m resolution on the QTP in 2020.

3.2 Continuous Annual Vegetation Mapping at 500 m Resolution from 2000 to 2022

This study used the CCD algorithm to detect "breaks" in the dense time series MODIS data of the QTP from 2000 to 2022, defining these breaks as potential changes (PC) (Fig. 5a). Subsequently, an RF model incorporating terrain, climate, and remote sensing data was applied to annually update and generate vegetation maps of the QTP, identifying changes between adjacent years as real changes (RC) (Fig. 5b).

Approximately 31.8% of the QTP is classified as PC areas. Regions with one or two instances of PC account for 18.2% and 7.12% of the area, respectively (Fig. 5g), primarily located at the junctions of AM, AG, and AD in the central part of the plateau (Fig. 5a). Notably, areas near Hala Lake in the northeast (ROI1) and Mapam Yumco in the southwest Ali region (ROI2) experienced more than five instances of change, covering approximately 0.82% of the total area of the QTP (Fig. 5c-d). Compared to the PC areas, the RC areas from the annual vegetation maps of the QTP are significantly smaller (Fig. 5b), encompassing only 3.17% of the total area. As shown in Fig. 5g, regions with one or two instances of RC account for 2.75% and 0.34% of the QTP, respectively. Areas with three or more instances of RC make up only 0.08%, with no regions experiencing five or more instances of change. In all categories of change frequency, the area of RC is consistently smaller than that of PC. In the RC regions of ROI1 and ROI2, only a few areas show one or two instances of change. Notably, in ROI2, no areas with five instances of PC (red) appear in the corresponding RC regions; only a small portion shows a single instance of RC change.

Although the CCD model, which relies on dense spectral data, often detects many false changes—meaning that many "breaks" detected by CCD do not necessarily represent land type transformations (Zhu et al., 2019)—the RF model based on climate, terrain, and remote sensing data produces annual vegetation maps that effectively reduce these false changes. The frequency and area of detected change regions are significantly smaller compared to the CCD model results.

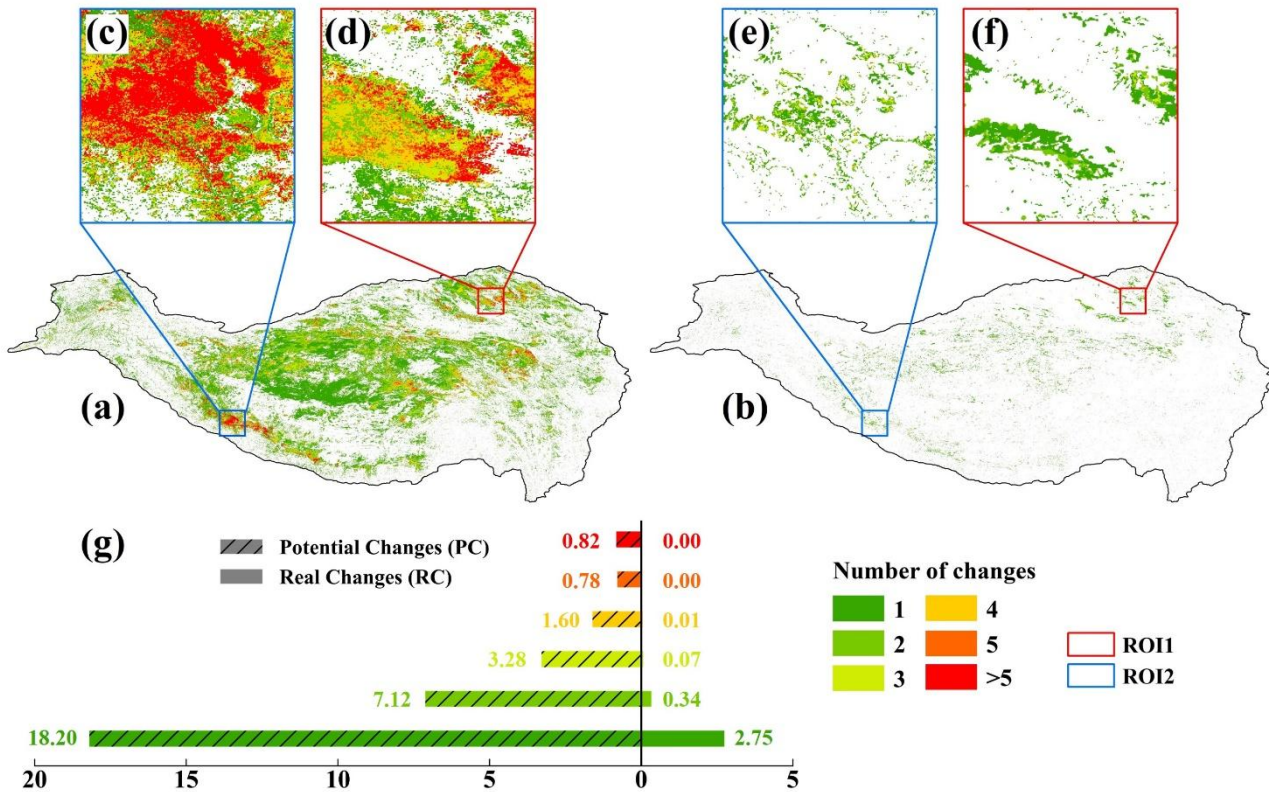



Figure 5. Overview of potential change (PC) and real change areas (RC) on the QTP from 2000 to 2022. (a-b) Distribution of PC and RC on the QTP; (c-f) PC and RC areas in ROI1 and ROI2; (g) Area statistics of PC and RC on the QTP.

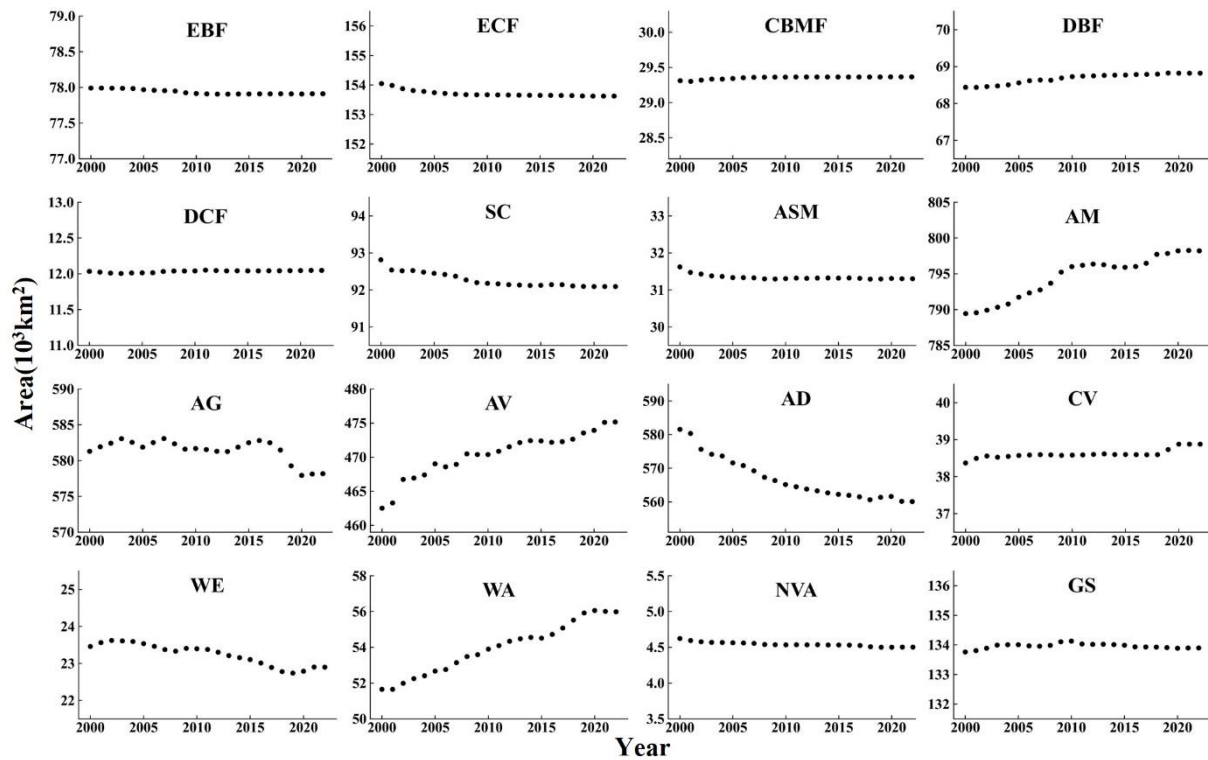



Figure 6. Annual area changes for 16 vegetation types on the QTP from 2000 to 2022.

