



TPLake-MED: A Monthly Extent Dataset for Lakes on the Tibetan Plateau

Siyu Zhao^{1,2}, Xiang Zhao^{1,2}, Jiacheng Zhao³, Xin Zhang⁴, Xingyu Liu^{1,2}, Chengzhi Yao^{1,2}

¹State Key Laboratory of Remote Sensing and Digital Earth, Beijing Normal University, Beijing 100875, China

5 ²Beijing Engineering Research Center for Global Land Remote Sensing Products, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

³School of Ecology and Applied Meteorology, Nanjing University of Information Science and Technology, Nanjing, Jiangsu 210044, China

⁴Department of Computing and Mathematics, Manchester Metropolitan University, Manchester M15GD, UK

10 Correspondence to: Xiang Zhao (zhaoxiang@bnu.edu.cn)

Abstract. Lakes on the Tibetan Plateau have expanded markedly over recent decades, reflecting complex interactions between the regional water cycle and the cryosphere. Whereas annual datasets capture long-term trends, they often overlook short-term hydrological responses and seasonal transitions that are resolved by monthly observations. Consequently, a systematic understanding of intra-annual lake variability remains limited, largely because most existing datasets are designed for interannual scales, which makes monthly variations and seasonal patterns difficult to characterise. These limitations hinder investigations into the driving mechanisms and complicate assessments of climate-change impacts. To address this gap, we utilised Google Earth Engine (GEE) and the MODIS Surface Reflectance product MOD09A1 (500 m) to construct a monthly vector boundary dataset for lakes larger than 10 km² across the Tibetan Plateau for 2000–2024. Within this dataset, the number of large lakes larger than 50 km² ranged from 142 to 175, and the number of smaller lakes (10–50 km²) varies between 232 and 260 across the study period. A random forest classifier based on spectral indices was developed and validated with 533 balanced water/non-water samples, achieving an overall accuracy of 93.21% and an F1 score of 0.927. To enhance spatial precision, we implemented a boundary optimisation workflow integrating filtering, morphological operations, and geometric rectification, thereby improving agreement between extracted and actual lake extents. Aggregate lake area on the Plateau increased at 34.91 km² per year, and typically reached its annual maximum in September or October. The relative monthly rate of area change showed higher values in the west, lower in the east, and stronger variability centrally; for individual lakes the maximum monthly relative change reached 28.43% from 2000 to 2024. In addition, smaller lakes were more sensitive to environmental change than larger lakes. To our knowledge, this is the first monthly resolution vector dataset of Tibetan Plateau lakes that couples multi-temporal classification with morphological optimisation. The dataset provides critical support for climate-change research, ecological conservation, and policy formulation, and is publicly available at <https://doi.org/10.12443/BNU.RSEC.TPLake-MED20251028>.



1 Introduction

The Tibetan Plateau (TP), often termed the “Roof of the World” and the “Water Tower of Asia”, is the world’s highest and largest plateau and exhibits distinctive environmental characteristics (Qiu, 2018; Bibi et al., 2018). Its ecological significance is paramount, as it contains the headwaters of the Yangtze and Yellow Rivers and supplies major transboundary watercourses to South Asia, Southeast Asia, and the Indochinese Peninsula, supporting the livelihoods and socioeconomic development of vast downstream populations (Immerzeel et al., 2010). The TP also hosts exceptionally diverse alpine biota that underpin unique ecosystems and evolutionary processes (Ding et al., 2020). Through its extreme elevation and strong snow-albedo feedbacks, the TP exerts substantial influence on the Asian monsoon system and even global atmospheric circulation patterns. It is among the world’s most responsive climate regions, shaping key hydrological processes and distinctive ecological regimes (Li and Guo, 2022; Wang, 2016). Moreover, the Plateau sustains rich cultural diversity and plays an irreplaceable strategic role in regional sustainable development and ecological security (Luo and Yang, 2011).

Lakes are integral to the terrestrial water cycle; they modulate hydrological processes, respond to climatic variability, and archive signals of past climate and human activity. They are widely regarded as indicators of global climate (Li et al., 2018; Song et al., 2020). The Tibetan Plateau hosts the world’s largest, most numerous, and highest-elevation cluster of plateau lakes (Zhao et al., 2022). Driven by climate warming, glacier melt, and permafrost degradation, lake area on the Plateau has expanded substantially. Changes in lake number and surface area are now key indicators of regional hydrological processes and climate change (Qiao et al., 2019; Wan et al., 2014). Since the late twentieth century, lakes in endorheic basins have expanded faster than those in exorheic basins, against a backdrop of overall expansion across the Plateau (Zhao et al., 2021; Li et al., 2022). This spatially distinct expansion process not only introduces uncertainties in how different watersheds and lake basins respond to climate change and glacier melt (Chen et al., 2022; Deng et al., 2018), but also directly impacts regional water balance, wetland ecosystem stability, and the safety of surrounding engineering infrastructure through fluctuations in lake area (Xu et al., 2025; Zhu et al., 2025).

