https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

Open Access Earth System Science Data Discussions

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 time continuity of vegetation to presence, and leads to a gradual increase in the number of misclassified pixels and the uncertainty of 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 80.9% based on 733 samples from literature, with the reference annual 2020 reaching an accuracy of 86.5% and a Kappa coefficient of 0.85. The study supports the use of remote sensing data to mapping a long-term continuous annual vegetation.

1 Introduction

Vegetation, an integral component of Earth's ecosystems, plays an irreplaceable role in maintaining climate stability, preserving biodiversity, and supplying vital human resources. 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

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

GlobCover_2005 and GlobCover_2009 (300 m) (Bontemps et al., 2010), and NASA's annual MCD12Q1 product since 2000 (500 m) (Friedl et al., 2010; Sulla-Menashe et al., 2019). In China, representative products include GlobeLand30 by Chen et al. (2015), FROM_GLC10 and FROM_GLC30 by Tsinghua University (Gong et al., 2013; Gong et al., 2019), GLC_FCS30 by the Chinese Academy of Sciences (Zhang et al., 2019), and CLCD by Wuhan University (Yang and Huang, 2021). Building on 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).

At present, the preparation of long-time series LUSV dataset typically adopts independent classification methods for each period's products, such as the CLCD (Yang and Huang, 2021) and the GlobeLand30 (Chen et al., 2015). These methods often lead to misclassified pixels and the uncertainty of their locations in each period's land cover and vegetation types, because they do not consider the time continuity of vegetation presence. Over time, the trend of increasing misclassified pixels and their positional uncertainties ultimately reduces the reliability of remote sensing interpretation for LCSV types. Therefore, improving the precision of remote sensing mapping for long-time series of LUSV types is a pressing issue that needs to be addressed.

The QTP, known as the "Roof of the World" and the "Water Tower of Asia," plays a crucial role in 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, and vegetation survey is 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 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 with 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., 2023), 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., 2022). 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 LUSV data of the QTP, such as wetland, water, and glacier (Li et al., 2023; Zhang et al., 2019; Hu et al., 2022). However, these long-time series products, created with independent classification methods, are primarily focused on specific LUSV types, overlooking the need to identify temporal changes in 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 from remote sense imagery, and to map vegetation of the entire QTP at 500 m from 2000 to 2022, using the MOD09A1 remote sensing data.

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

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 4320m and a total area of 3.08 million km² (Zhang et al., 2021). This study covers the entire QTP, involving 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 (Zhang et al., 2021). 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 that ranges from forests, shrubs, and meadows in the southeast to grasslands and deserts in the northwest. Additionally, as altitude increases and temperatures decrease, 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 Vegetation Type Sample Data

The vegetation of the QTP is primarily categorized into 15 types (Zhou et al., 2023). Due to 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 vegetation type sample data 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., 2022). 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). Following these rules, random sampling was carried out in areas where purity exceeded 70%, representing approximately 62.34% of the QTP. The vegetation type sample data for the QTP outside of China's borders were derived from high-resolution Google online imagery from the year 2020. By visually interpreting these images, areas within a 500 m range where the vegetation type was consistent were selected as samples for vegetation types. A total of 9,920 pixels from both within and outside the QTP were obtained, of which 90% (8,937 points) were used as training samples, while

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

the remaining 10% (983 points) were reserved for validation (Fig. 1). Furthermore, all pixel data (Table 2), encompassing all vegetation types, were evenly distributed across the study area.

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

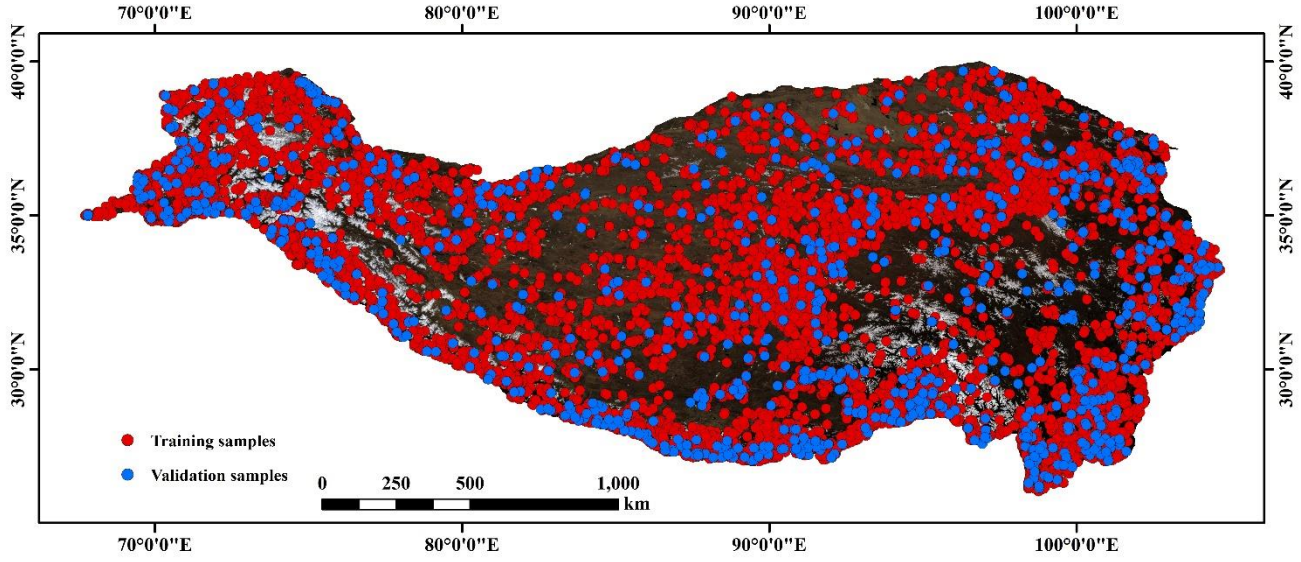

Figure 1. Spatial distribution of training and validation samples for vegetation mapping at 500 m resolution on the QTP in 2020.

Table 2. Training and Validation Sample Counts for Vegetation Mapping on the QTP.