Fig. 6 illustrates the area changes of 16 vegetation types on the QTP from 2000 to 2022. All forest types remained relatively stable, with changes within 0.3%. Among the five forest types, ECF showed a small decrease, reducing by approximately 400 km² (0.28%), while DCF had a small increase by about 14 km² (0.12%). Except for DBF, SC and ASM have limited distributions on the plateau. ASM decreased by about 300 km² (1.02%), primarily before 2005. AM, AG, AV, and AD are the four most widespread types on the plateau. AM and AV showed significant increases, with AM rising by about 8,800 km² (1.11%) and AV by about 12,600 km² (2.73%). In contrast, AD significantly decreased by about 21,500 km² (3.69%). AG fluctuated but notably declined by about 3,100 km² (0.54%) after 2016. CV increased by approximately 500 km² (1.32%), mainly after 2018. WA showed the most significant change, increasing by about 4,300 km² (8.40%) between 2000 and 2022. GS decreased by approximately 200 km² (0.17%), primarily after 2010. Both WE and NVA showed a decreasing trend, with reductions of about 600 km² and 100 km², respectively, from 2000 to 2022.

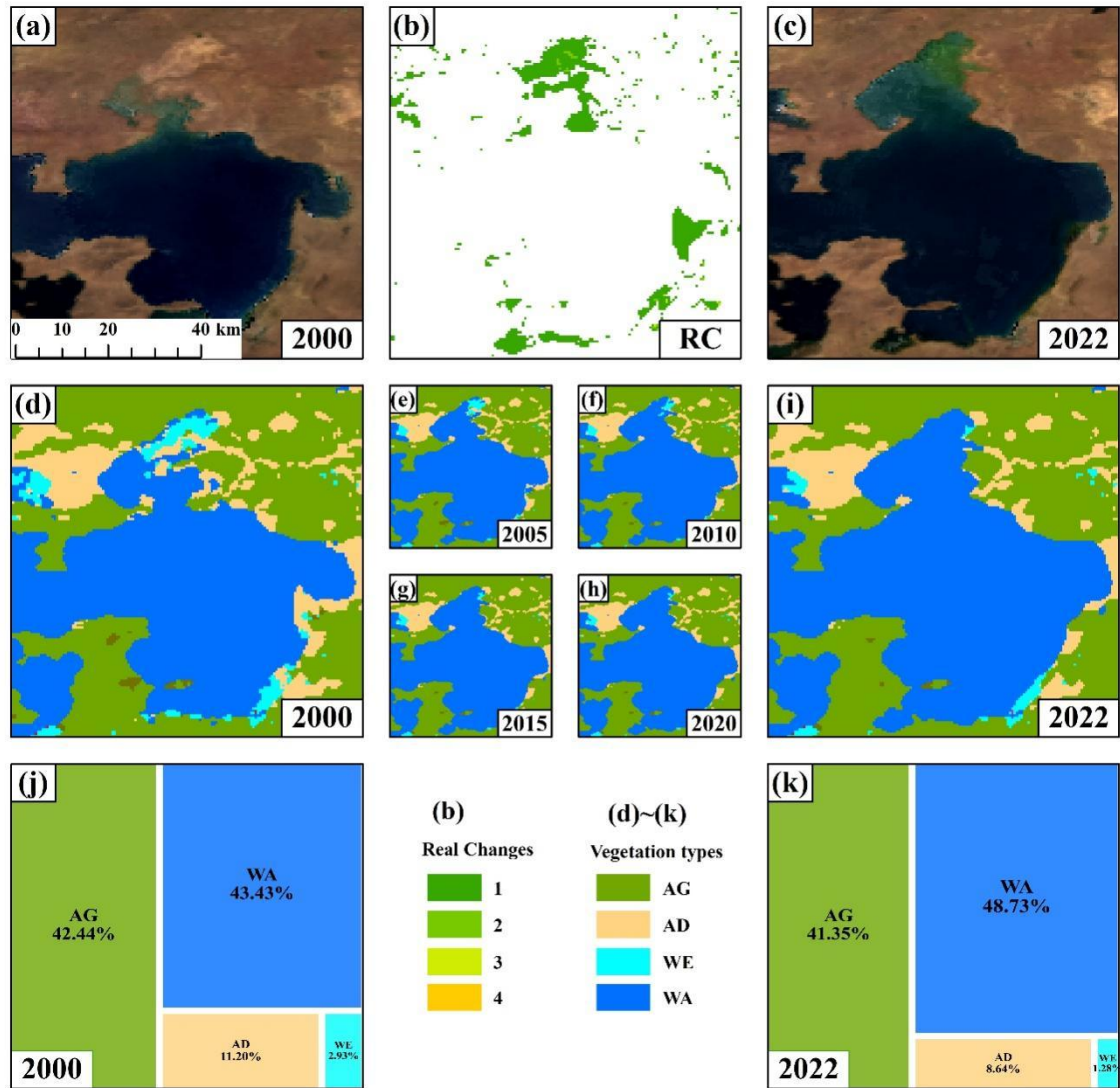



Figure 7. Annual vegetation cover changes in Area 1 [88.6985E, 31.5002N]. (a, c): Remote sensing images from 2000 and 2022, respectively. (b): Distribution and frequency of RC. (d-i): Annual vegetation maps for 2000, 2005, 2010, 2015, 2020, and 2022. (j, k): Area proportions of major vegetation types in 2000 and 2022, respectively.

365   As shown in Fig. 7, most RC occurred only once, predominantly in the northern, eastern, and southwestern parts of Selin Co. In 2000, these areas were classified as WE, AD, and AG, respectively. From 2000 to 2022, these regions gradually transformed into WA, as evident in the remote sensing images. Specifically, the northern WE transitioned to WA between 2000 and 2015, the eastern AD between 2000 and 2010, and the southern AG between 2000 and 2005. The WA in this region increased from approximately 43.43% in 2000 to 48.73% in 2022, marking a 5.3% rise. Simultaneously, AG, AD, and

370 WE decreased by 1.09%, 2.56%, and 1.65%, respectively. The eastern AD was the most significant contributor to the expansion of Selin Co's water.

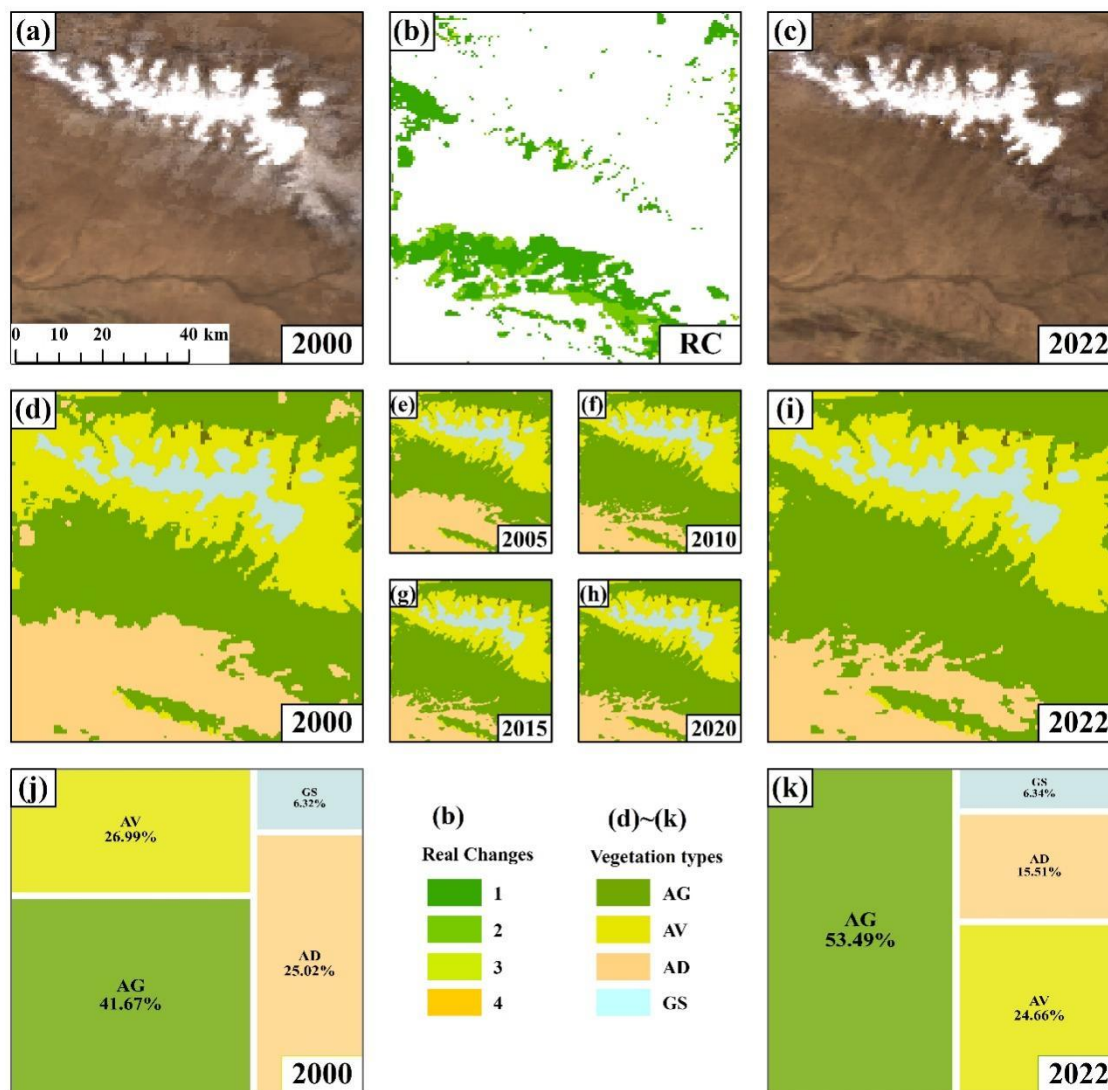



Figure 8. Annual vegetation cover changes in Area 2 [95.8644E, 37.6311N]. (a, c): Remote sensing images from 2000 and 2022,
respectively. (b): Distribution and frequency of RC. (d-i): Annual vegetation maps for 2000, 2005, 2010, 2015, 2020, and 2022. (j, k):
375 **Area proportions of major vegetation types in 2000 and 2022, respectively.**

In the northeastern part of Hala Lake, close to ROI1, the area primarily consists of AG, AV, AF, and GS. Fig. 8b shows that most changes in this region from 2000 to 2022 occurred once or twice. The area of GA changed very little over the past 23 years, increasing slightly from 6.32% to 6.34%. In contrast, AG expanded significantly, from 41.67% to 53.49%, an increase of about 11.82%. Meanwhile, the areas of AD and AV decreased by 9.51% and 2.33%, respectively. These changes

 mainly occurred in transition zones, particularly between AG and AD, where most conversion from AD to AG happened between 2000 and 2010. The shift from AV to AG mostly occurred between 2000 and 2005. Notably, some areas in Fig. 8b experienced two changes. These changes mainly reflect fluctuations between AD and AG (Fig. 8c-d), likely due to similar climate and spectral characteristics in the transition zones, causing some pixels to switch between the two types multiple times.

3.3 Accuracy Validation

The annual vegetation map of the QTP is dynamically updated using 2020 as the reference year, making its accuracy crucial for long-term product reliability. Due to minimal differences in vegetation distribution between adjacent years, samples from 2019, 2020, and 2021 were used to validate the reference map. This validation set included 1,175 samples. The results showed an OA of 83.32% for 2020, with a Kappa coefficient of 0.82 (Table 3). For most vegetation types, the UA for DCF, AM, AV, AD, and WA exceeded 90%, and the PA for DCF, WE, WA, and GS was also above 90%. However, the UA and PA for CBMF were only 57.14% and 48.00%, respectively. This lower accuracy is likely because CBMF, as a mix of coniferous and broadleaf forests, is often misclassified by the RF model as EBF, ECF, or DBF due to spatial similarities. Overall, the 2020 vegetation map, validated with 1,175 samples, achieved sufficient accuracy to be used as the reference for dynamic updates.

Table 3. Confusion matrix, PA, and UA of vegetation mapping at 500 m resolution on the QTP in 2020

	EBF	ECF	CBMF	DBF	DCF	SC	ASM	AM	AG	AV	AD	CV	WE	WA	NVA	GS	Total	UA (%)
EBF	56	11	8	1	0	0	0	0	0	0	0	0	0	0	0	0	76	73.68
ECF	2	74	3	2	1	1	0	3	0	0	0	0	0	0	0	0	86	86.05
CBMF	0	4	12	4	0	0	0	0	0	0	0	1	0	0	0	0	21	57.14
DBF	5	4	2	51	0	0	0	0	0	0	0	0	0	0	0	0	62	82.26
DCF	0	0	0	0	87	0	0	0	2	0	2	0	0	0	0	0	91	95.60
SC	0	0	0	0	0	46	2	6	1	0	1	2	0	0	0	0	58	79.31
ASM	5	0	0	0	0	8	29	4	1	0	0	1	0	0	0	0	48	60.42
AM	0	0	0	0	0	0	1	54	2	0	0	0	0	0	0	0	57	94.74
AG	0	1	0	0	0	4	0	8	61	0	7	2	0	0	0	0	83	73.49
AV	0	0	0	0	0	0	1	2	5	89	0	0	0	0	0	0	97	91.75
AD	0	0	0	0	5	1	0	1	1	0	85	0	0	0	0	0	93	91.40
CV	0	1	0	1	0	1	0	6	0	0	2	87	0	0	7	0	105	82.86
WE	0	0	0	0	2	0	0	6	1	0	4	3	60	0	0	0	76	78.95
WA	0	0	0	0	0	1	0	3	0	0	0	0	0	83	0	0	87	95.40
NVA	0	2	0	0	0	1	0	2	0	0	0	9	0	0	22	0	36	61.11
GS	0	0	0	0	0	0	0	0	0	16	0	0	0	0	0	83	99	83.84
Total	68	97	25	59	95	63	33	95	74	105	101	105	60	83	29	83	1175	
PA (%)	82.35	76.29	48.00	86.44	91.58	73.02	87.88	56.84	82.43	84.76	84.16	82.86	100.00	100.00	75.86	100.00		83.32

The third-party samples (1,014 samples) and visual interpretation samples (1,002 samples) were used to validate the accuracy of annual vegetation maps. The third-party samples mainly include grasslands, AD, and GS. Because the First All-season Sample Set and the Global Land Cover Validation Samples did not differentiate between AM and AG, these two

400   types were combined under the grasslands category during the validation process. The third-party samples mainly originate from 2014 (190 samples) and 2015 (666 samples), with smaller numbers from other years. The visual interpretation samples exhibit stable characteristics across multiple years, making them suitable for validating annual vegetation maps from 2000 to 2019.

Validation results indicated that accuracy from 2000 to 2019 remained above 80%, with the lowest being 82.07% in 2015
405   (Fig. 9). Upon inspection, most of the misclassified samples were located in the Barren land (vegetation cover <10%) category in the First All-season Sample Set, which corresponds to AD in the vegetation maps. However, some samples of this category were classified as AG in the vegetation map. There are definitional differences between Barren land and AD. Additionally, AD and AG represent two largest vegetation types on the QTP, and there are significant transitional zones between these two types (Fig. 4). As a result, the accuracy in 2015, when a large number of third-party samples were used,
410   was slightly lower than in other years. Moreover, because there were insufficient validation samples for 2021 and 2022, and 2020 has already been validated as the reference year, the overall accuracy of the annual vegetation maps is estimated to be 83.27%, based on the average accuracy from 2000 to 2019.

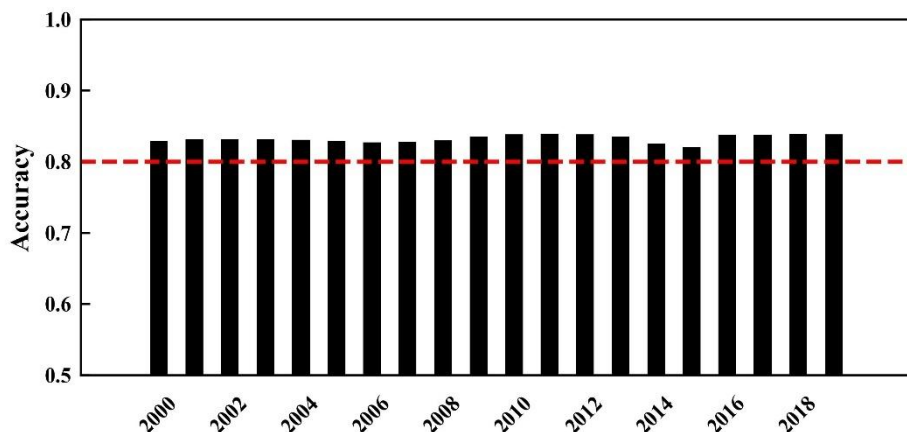



Figure 9. Annual overall accuracy of vegetation maps from 2000 to 2019.

415   **4 Discussion**

4.1 Evaluating the Efficacy of the CCD Algorithm in Annual Vegetation Mapping from 2000 to 2022

This study proposed a method for long-time continuous annual vegetation mapping. Specifically, the CCD algorithm was applied to MODIS data from 2000 to 2022 to detect breakpoints. Subsequent processes involved identifying potential change areas, recognizing true vegetation types, and spatial-temporal consistency. This enabled consistent mapping of vegetation on
420   the QTP annually from 2000 to 2022. The CCD algorithm in this study used harmonic functions to fit long-term remote sensing images, thereby identifying breakpoints and determining the timing of these breakpoints.

For instance, in Fig. 10, the CCD algorithm was applied to detect changes in the RED, NIR, and SWIR bands of the sampling site from 2000 to 2022. The results indicated that there was a breakpoint in 2011 (highlighted in yellow), dividing the period into Fit1 (from 2000 to 2011) and Fit2 (from 2011 to 2022). The annual amplitude of the three bands exhibited small fluctuations with stable interannual patterns. In contrast, the amplitude of the RED and NIR bands in Fit2 far exceeded that of Fit1, showing significant differences in seasonal patterns within the year. Additionally, for the SWIR band, the overall reflectance in Fit2 was substantially lower than in Fit1, reflecting distinct seasonal fluctuations within the year. Based on the annual Landsat images from 2009 to 2013, there was a noticeable expansion of WA in the selected area, marking a transition from AD prior to 2011 to WA thereafter. The annual vegetation maps accurately captured this transition, with the area represented as brown-yellow (AD) before and up to 2010, and shifting to deep blue (WA) from 2011 onward. This change was further corroborated by the CCD fitting results and visual interpretation of the Landsat images.