Vegetation types	Training samples	Validation samples	Total
EBF	936	106	1042
ECF	768	96	864
CBMF	475	61	536
DBF	570	71	641
DCF	370	41	411
SC	397	36	433
ASM	213	21	234
AM	1375	148	1523
AG	872	86	958

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

AV	503	59	562
AD	499	51	550
CV	469	65	534
WE	246	24	270
WA	377	31	408
NVA	204	24	228
GS	663	63	726
Total	8937	983	9920

2.2.2 Remote Sensing Data

Remote sensing data in this study were sourced from the 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 Google Earth Engine (GEE) platform, having undergone atmospheric and topographic corrections. The MOD09A1 dataset, which provides surface reflectance in seven spectral bands including Red, Blue, Green, NIR, MIR, SWIR 1, and SWIR 2 at 500 m spatial resolution, with all cloud-contaminated pixels removed. There were 1051 MOD09A1 remote sensing images from January 1, 2000, to December 31, 2022, selected for this study, which were converted to the WGS84 geographic coordinate grid from original sinusoidal projection for the preparation of annual vegetation maps at 500 m of the QTP from 2000 to 2022.

2.2.3 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 was obtained from the National Tibetan Plateau Data Center. In contrast, climate data for areas of the QTP outside China was derived from the CRU high-resolution gridded dataset, featuring a spatial resolution of approximately 50,000 m. The terrain data employed comes 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.

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

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 3): 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 (Table 3). These features were derived from the MOD09A1 remote sensing imagery data spanning from January 1, 2020, to December 31, 2020. Additionally, 6 percentiles: 5%, 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. A total of 131 features were formulated from these 4 categories for vegetation mapping on the QTP.

Table 3. 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	$\frac{N-R}{N+R}$
	EVI	$2.5\frac{N-R}{N+6R-7.5B+1}$

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

	RVI	$\frac{N}{R}$
	DVI	$N - R$
	SAVI	$\frac{(N - R) * 1.5}{N + R + 0.5}$
	GCVI	$\frac{N}{G} - 1$
	NIRV	$\frac{(N - R) * N}{N + R}$
Urban Index	NDBI	$\frac{S1 - N}{S1 + N}$
	IBI	$\frac{NDBI - (SAVI + (G - S1)/(G + S1))/2}{NDBI + (SAVI + (G - S1)/(G + S1))/2}$
Water Index	NDWI	$\frac{G - N}{G + N}$
	LSWI	$\frac{N - S1}{N + S1}$
Snow Index	NDSI	$\frac{G - S1}{G + S1}$
	NDGlaI	$\frac{G - R}{G + R}$
Soil Index	BI	$\frac{(S1 + R) - (N + B)}{(S1 + R) + (N + B)}$

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 may suffer from 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:

$$VIF_j = \frac{1}{1\text{-}R_j^2} \quad (1)$$

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 best 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

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

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), MA (Eq. 4), and UA (Eq. 5).

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

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

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

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

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 vegetation map of the QTP with various algorithms: time series fitting (Eq. 6), breakpoint detection (Eq. 7), potential vegetation change area identification (Eq. 8), and actual vegetation type identification (Eq. 9-10). 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 applied, enabling the dynamic updating of annual vegetation maps for the years 2000 to 2019, as well as for 2021 and 2022.

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

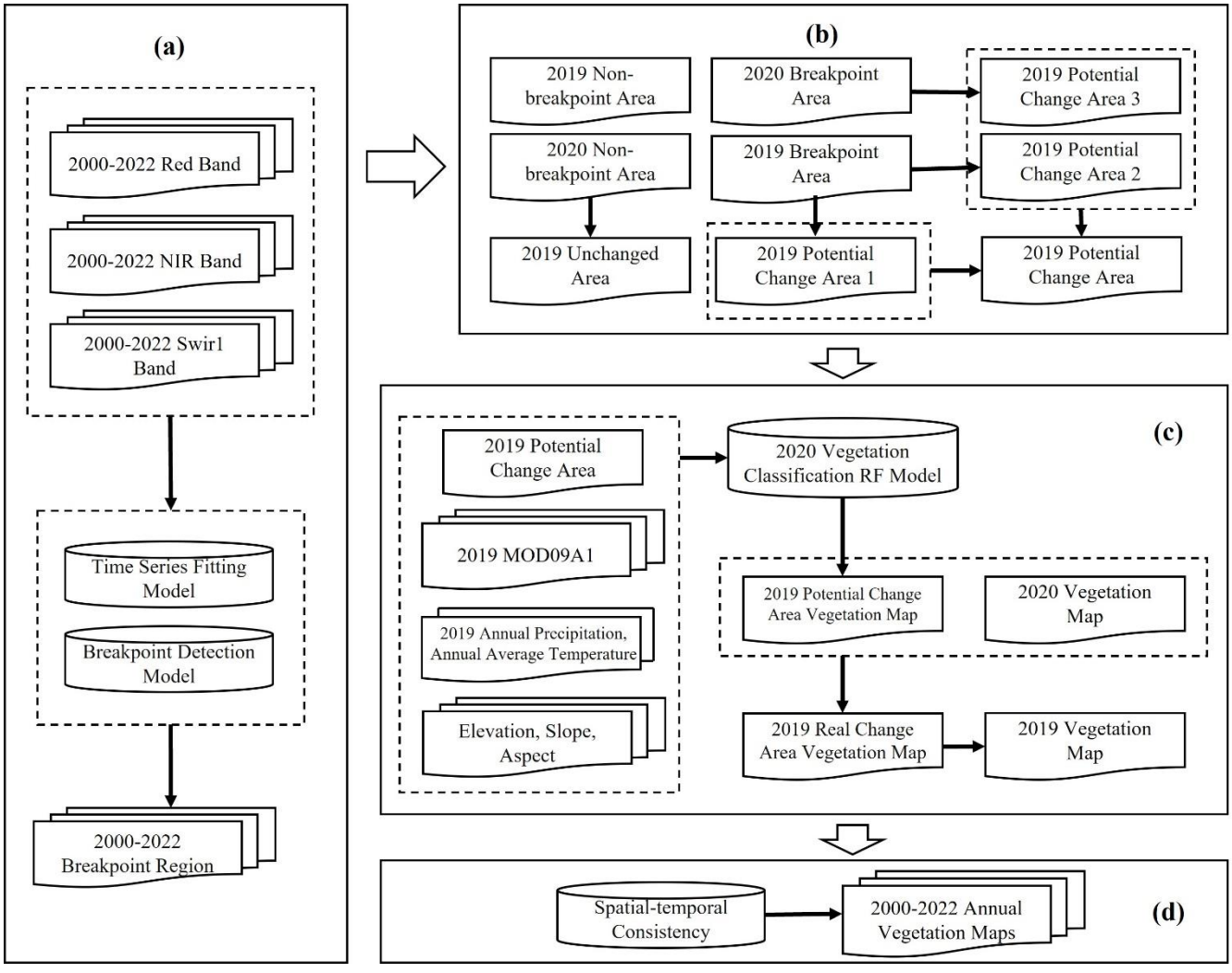

Figure 2. Vegetation mapping technology route at 500 m resolution on the QTP from 2000 to 2022 (dynamically updating the 2019 vegetation map with the 2020 vegetation map as a reference).