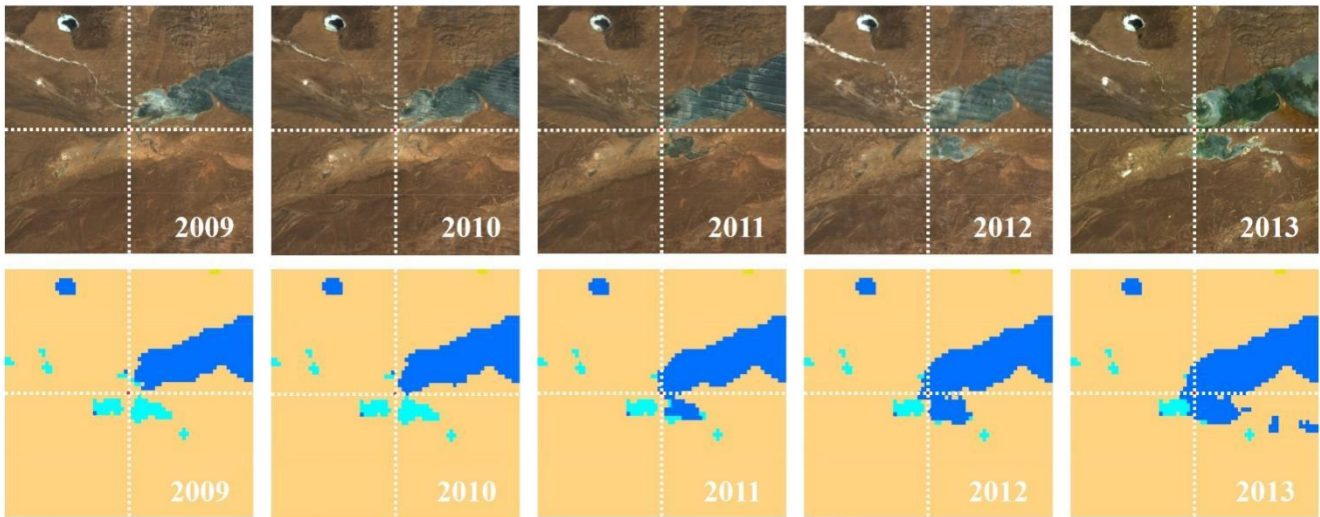
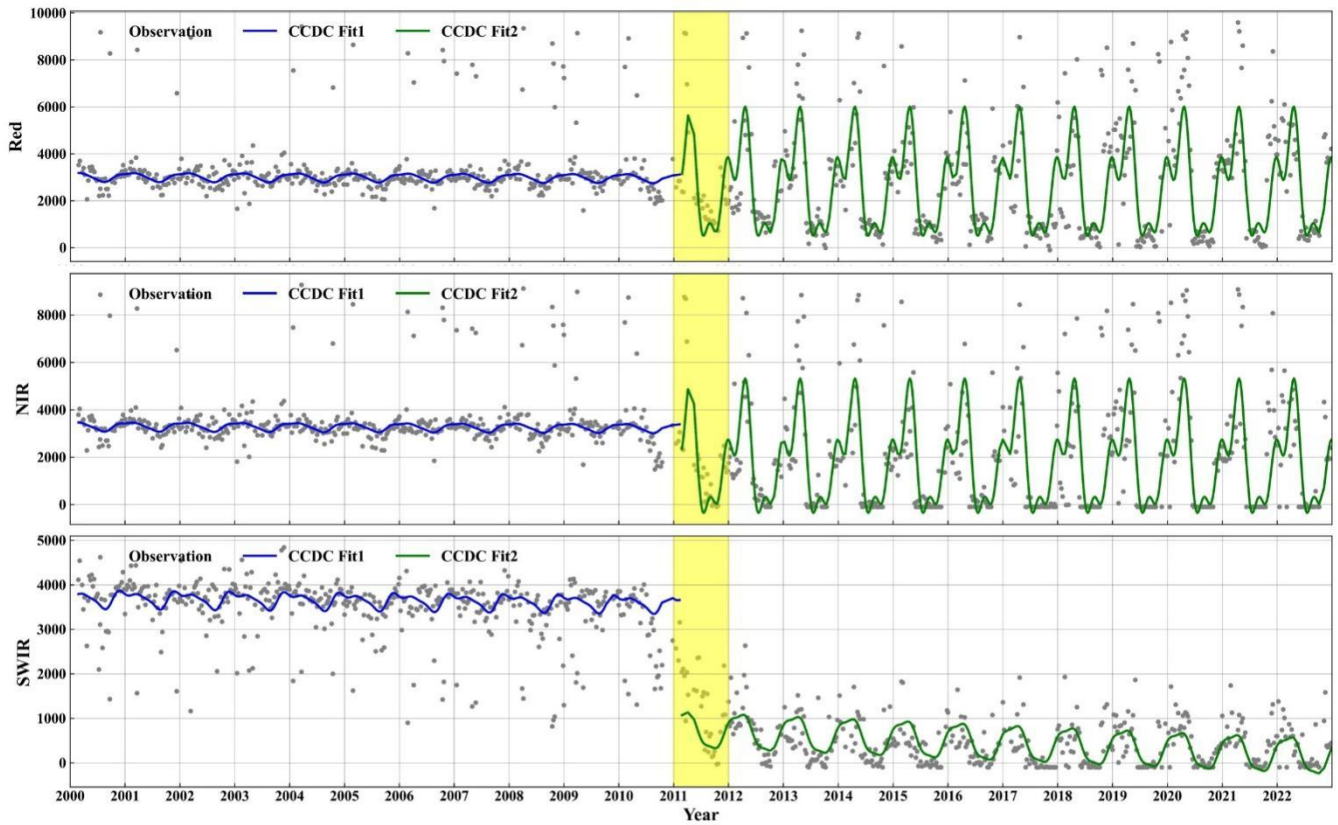



Figure 10. Validation of vegetation mapping: consistent change detection in WA sample [82.9294E, 35.2425N] by CCD, Landsat, and annual vegetation maps.

435   In Fig. 11, the CCD algorithm indicated that there were no breakpoints from 2000 to 2022. The annual and interannual variations in the RED, NIR, and SWIR bands were stable over the years. Combined with the Landsat images, the selected

area was consistently classified as EBF without any detected changes. The long-time annual vegetation maps consistently reflected this characteristic.

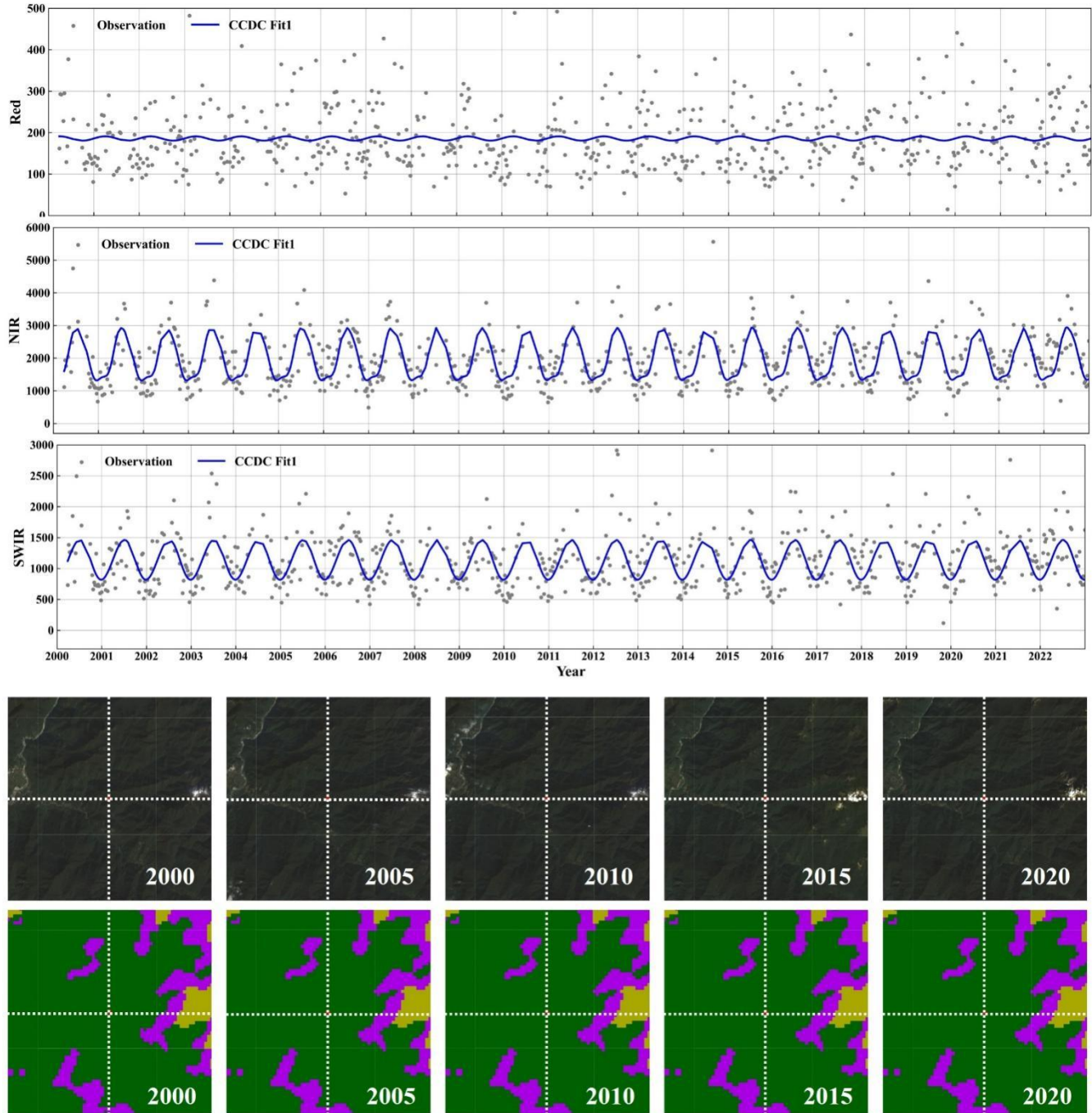



Figure 11. Validation of vegetation mapping: consistency of no change detection in forest sample [95.2794E, 28.7617N] by CCD, Landsat, and annual vegetation maps.

However, abrupt changes detected by the CCD algorithm may not always accurately reflect real changes on the ground (Zhu and Woodcock, 2014; Du et al., 2023). In Fig. 12, an area of CV on the QTP was examined. The CCD results indicated a breakpoint in 2018, with data from 2000 to 2018 categorized as Fit1 and post-2018 as Fit2. Although Fit2 shows greater amplitude in the RED, NIR, and SWIR bands compared to Fit1, the waveform remains similar. Despite a breakpoint was detected, Landsat images from 2016 to 2020 confirm that the area consistently featured CV. The detected changes were likely caused by variations in cultivation practices in 2018 or similar factors, rather than actual changes in vegetation. As a result, the annual vegetation maps did not capture the change detected by the CCD in 2018.

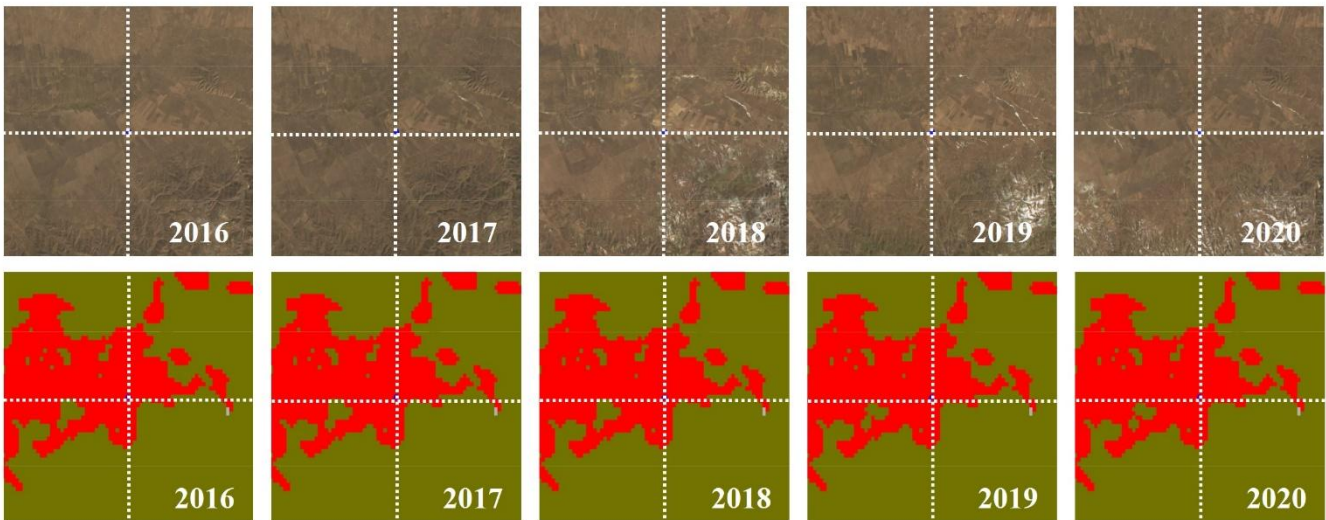
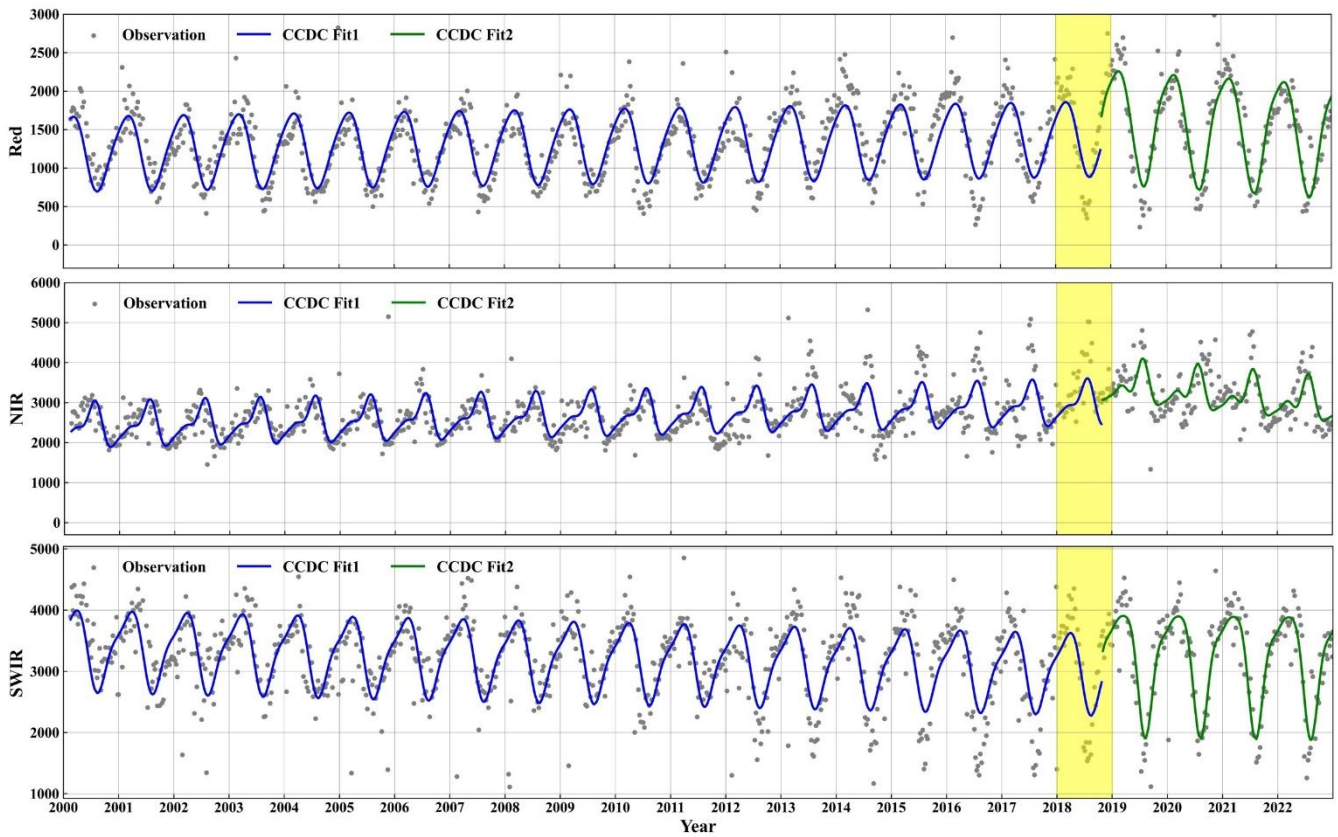
Given its performance, the CCD algorithm effectively identifies the regions and timings of breakpoints in long-time remote sensing imagery, making it a suitable foundational method for detecting potential change areas in annual vegetation mapping. Although the CCD algorithm is susceptible to false positives due to factors such as changes in cultivated species (Fig. 12), the subsequent methods employed in this study, including potential area identification, true vegetation type recognition, and spatial-temporal constraints, help mitigate these false positive errors (Fig. 12).

455

Figure 12. Validation of vegetation mapping: CV sample [100.8570E, 35.2474N] changes detected by CCD not reflected in Landsat and annual vegetation maps.