2.4.1 Time Series Fitting Algorithm

The Continuous Change Detection and Classification (CCDC) model (Eq. 6) was employed to analyze three MODIS surface reflectance bands – Red, NIR, and SWIR 1 – for the period from 2000 to 2022 (Zhu and Woodcock, 2014). Each band is associated with specific vegetation characteristics: the red band correlates with chlorophyll content, the NIR band with leaf structure, and the SWIR 1 band with water content. The CCDC model effectively characterizes seasonal patterns, trends, and abrupt changes in vegetation, estimating model coefficients through the least squares method from actual observations.

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

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

where $\hat{\rho}(i,t)$ is the predicted value of the i[th] 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 n[th] order for the i[th] band i; c_{0i} and c_{1i} are the intercept and slope coefficients, respectively.

2.4.2 Breakpoint Detection Algorithm

The CCDC model (Zhu and Woodcock, 2014) predicts values for each pixel across all time points, from which it then calculates the average value, standard deviation, and residuals for each pixel. A pixel's vegetation type is considered unchanged if the residual value at a time point is less than three times the standard deviation of the current model. Changes in vegetation type are determined when residual values at three consecutive time points each exceed three times the model's RMSE (Eq. 7), a criterion reflecting the gradual nature of such changes. The Red, NIR, and SWIR 1 bands are treated as individual time series, where the residual values and averages are calculated for three consecutive time points per band. The period is identified as a breakpoint in the time series (Fig. 2a) when the average residual value across the three bands at consecutive time points exceeds a specific threshold, indicating potential changes in vegetation type. Employing this approach, this study calculated annual breakpoint areas for the QTP from 2000 to 2022, ensuring that the vegetation type sample data for 2020 did not include any breakpoints.

$$\frac{1}{3}\Sigma_{i=1}^{3}\frac{|\rho(i,t)-\hat{\rho}(i,t)|}{3\times \mathrm{RMSE_i}} > 1 \quad (7)$$

where i is the i[th] spectral band; t is the Julian day; $\mathrm{RMSE_i}$ is the root mean square error for the i[th] band; $\rho(i,x)$ and $\hat{\rho}(i,x)$ are the observed and predicted values of the i[th] band on Julian day t, respectively.

2.4.3 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: areas without breakpoints in both 2019 and 2020 (unchanged areas), areas with breakpoints in both 2019 and 2020 (potential change area 1), areas with breakpoints in 2019 but not in 2020 (potential change area 2), and areas without breakpoints in 2019 but with breakpoints in 2020 (potential change area 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.

$$S_{T-1} = (B_{T-1} \cap B_T) \cup (B_{T-1} \cap \overline{B_T}) \cup (\overline{B_{T-1}} \cap B_T) \quad (8)$$

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.

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

2.4.4 Actual Vegetation Type Identification Algorithm

The potential vegetation change areas merely indicate that the vegetation types in these areas 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).

$$R'_{T-1} = M_{2020}(F_{T-1}, S_{T-1}) \tag{9}$$

$$R_{T-1} = \begin{cases} R'_{T-1}, & S_{T-1} \text{ and } R'_{T-1} \neq R_T \\ R_T, & \text{otherwise} \end{cases} \tag{10}$$

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.

2.4.5 Spatial-temporal Consistency

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 produce salt-and-pepper noise. To mitigate this issue, we employ 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):

$$C_{x,y,t} = \frac{1}{27}\sum_{i=x-1}^{x+1}\sum_{j=y-1}^{y+1}\sum_{k=t-1}^{t+1} I(Label_{x,y,t} = Label_{i,j,k}) \tag{11}$$

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.

2.5 Evaluation of Annual Vegetation Mapping Accuracy at 500 m Resolution

To evaluate the accuracy of the annual vegetation maps of the QTP at 500 m from 2000 to 2022, this study performed a literature review using the CNKI (https://www.cnki.net/) and Web of Science (https://www.webofscience.com/) databases with keywords including 'Qinghai-Tibet Plateau,' 'vegetation,' and 'vegetation cover.' This search identified 116 papers, providing a total of 733 vegetation type sample points across various years. Among these, the samples were predominantly composed of AM (504 points), followed by AG (72 points), and AD (39 points). The dataset also included ECF (27 points),

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

CBMF (7 points), DCF (7 points), SC (45 points), ASM (18 points), CV (8 points), and WE (6 points). Sample points were distributed throughout the years, excluding 2000, 2021, and 2022, with notable concentrations in 2014 (122 points), 2015 (95 points), and 64 points each in 2006 and 2009.

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%, sur_refl_b05 90%, and IBI 45%. These were combined with two climate factors (AT and AP) and three terrain features (elevation, slope, and aspect) to create an 18-feature set for out-of-bag error analysis. This analysis indicated that the out-of-bag accuracy improved with an increasing number of features, reaching a peak accuracy of 0.86 with 11 features, beyond which there was a slight decrease. Accordingly, the study selected these top 11 features for the RF model: AP, elevation, NDVI 90%, annual average temperature, slope, sur_refl_b05 90%, EVI 90%, NDBI 90%, NDGlaI 90%, NDGlaI 15%, and IBI 75%, which were then used on the GEE platform to produce the 2020 vegetation map of the QTP at a 500 m spatial resolution (Fig. 3).

Table 4. Confusion Matrix, MA, 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	MA
EBF	88	4	6	8	0	0	0	0	0	0	0	0	0	0	0	0	106	83.0%
ECF	3	83	6	2	0	1	0	1	0	0	0	0	0	0	0	0	96	86.5%
CBMF	12	12	36	1	0	0	0	0	0	0	0	0	0	0	0	0	61	59.0%
DBF	3	0	0	66	0	0	0	0	0	0	0	1	1	0	0	0	71	93.0%
DCF	0	0	0	0	41	0	0	0	0	0	0	0	0	0	0	0	41	100.0%
SC	1	1	0	4	0	28	0	2	0	0	0	0	0	0	0	0	36	77.8%
ASM	1	1	0	0	0	0	15	4	0	0	0	0	0	0	0	0	21	71.4%
AM	0	0	0	0	0	0	1	142	3	1	1	0	0	0	0	0	148	95.9%
AG	0	1	0	0	0	0	0	2	80	1	1	1	0	0	0	0	86	93.0%
AV	0	0	0	0	1	0	0	5	1	47	5	0	0	0	0	0	59	79.7%
AD	0	0	0	0	0	0	0	0	2	8	41	0	0	0	0	0	51	80.4%
CV	0	0	0	5	0	1	0	2	0	0	0	51	0	0	6	0	65	78.5%
WE	0	0	0	0	0	0	0	2	1	0	1	0	20	0	0	0	24	83.3%
WA	0	0	0	0	0	0	0	0	1	0	0	0	0	29	1	0	31	93.5%
NVA	0	0	0	0	0	0	0	0	1	0	0	2	0	0	21	0	24	87.5%
GS	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	62	63	98.4%
Total	108	102	48	86	42	30	16	161	89	57	49	55	21	29	28	62	983	
UA	81.5%	81.4%	75.0%	76.7%	97.6%	93.3%	93.8%	88.2%	89.9%	82.5%	83.7%	92.7%	95.2%	100.0%	75.0%	100.0%		