4.2 Inter-comparison with other products

MCD12Q1 (Sulla-Menashe et al., 2019), CLCD (Yang and Huang, 2021), and GLC-FCS30D (Zhang et al., 2024) were
460   selected for cross-comparison with the annual vegetation maps. MCD12Q1 provides annual global land cover data at a 500
m resolution from 2001 to 2022, using the International Geosphere-Biosphere Programme (IGBP) classification system,
which includes 17 categories such as evergreen needleleaf forest, grassland, and cropland. CLCD offers 30 m land cover
data for China from 1990 to 2022, classifying land into nine types: cropland, forest, shrub, grassland, water, snow/ice, barren,
impervious, and wetland. GLC-FCS30D delivers land cover data from 1985 to 2022, with five-year intervals before 2000
465   and annual intervals thereafter, covering 35 detailed land cover categories. Based on the temporal coverage of these products,
annual vegetation maps from 2001, 2011, and 2021 were selected for cross-validation, with a regional focus on four main
types: WA, GS, grassland, and forest.

Landsat remote sensing images reveal significant shrinkage of water in this region from 2001 to 2021 (Fig. 13a). In 2001,
all four products accurately depicted the distribution of WA, with clear boundaries and complete representation. By 2011,
470   GLC-FCS30D identified part of the WA as WE, while the other three products consistently showed it as WA. By 2021, the
WA had markedly shrunk compared to their extent in 2001. Compared to the Landsat images, CLCD showed an omission of
WA, with most areas transforming into AD, while GLC-FCS30D misclassified part of the WA as WE and identified some
non-water areas as WA.

For GS areas, all four products demonstrated good identification in 2001, 2011, and 2021 (Fig. 13b). Uniquely, the annual
475   vegetation maps identified additional "alpine vegetation". Consequently, in the vicinity of GS, MCD12Q1 categorized these
areas as AD, while CLCD and GLC-FCS30D largely classified them as grassland. The annual vegetation maps not only
preserved the distribution details of GS, AD, and AG from the comparison products but also provided a more detailed
classification of AV.

However, the three comparison products classified this region simply as grassland without further differentiation (Fig.
480   14a). Grassland is a dominant vegetation type on the QTP, covering approximately 57% of the QTP. According to Zhou et al.
(2022a), grasslands on the QTP can be divided into AG, AM, ASM, and AD. Therefore, it is essential to differentiate
between AG, AM, ASM, and AD. In Fig. 14a, the overall distribution of AG and AM has remained consistent over time,
though changes have occurred in their transition zones. For example, in the southwestern AM region of Fig. 14a, some AG
areas have gradually transitioned to AM over the past 23 years.

485   Forests are sparsely distributed on the QTP, found mainly on the southern slopes of the Himalayas and in the southeastern
region (Fig. 14b). Their distribution has remained stable over the years with minimal change. Among the four products,
CLCD did not differentiate forest types, and thus represents them under the general category of 'Forest'. In contrast, the other
three products classified the forested areas as EBF. MCD12Q1 shows a smaller forest extent and includes a notable
proportion of CBMF within this range. The forest extent in the annual vegetation maps, CLCD, and GLC-FCS30D shows

490   high consistency with the Landsat images. However, the annual vegetation maps identified some ASM in the meadow areas adjacent to forests.

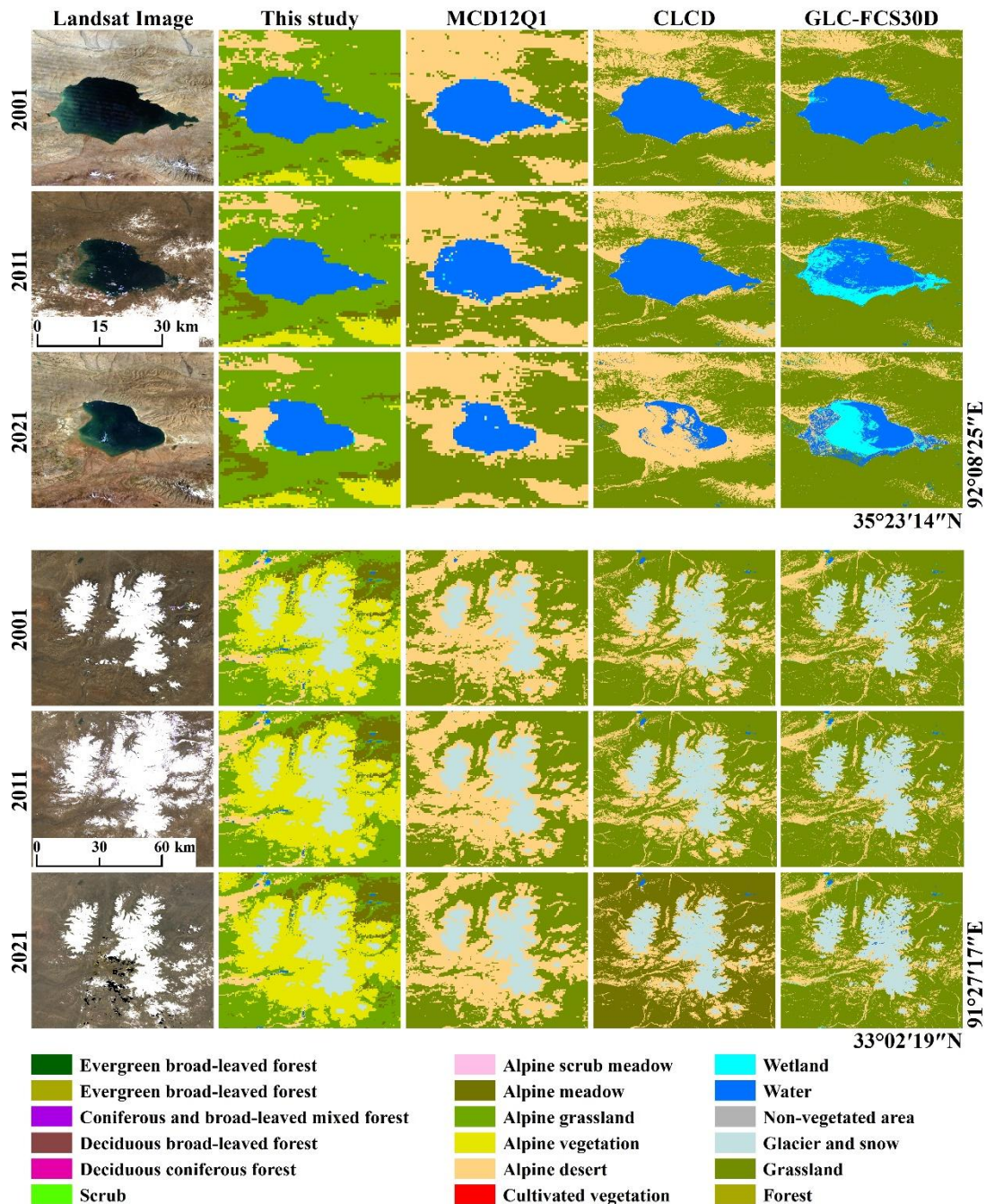



Figure 13. Comparison with the annual vegetation maps, MCD12Q1, CLCD, and GLC-FCS30D. (a): Region dominated by WA and AG. (b): Region dominated by GA and AM.

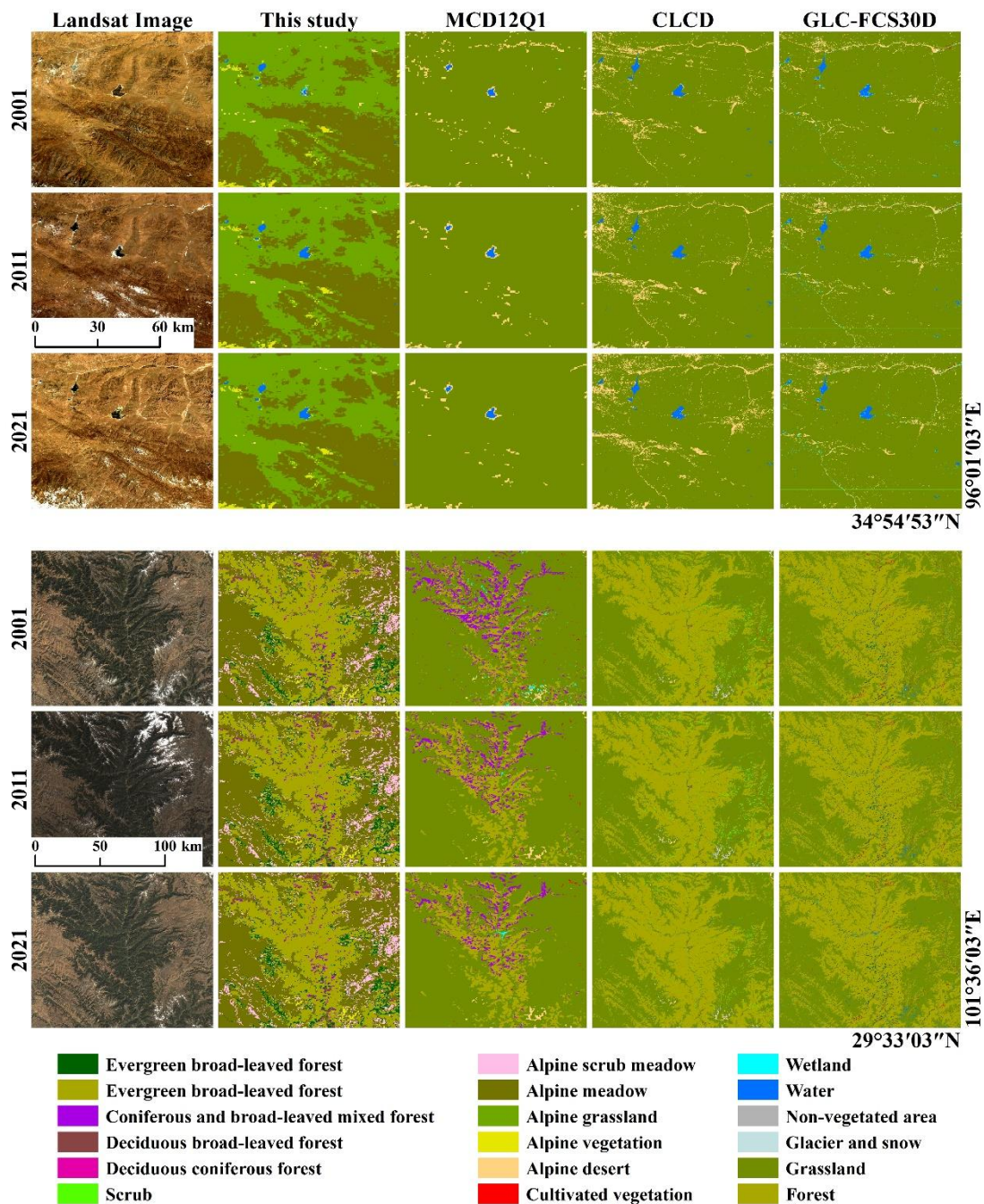



Figure 14. Comparison with the annual vegetation maps, MCD12Q1, CLCD, and GLC-FCS30D. (a): Region dominated by AM and AG. (b): Region dominated by EBF and AM)

5 Data availability

500 The 500 m annual vegetation maps of QTP from 2000 to 2022 are available at https://data.tpdc.ac.cn/en/disallow/6304c1a4-efc0-4766-bae3-4148bdf7bcfd (Zhou et al., 2024). The vegetation maps are stored in TIFF format, with the file name "QTP_Vegetation_Map_XXXX.tif," where XXXX represents the year. All files can be opened and reprocessed using software such as ArcGIS, QGIS, and ENVI. Each TIFF dataset contains values from 0 to 16, where 0 represents invalid values, and 1 to 16 correspond to the 16 vegetation types listed in Table 2. The MOD09A1, MCD12Q1, and SRTM data

505 used in this study were obtained from GEE (http://code.earthengine.google.com, last access: 23 July 2024). The "1-km monthly precipitation dataset for China (1901-2023)", the "1-km monthly mean temperature dataset for China (1901-2023)", the "Vegetation map of Qinghai Tibet Plateau in 2020 with 10 m spatial resolution", and QTP boundary dataset were provided by the National Tibetan Plateau / Third Pole Environment Data Center (https://data.tpdc.ac.cn, last access: 29 September 2024). The First All-season Sample Set can be download from https://data-starcloud.pcl.ac.cn/resource/54 (last

510 access: 29 September 2024). The Global Land Cover Validation Samples can be download from https://zenodo.org/records/3551995 (last access: 29 September 2024). The CLCD can be download from https://zenodo.org/records/8176941 (last access: 23 July 2024), and the GLC-FCS30D can be download from https://zenodo.org/records/8239305 (last access: 23 July 2024).

6 Conclusions

515 Long-time series of annual regional vegetation types and geographic distribution data are vital for examining the impact of climate change on vegetation and its evolutionary trends. In this study, annual vegetation of the QTP from 2000 to 2022 at a 500 m spatial resolution was mapped through the MOD09A1 product, together with a reference year vegetation classification map and a breakpoint detection algorithm. The study achieved an overall accuracy of 83.27% for continuous annual vegetation mapping at a 500 m resolution from 2000 to 2022. The study supports the use of remote sensing data to

520 mapping a long-term continuous annual vegetation. Furthermore, it facilitates the elucidation of the spatial and temporal evolution of regional and global vegetation under the background of global warming.

Author contributions

GZ and HR conceived the study. LZ and HR analysed the data and drafted the paper. XL and MZ supported the generation of the dataset and the analysis of the results.

525 Competing interests

The authors declare that they have no conflict of interest.

Financial support

This work was supported by the Second Tibetan Plateau Scientific Expedition and Research Program (grant number 2019QZKK0106) and the National Natural Science Foundation of China (grant number 42141007).

530 **References**

Arévalo, P., Bullock, E. L., Woodcock, C. E., and Olofsson, P.: A suite of tools for continuous land change monitoring in google earth engine, Front. Clim., 2, 111051, https://doi.org/10.3389/fclim.2020.576740, 2020.

Awty-Carroll, K., Bunting, P., Hardy, A., and Bell, G.: An evaluation and comparison of four dense time series change detection methods using simulated data, Remote Sens., 11, 2779, https://doi.org/10.3390/rs11232779, 2019.

535 Bontemps, S., Defourny, P., Bogaert, E. V., Arino, O., Kalogirou, V., and Perez, J. R.: GLOBCOVER 2009 Products Description and Validation Report, available at: http://due.esrin.esa.int/files/ GLOBCOVER2009_Validation_Report_2.2.pdf (last access: 29 September 2024), 2010.Breiman, L.: Random forests, Machine learning, 45, 5-32, https://doi.org/10.1023/A:1010933404324, 2001.

Brown, J. F., Tollerud, H. J., Barber, C. P., Zhou, Q., Dwyer, J. L., Vogelmann, J. E., Loveland, T. R., Woodcock, C. E.,
540 Stehman, S. V., Zhu, Z., Pengra, B. W., Smith, K., Horton, J. A., Xian, G., Auch, R. F., Sohl, T. L., Sayler, K. L., Gallant, A. L., Zelenak, D., Reker, R. R., and Rover, J.: Lessons learned implementing an operational continuous United States national land change monitoring capability: The Land Change Monitoring, Assessment, and Projection (LCMAP) approach, Remote Sens. Environ., 238, 111356, https://doi.org/10.1016/j.rse.2019.111356, 2020.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., and Lu, M.: Global land cover mapping
545 at 30 m resolution: A POK-based operational approach, ISPRS J. Photogramm. Remote Sens., 103, 7-27, https://doi.org/10.1016/j.isprsjprs.2014.09.002, 2015.

Cohen, W. B., Healey, S. P., Yang, Z., Zhu, Z., and Gorelick, N.: Diversity of algorithm and spectral band inputs improves Landsat monitoring of forest disturbance, Remote Sens., 12, 1673, https://doi.org/10.3390/rs12101673, 2020.

Ding, Y.X., and Peng, S.Z.: Spatiotemporal trends and attribution of drought across China from 1901–2100, Sustainability,
550 12(2), 477, https://doi.org/10.3390/su12020477, 2020.

Du, Z., Yu, L., Li, X., Zhao, J., Chen, X., Xu, Y., Yang, P., Yang, J., Peng, D., and Xue, Y.: Integrating remote sensing temporal trajectory and survey statistics to update land use/land cover maps, Int. J. Digit. Earth, 16, 4428-4445, https://doi.org/10.1080/17538947.2023.2274422, 2023.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., and Roth, L.:
555 The shuttle radar topography mission, Rev. Geophys., 45, https://doi.org/10.1029/2005RG000183, 2007.

Feng, M., Huang, C., Channan, S., Vermote, E. F., Masek, J. G., and Townshend, J. R.: Quality assessment of Landsat surface reflectance products using MODIS data, Comput. Geosci., 38, 9– 22, https://doi.org/10.1016/j.cageo.2011.04.011, 2012.

Feng, M. and Li, X.: Land cover mapping toward finer scales, Sci. Bull., 65, 1604-1606,
560 https://doi.org/10.1016/j.scib.2020.06.014, 2020.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X.: MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets, Remote Sens. Environ., 114, 168-182, https://doi.org/10.1016/j.rse.2009.08.016, 2010.

Gong, P., Liu, H., Zhang, M., Li, C., Wang, J., Huang, H., Clinton, N., Ji, L., Li, W., and Bai, Y.: Stable classification with
565 limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017, Sci. Bull., 64, 370-373, https://doi.org/10.1016/j.scib.2019.03.002 2019.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., and Liu, S.: Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data, Int. J. Remote Sens., 34, 2607-2654, https://doi.org/10.1080/01431161.2012.748992, 2013.

570 Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R.: Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sens. Environ., 202, 18-27, https://doi.org/10.1016/j.rse.2017.06.031, 2017.

Hu, M., Zhou, G., Lv, X., Zhou, L., Wang, X., He, X., and Tian, Z.: Warming Has Accelerated the Melting of Glaciers on the Tibetan Plateau, but the Debris-Covered Glaciers Are Rapidly Expanding, Remote Sens., 15, 132, https://doi.org/10.3390/rs15010132, 2022.

575 Immerzeel, W. W., Van Beek, L. P., and Bierkens, M. F.: Climate change will affect the Asian water towers, Science, 328, 1382-1385, https://doi.org/10.1126/science.1183188, 2010.