With an OA of 86.5% and a Kappa coefficient of 0.85, the 2020 vegetation map of the QTP at a 500 m spatial resolution was suitable as the reference vegetation map. The details are presented in Table 4, which includes the confusion matrix, MA, and UA. The MA for CBMF, SC, ASM, AV, and CV was all below 80%. Specifically, the MA for CBMF and ASM were

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

59.0% and 71.4%, respectively. According to the confusion matrix, the CBMF was often misclassified as EBF and ECF, while ASM was misclassified as AM. The UA shows that CBMF, DBF, and NVA all have UAs below 80%. Among these, CBMF and NVA have a UA of 75.0%, and DBF has a UA of 76.7%. According to the confusion matrix, there is noticeable confusion between EBF and ECF, as well as between CBMF. A similar issue of misclassification is observed between ASM and AM. The root cause of this confusion lies in the significant overlap in characteristics such as canopy shape, tree height, growth environment, and vegetation composition, especially since CBMF are inherently a combination of both tree types. In parallel, ASM and AM share many attributes, including their vegetation structure, physical features, and spectral reflectance, leading to these classification challenges.

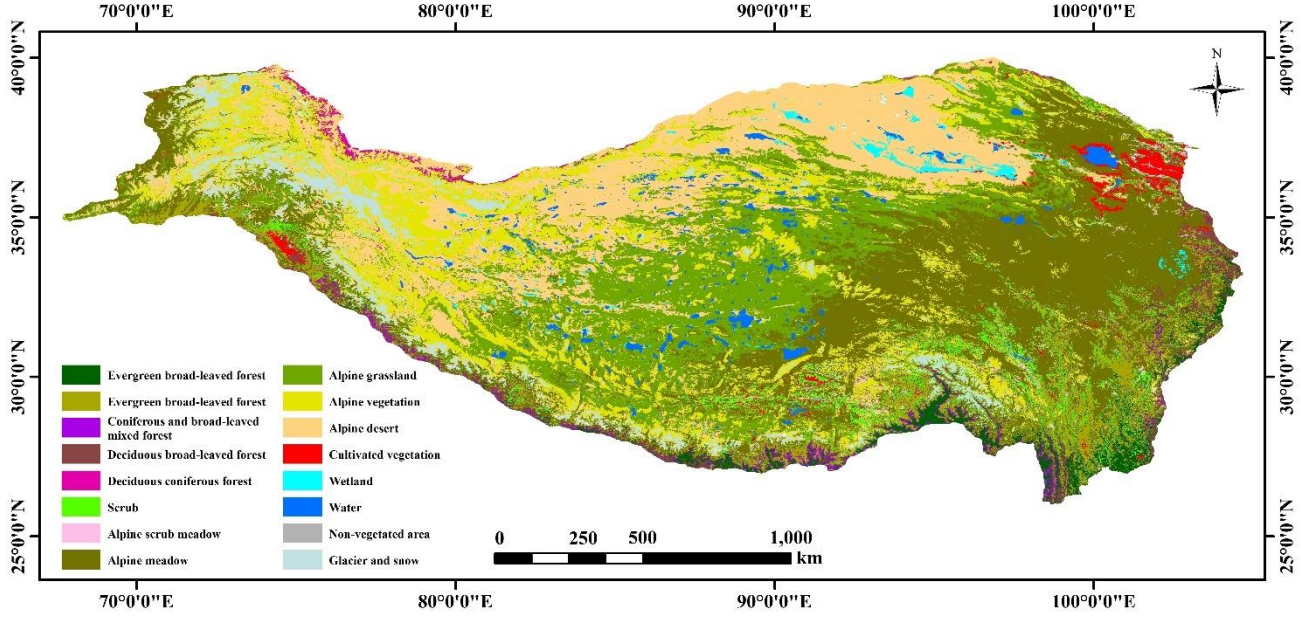

Figure 3. Vegetation types and spatial distribution at 500 m resolution on the QTP in 2020.

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

By integrating the CCDC model, breakpoint detection, and identifying potential change areas, this study identified the annual potential vegetation changes on the QTP from 2000 to 2022. These changes were identified at 500 m using three reflective bands from the MOD09A1 data: Red, NIR, and SWIR 1. Building on this data, annual vegetation mapping of the QTP was conducted at the same resolution. The 2020 vegetation map served as a reference, and the RF classification model was employed to support this process (Fig. 4).

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

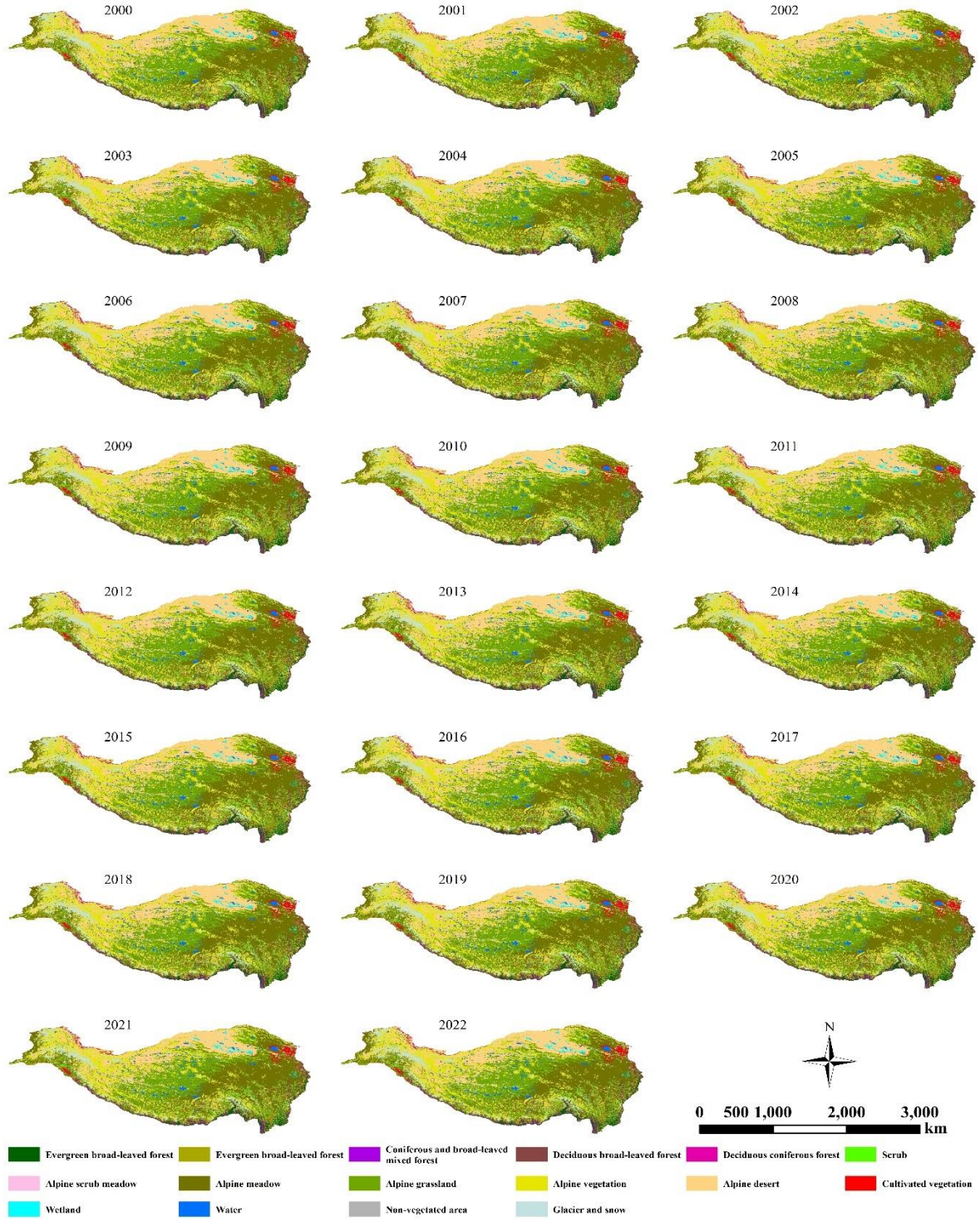

Figure 4. Vegetation types and spatial distribution at 500 m resolution on the QTP from 2000 to 2022.

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

3.3 Accuracy Validation

The validation of vegetation types using 733 samples from literature indicated an overall accuracy of 80.9%, with 593 samples correctly classified. On the QTP, of the 504 AM samples, 462 were accurately classified (91.7% accuracy), and 56 of 72 AG samples were correctly identified, resulting in a 77.8% accuracy rate. Excluding the years 2000, 2021, and 2022, which had insufficient validation samples, the accuracy for other years remained above 70%. The validation accuracy based on samples collected from literature was 5.6% lower compared to the 2020 vegetation map of the QTP. This discrepancy may due to the fact that the validation samples for the 2020 map primarily came from areas with pixel purity above 70%, while the 733 sample pixels from the literature had lower purity and were affected by mixed pixel issues. AM and AG, with the wide distribution across the QTP, tend to have higher pixel purity at a 500 m spatial resolution, contributing to higher validation accuracy. Conversely, smaller, patchily distributed types like SC and ASM have lower pixel purity, potentially blending various vegetation types and thus resulting in lower validation accuracy.

4 Discussion

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

This study proposed a method for long-time continuous annual vegetation mapping. Specifically, the CCDC 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 the QTP annually from 2000 to 2022. The CCDC algorithm in this study used harmonic functions to fit long-term remote sensing images, thereby identifying breakpoints and precisely determining the timing of these breakpoints.

For instance, in Fig. 5, the CCDC 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), where the period from 2000 to 2011 was categorized as Fit1, and from 2011 to 2022 as Fit2. The annual amplitude of the three bands in Fit1 was relatively small, with stable interannual fluctuations. 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, with noticeable differences in 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. The region transitioned from AD before 2011 to WA afterwards. Fortunately, our annual vegetation map also reflected this characteristic. The area was represented as brown-yellow (AD) before and up to 2010, and as deep blue (WA) from 2011 onward. This change aligned with both the CCDC fitting results and the visual interpretation of Landsat images.