James, G., Witten, D., Hastie, T., and Tibshirani, R.: An introduction to statistical learning, Springer, https://doi.org/10.1007/978-1-0716-1418-1, 2013.

Li, C., Peng, G., Wang, J., Zhu, Z., Biging, G. S., Yuan, C., Hu, T., Zhang, H., Wang, Q., and Li, X.: The first all-season
580 sample set for mapping global land cover with Landsat-8 data, Sci. Bull., 62, 508–515, https://doi.org/10.1016/j.scib.2017.03.011, 2017.

Li, X., Gong, P., and Liang, L.: A 30-year (1984–2013) record of annual urban dynamics of Beijing City derived from Landsat data, Remote Sens. Environ., 166, 78-90, https://doi.org/10.1016/j.rse.2015.06.007, 2015.

Li, Y., Hou, Z., Zhang, L., Song, C., Piao, S., Lin, J., Peng, S., Fang, K., Yang, J., and Qu, Y.: Rapid expansion of wetlands
585 on the Central Tibetan Plateau by global warming and El Niño, Sci. Bull., 68, 485-488, https://doi.org/10.1016/j.scib.2023.02.021, 2023.

Liu, H., Gong, P., Wang, J., Wang, X., Ning, G., and Xu, B.: Production of global daily seamless data cubes and quantification of global land cover change from 1985 to 2020-iMap World 1.0, Remote Sens. Environ., 258, 112364, https://doi.org/10.1016/j.rse.2021.112364, 2021.

590 Liu, L., Zhang, X., Gao, Y., Chen, X., Shuai, X., and Mi, J.: FinerResolution Mapping of Global Land Cover: Recent Developments, Consistency Analysis, and Prospects, J. Remote Sens., 2021, 5289697, https://doi.org/10.34133/2021/5289697, 2021.

Pasquarella, V. J., Arévalo, P., Bratley, K. H., Bullock, E. L., Gorelick, N., Yang, Z., and Kennedy, R. E.: Demystifying LandTrendr and CCDC temporal segmentation, Int. J. Appl. Earth Obs., 110, 102806,
595 https://doi.org/10.1016/j.jag.2022.102806, 2022.

Peng, S.: 1-km monthly mean temperature dataset for China (1901-2023), TPDC [dataset], https://doi.org/10.11888/Meteoro.tpdc.270961, 2019.

Peng, S.: 1-km monthly precipitation dataset for China (1901-2023), TPDC [dataset], https://doi.org/10.5281/zenodo.3114194, 2020.

600 Peng, S., Gang, C., Cao, Y., and Chen, Y.: Assessment of climate change trends over the Loess Plateau in China from 1901 to 2100, Int. J. Climatol., https://doi.org/10.1002/joc.5125, 2017a.

Peng, S.Z., Ding, Y.X., Liu, W.Z., and Li, Z.: 1 km monthly temperature and precipitation dataset for China from 1901 to 2017, Earth Syst. Sci. Data, 11, 1931–1946, https://doi.org/10.5194/essd-11-1931-2019, 2019.

Peng, S.Z., Ding, Y.X., Wen, Z.M., Chen, Y.M., Cao, Y., and Ren, J.Y.: Spatiotemporal change and trend analysis of
605 potential evapotranspiration over the Loess Plateau of China during 2011-2100, Agric. For. Meteorol., 233, 183-194, https://doi.org/10.1016/j.agrformet.2016.11.129, 2017b.

Ramosaj, B. and Pauly, M.: Consistent estimation of residual variance with random forest Out-Of-Bag errors, Stat. Probab. Lett., 151, 49-57, https://doi.org/10.1016/j.spl.2019.03.017, 2019.

Shimizu, K., Ota, T., and Mizoue, N.: Detecting Forest Changes Using Dense Landsat 8 and Sentinel-1 Time Series Data in
610 Tropical Seasonal Forests, Remote Sens., 11, 1899, https://doi.org/10.3390/rs11161899, 2019.

Sulla-Menashe, D., Gray, J. M., Abercrombie, S. P., and Friedl, M. A.: Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product, Remote Sens. Environ., 222, 183-194, https://doi.org/10.1016/j.rse.2018.12.013, 2019.

Tang, X., Bullock, E. L., Olofsson, P., Estel, S., and Woodcock, C. E.: Near real-time monitoring of tropical forest
615 disturbance: New algorithms and assessment framework, Remote Sens. Environ., 224, 202-218, https://doi.org/10.1016/j.rse.2019.02.003 , 2019.

Wu, F., Ren, H., and Zhou, G.: The 30 m vegetation maps from 1990 to 2020 in the Tibetan Plateau, Sci. Data, 11, 804, https://doi.org/10.1038/s41597-024-03649-7, 2024.

Xian, G. Z., Smith, K., Wellington, D., Horton, J., Zhou, Q., Li, C., Auch, R., Brown, J. F., Zhu, Z., and Reker, R. R.:
620 Implementation of the CCDC algorithm to produce the LCMAP Collection 1.0 annual land surface change product, Earth Syst. Sci. Data, 14, 143–162, https://doi.org/10.5194/essd-14-143-2022, 2022.

Yang, J. and Huang, X.: The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019, Earth Syst. Sci. Data, 13, 3907–3925, https://doi.org/10.5194/essd-13-3907-2021, 2021.

Yu, B., Lv, C., Lv, T., Yang, A., and Liu, C.: Datasets of the boundary and area of the Tibetan Plateau, Acta Geogr. Sin., 69(s1), 65-68, https://doi.org/10.11821/dlxb2014S012, 2014.

Zhang, G., Luo, W., Chen, W., and Zheng, G.: A robust but variable lake expansion on the Tibetan Plateau, Sci. Bull., 64, 1306-1309, https://doi.org/10.1016/j.scib.2019.07.018, 2019a.

Zhang, H., Zhao, C., and Zhu, W.: A new vegetation map for Qinghai-Tibet Plateau by integrated classification from multi-source data products(2020), TPDC [dataset], https://doi.org/10.11888/Terre.tpdc.271997, 2022.

Zhang, X., Liu, L., Chen, X., Gao, Y., Xie, S., and Mi, J.: GLC_FCS30: global land-cover product with fine classification system at 30m using time-series Landsat imagery, Earth Syst. Sci. Data, 13, 2753–2776, https://doi.org/10.5194/essd-13-27532021, 2021a.

Zhang, X., Liu, L., Chen, X., Xie, S., and Gao, Y.: Fine land-cover mapping in China using Landsat datacube and an operational SPECLib-based approach, Remote Sens., 11, 1056, https://doi.org/10.3390/rs11091056, 2019b.

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 Syst. Sci. Data, 16, 1353–1381, https://doi.org/10.5194/essd-16-1353-2024, 2024.

Zhang, Y., Li, B., Liu, L., and Zheng, D.: Redetermine the region and boundaries of Tibetan Plateau, Geogr. Res, 40, 1543-1553, https://doi.org/10.11821/dlyj020210138 ,2021b.

Zhang, Y., Liu, L., Li, B., and Zheng, D.: Boundary Data of the Tibetan Plateau (2021 Version), Digital Journal of Global Change Data Repository [J/DB/OL], https://doi.org/10.3974/geodb.2021.07.10.V1, 2021c.

Zhang, Y., Liu, L., Li, B., and Zheng, D.: Comparison of boundary datasets covering Tibetan Plateau between 2021 and 2014 versions, J. Global Change Data Discov., 5(3), 322-332, https://doi.org/10.3974/geodp.2021.03.10, 2021d.

Zhang, Y.: Integration dataset of Tibet Plateau boundary, TPDC [dataset], https://doi.org/10.11888/Geogra.tpdc.270099, 2019.

Zhou, G., Ren, H., Liu, T., Zhou, L., Ji, Y., Song, X., and Lv, X.: Vegetation map of Qinghai Tibet Plateau in 2020 with 10 m spatial resolution, TPDC [dataset], https://doi.org/10.11888/Terre.tpdc.272408, 2022a.

Zhou, G., Ren, H., Liu, T., Zhou, L., Ji, Y., Song, X., and Lv, X.: A new regional vegetation mapping method based on terrain-climate-remote sensing and its application on the Qinghai-Xizang Plateau, Sci. China Earth Sci., 66, 237-246, https://doi.org/10.1007/s11430-022-1006-1, 2023.

Zhou, G., Ren, H., Zhang, L., Lv, X., and Zhou, M.: 500 m annual vegetation maps of Qinghai Tibet Plateau (2000-2022), TPDC [dataset], https://data.tpdc.ac.cn/en/disallow/6304c1a4-efc0-4766-bae3-4148bdf7bcfd, 2024.

Zhou, J., Zheng, Y., Song, C., Cheng, C., Gao, P., Shen, S., and Ye, S.: Vegetation map of Qinghai Tibet Plateau in 1980s, TPDC [dataset], https://doi.org/10.11888/Terre.tpdc.272385, 2022b.

Zhu, Z. and Woodcock, C. E.: Continuous change detection and classification of land cover using all available Landsat data, Remote Sens. Environ., 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 Sens. Environ., 238, 111116, https://doi.org/10.1016/j.rse.2019.03.009, 2019.