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

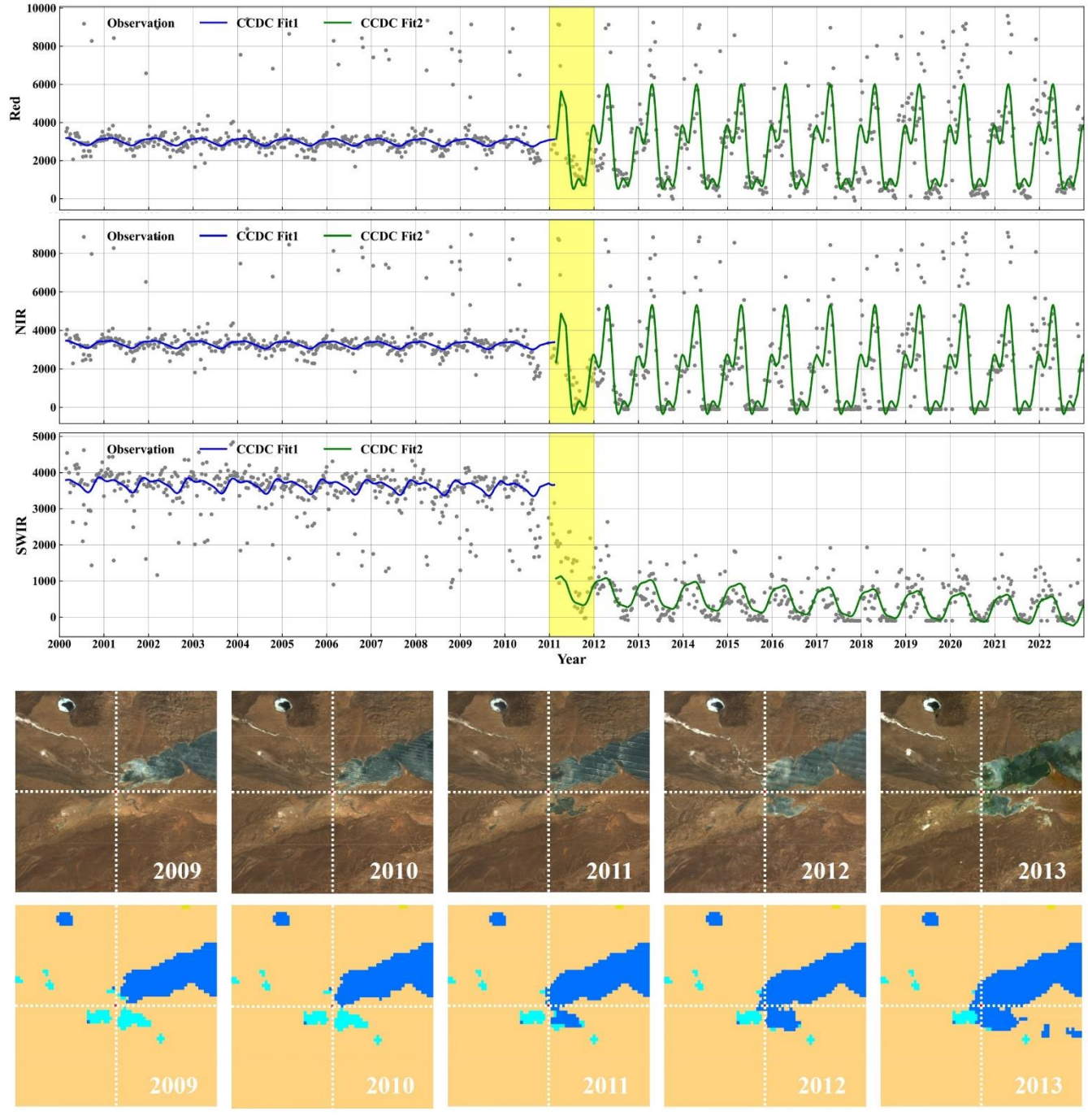

Figure 5. Validation of vegetation mapping: consistent change detection in WA sample [82.9294E, 35.2425N] by CCDC, Landsat, and annual vegetation maps.

290 In Fig. 6, the CCDC 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, it was evident

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

Open Access Earth System Science Data Discussions

that the selected area had consistently been EBF without any change. Our long-time annual vegetation maps also reflected this characteristic.

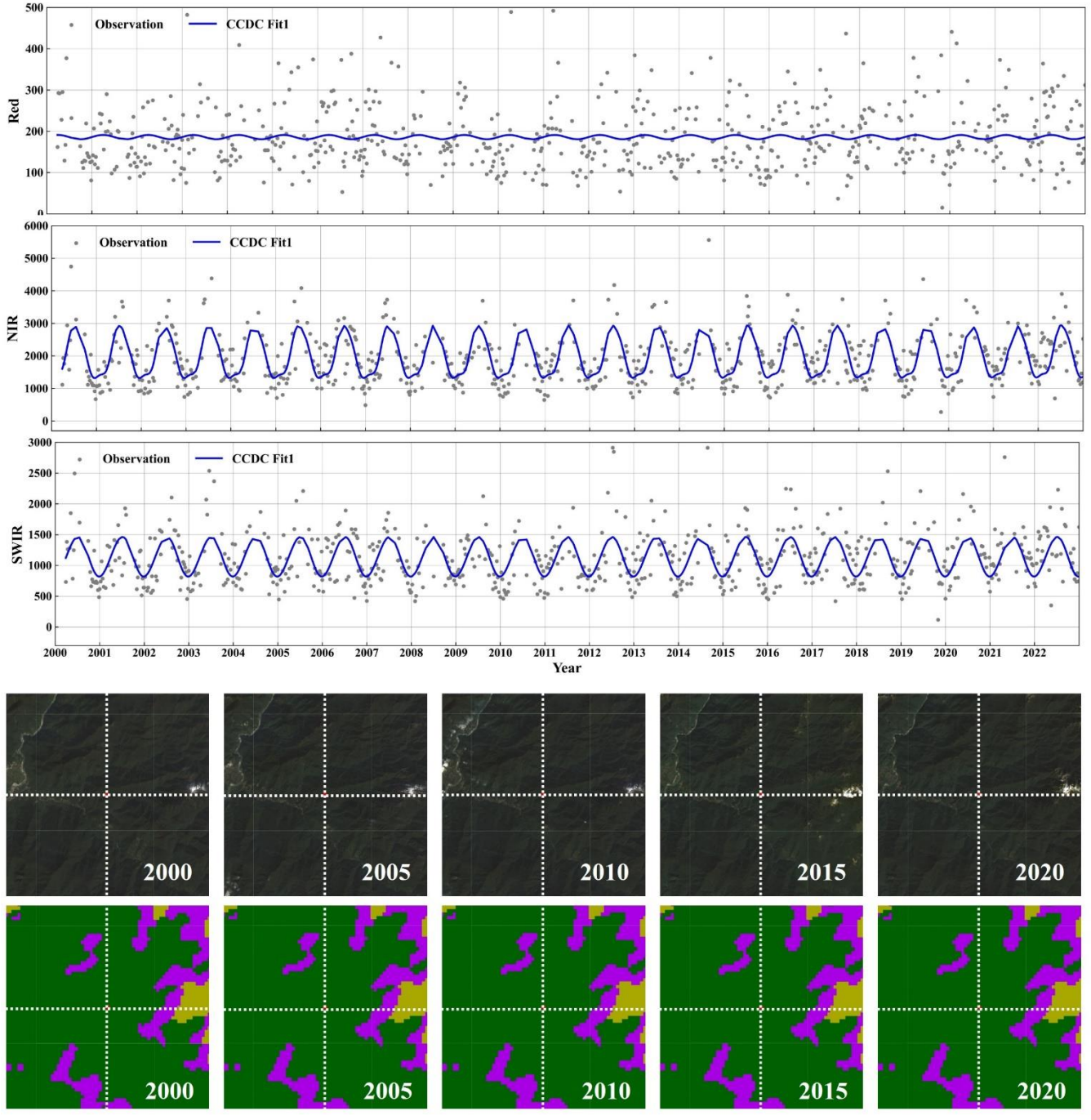

Figure 6. Validation of vegetation mapping: consistency of no change detection in forest sample [95.2794E, 28.7617N] across CCDC, Landsat, and annual vegetation maps.

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

However, abrupt changes detected by the CCDC algorithm do not always accurately reflect real changes on the ground (Zhu and Woodcock, 2014; Du et al., 2023). In Fig. 7, we checked an area of CV on the QTP. The CCDC results indicate a breakpoint in 2018, with the 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 this detected breakpoint, Landsat images from 2016 to 2020 confirms that the area consistently featured CV. The detected changes were likely caused by variations in cultivation practices in 2018 or similar factors, rather than true changes in vegetation. Notably, our annual vegetation maps did not reflect this erroneous detection by the CCDC in 2018.

Therefore, the CCDC algorithm effectively identifies the regions and timings of breakpoints in long-time remote sensing imagery. Choosing the CCDC algorithm as the foundational method for identifying potential change areas in our long-time vegetation mapping approach is both reasonable and appropriate. Although the CCDC algorithm is susceptible to false positives due to factors such as changes in cultivated species (Fig. 7), the subsequent methods employed in this study, including potential area identification, true vegetation type recognition, and spatial-temporal constraints, effectively reduce these false positive errors (Fig. 7).

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

Open Access Earth System Science Data Discussions

310

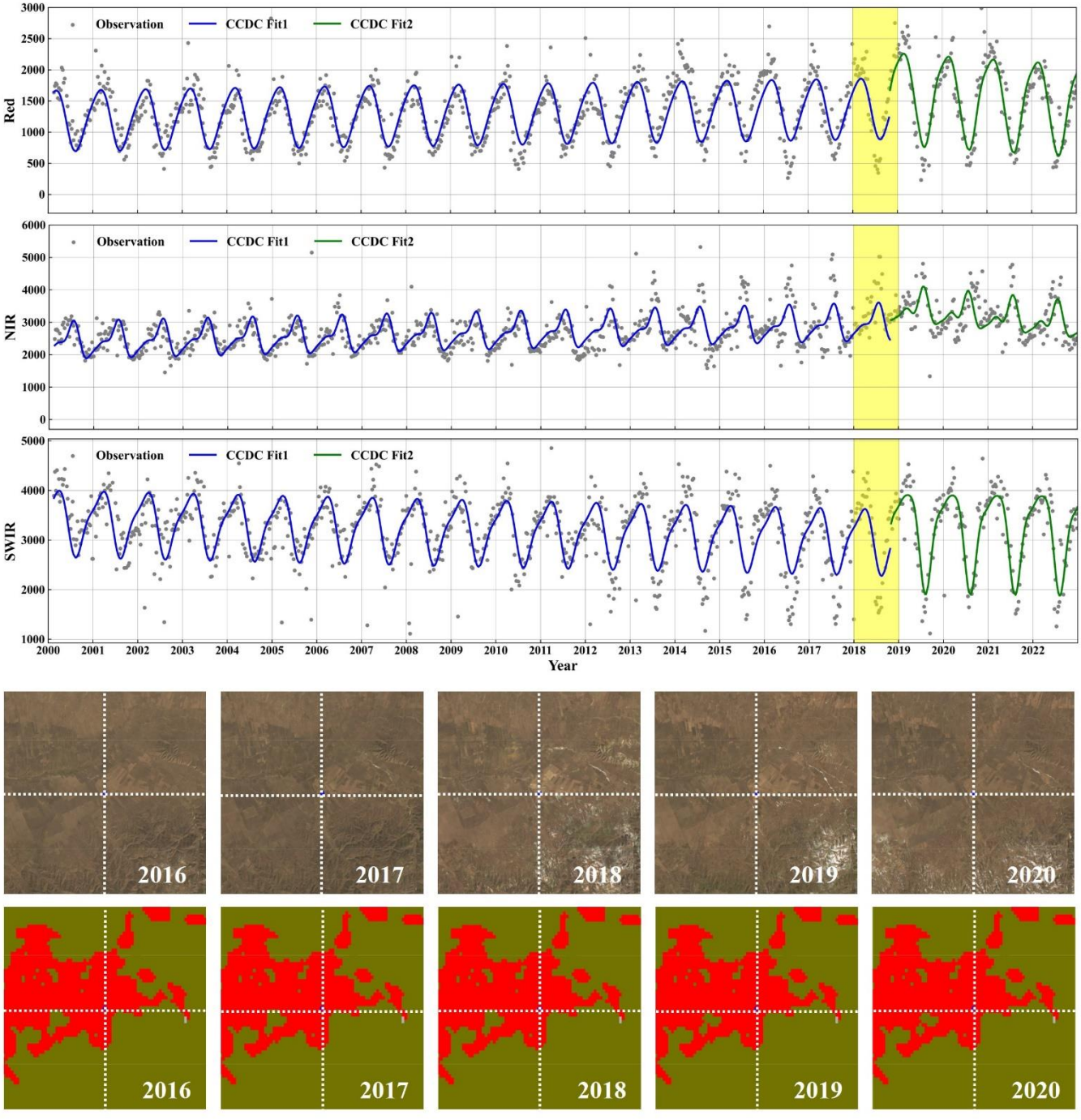

Figure 7. Validation of vegetation mapping: CV sample [100.8570E, 35.2474N] changes detected by CCDC not reflected in Landsat and annual vegetation maps.

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

4.2 Quantitative Assessment of Vegetation Changes on the QTP from 2000 to 2022

This study quantified both potential and actual vegetation changes on the QTP from 2000 to 2022 in terms of area and percentage. Areas where changes were detected by the CCDC algorithm were categorized as potential vegetation changes (Fig. 8). The findings reveal that from 2000 to 2022, the areas without potential unchanged were predominantly found in the forested regions of the southern and southeastern parts of the QTP, comprised approximately 68.21% of the total area (Table 5). Vegetation changes, occurring mostly 1 to 5 times and accounting for about 30.97% of the total area, were more frequent at the junctions of AM, AG, and AD in the central area of QTP. Notably, in the southwestern Ali region's Mapam Yumco area and the northeastern part near Hala Lake, more than 5 changes were detected, making up about 0.82% of the QTP's total area (Table 5).

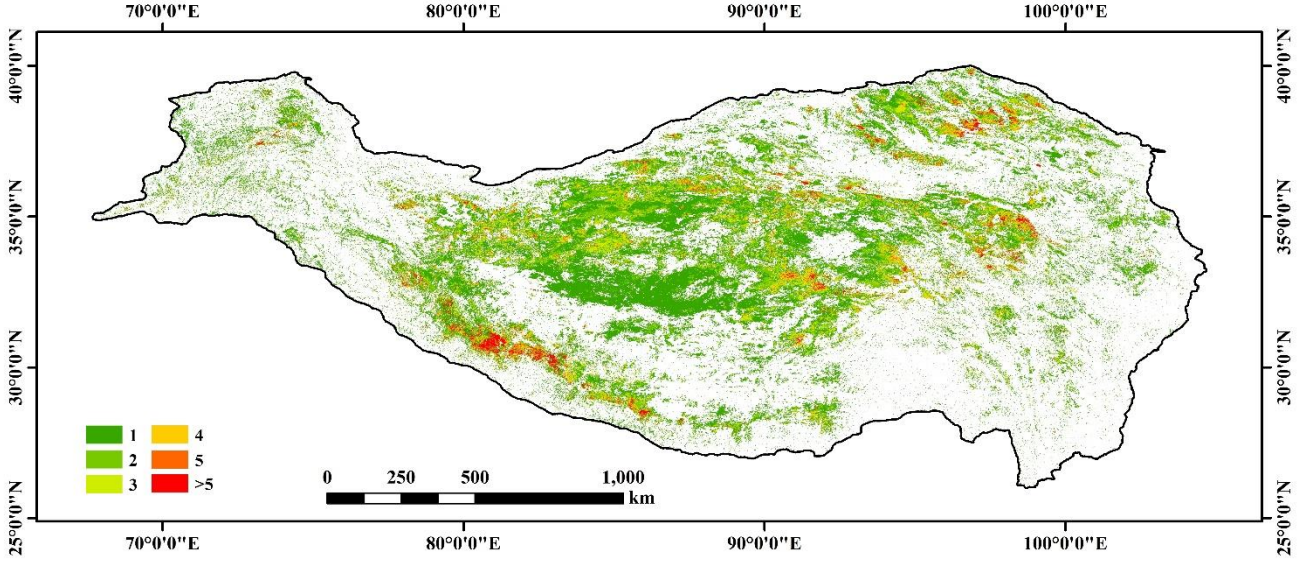

Figure 8. Potential vegetation change areas and number of changes on the QTP from 2000 to 2022.

Based on the annual vegetation maps of the QTP from 2000 to 2022, this study calculated the changes in vegetation types between consecutive years, thereby determining the actual changes in area and the frequency of these changes across the QTP during this period (Fig. 9). Compared to potential unchanged areas, the actual unchanged vegetation areas on the QTP making up about 96.82% of the total area, while the area with real changes (once or more) significantly decreased (Table 5). The majority of actual vegetation changes from 2000 to 2022 occurred 1-2 times, while regions with 3 or more changes accounted for just 0.08% of the QTP's area. These changes primarily involved a reduction in AD and GS areas, and an increase in WA and AM (Table 5).

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

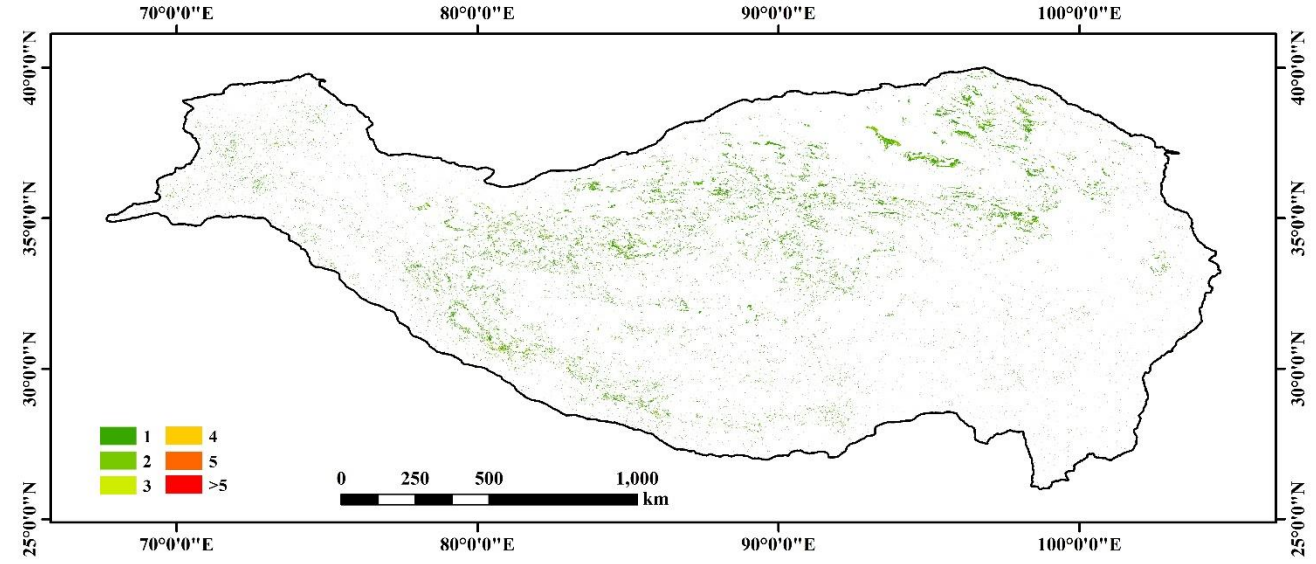

Figure 9. Actual vegetation change areas and number of changes on the QTP from 2000 to 2022.

Table 5. Area Proportion of Potential and Actual Vegetation Change Frequency on the QTP from 2000 to 2022.

Number of Changes	0	1	2	3	4	5	>5
Area Proportion of Potential Vegetation Change Regions (%)	68.21	18.20	7.12	3.28	1.60	0.78	0.82
Area Proportion of Actual Vegetation Change Regions (%)	96.82	2.75	0.34	0.07	0.01	0.00	0.00

In summary, only 3.18% of the QTP experienced changes between 2000 and 2022, a figure substantially lower than what was detected by the CCDC algorithm alone. Therefore, the long-time annual vegetation mapping method developed in this study enhances the temporal consistency of the results, effectively meeting the demands for long-term vegetation change detection.

5 Data availability

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 and SRTM data used in this study were obtained

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

from GEE (http://code.earthengine.google.com, last access: 14 May 2024). The annual precipitation dataset, annual average temperature dataset, and QTP boundary dataset were provided by the National Tibetan Plateau / Third Pole Environment Data Center (https://data.tpdc.ac.cn, last access: 14 May 2024).

6 Conclusions

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 overall accuracy of 80.9% for continuous annual vegetation mapping at a 500 m resolution from 2000 to 2022. The study supports the use of remote sensing data to 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

GS 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.

Competing interests

The authors declare that they have no conflict of interest.

Financial support

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

References

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, 2010.

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

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 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.

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

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.: The shuttle radar topography mission, Rev. Geophys., 45, https://doi.org/10.1029/2005RG000183, 2007.

Feng, M. and Li, X.: Land cover mapping toward finer scales, Sci. Bull., 65, 1604-1606, 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 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.

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.

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, 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 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.

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.

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.

Yang, J. and Huang, X.: 30 m annual land cover and its dynamics in China from 1990 to 2019, Earth Syst. Sci. Data Discuss., 2021, 1-29, https://doi.org/10.5194/essd-13-3907-2021, 2021.

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, 2019.

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., 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, 2019.

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 ,2021.

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, 2022.

https://doi.org/10.5194/essd-2024-193
Preprint. Discussion started: 4 July 2024

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, 2022.

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.