With advances in remote sensing technology and machine learning, methods for extracting lake extent from multi-temporal satellite imagery have progressed rapidly (Saha et al., 2024). Traditional water body index-based approaches have found application in large scale dynamic monitoring (Pekel et al., 2016; Han et al., 2019; Yan et al., 2018; Miao et al., 2025). Common indices such as NDWI, MNDWI, and AWEI are extensively utilised with multispectral data from MODIS, Landsat, and Sentinel-2 for water extraction. However, these index-based methods exhibit insufficient accuracy under conditions of high cloud cover, ice and snow coverage, or complex surface conditions (Huang et al., 2020; Lei et al., 2022). Conversely, water body classification methods integrating machine learning (e.g., random forest, support vector machine) with multi-source remote sensing features demonstrate favourable adaptability and generalisation capabilities in the Tibetan Plateau region (Li et al., 2021; Zhou et al., 2022), thereby improving the accuracy of lake boundary identification in complex



environments (Li et al., 2022; Liu et al., 2022; Yang et al., 2025). For instance, Li et al. integrated Landsat-8/9 multispectral data with DEM terrain features using a random forest algorithm to extract water bodies on the Tibetan Plateau, achieving an overall accuracy of 95.84%, significantly outperforming traditional methods like NDWI. Liu et al. integrated Landsat multispectral data, DEM texture features, and textural features using a random forest algorithm to generate a water body distribution map for the Tibetan Plateau, achieving an overall accuracy of 92.9%, markedly superior to single index methods. Yang et al. utilised the Google Earth Engine (GEE) platform to classify wetland types using a random forest approach, integrating Landsat time series data from 2000 to 2023 with DEM texture features. This achieved an average overall accuracy of 88.45%, effectively distinguishing lakes from marshy meadows.

Several vector datasets of lake area on the Tibetan Plateau are currently available and widely utilised (Zhang et al., 2018; Pang et al., 2021; Wang et al., 2023; Zhang et al., 2019; Zhang et al., 2020; Ran et al., 2023). For instance, Zhang et al. constructed triennial area time series for 364 lakes larger than 10 km² between 1970 and 2013; Pang et al. generated continuous area time series for 20 lakes larger than 100 km² from 1976 to 2019; Wang et al. (2023) provided annual mean area data for 180 lakes from 1986 to 2020; Zhang et al. acquired lake area data for 16 periods from 1970 to 2022; Zhang et al. extracted monthly scale area for lakes larger than 50 km² from 2015 to 2018 based on Sentinel-1 SAR data; Ran et al. generated a dataset of monthly lake area changes for lakes larger than 30 km² from 2015 to 2020 using multi-source remote sensing imagery. However, these interannual datasets typically select only a single specific phase (e.g., one summer scene per year) to represent annual lake extent, lacking representativeness and failing to capture intra-annual dynamics and seasonal variations. This limits their utility for analysing responses to extreme climate events and short-term hydrological processes (Yang et al., 2017). Secondly, although studies have begun to provide monthly scale data, their temporal coverage remains relatively short (e.g., 3–5 years), making it difficult to support in-depth analysis of long-term lake change trends and their driving factors (Li et al., 2025); Furthermore, existing datasets emphasise larger lakes, with insufficient coverage of small and medium lakes. Spatial resolution and temporal continuity are often difficult to reconcile (Ma et al., 2022). Simultaneously, inconsistencies in remote sensing data sources, extraction methods, and temporal phase selection across different datasets reduce consistency and comparability (Gu et al., 2023).

Few studies have systematically undertaken long-term, high-frequency, comprehensive monitoring of monthly lake area change across the Tibetan Plateau. High temporal resolution continuous data series (such as monthly data spanning over a decade) can effectively fill gaps in temporal granularity and sequence continuity within existing datasets. This provides crucial data support for high precision lake water volume dynamics modelling, analysis of driving mechanisms, and research into responses to climate change (Liu et al., 2024; Khandelwal et al., 2022). Consequently, this study aims to: (1) develop an automated extraction method for Tibetan Plateau lakes based on multi-temporal MODIS remote sensing imagery, incorporating random forest classification and morphological optimisation; (2) construct a high precision monthly scale vector boundary dataset for lakes exceeding 10 km² in area from 2000 to 2024, and systematically validate its accuracy; (3) compare



the dataset with existing products to evaluate advantages in accuracy, temporal continuity, and spatial consistency; (4) reveal
95 the spatiotemporal patterns of lake area change across the Tibetan Plateau over the past two decades, providing crucial
foundational data for regional hydrological process modelling, cryosphere change research, and climate change impact
assessments.

2 Study area

The Tibetan Plateau (26°00'–39°47' N, 73°19'–104°47' E) exceeds 4,000 metres in average elevation and is the highest and
100 largest plateau in the world. It is often referred to as the “Roof of the World” and the “Water Tower of Asia” (Long et al., 2022;
Zhang et al., 2023). The Tibetan Plateau is rich in lake resources, with over 1,500 lakes of various sizes, including
approximately 180 lakes larger than 50 square kilometres. The total lake area is approximately 44,993 square kilometres,
accounting for about 50% of the total lake area in China (Huang et al., 2025). These lakes are primarily saline lakes and salt
lakes, including the famous Nam Co, Qinghai Lake, and Qarhan Salt Lake (Yan et al., 2017). The distribution of lakes across
105 the Tibetan Plateau is influenced by both the altitude and geographical features. The larger lakes, those over 50 km² are
predominantly located in regions with relatively flat terrain and higher elevations, such as the central and northeastern parts of
the plateau. Smaller lakes (10–50 km²) are scattered more widely across areas with more complex terrain, including the
southern and southeastern regions of the plateau. These areas are characterized by rugged mountain systems, such as the
Himalayas and Hengduan Mountains, which shape the hydrological landscape and influence lake distribution.

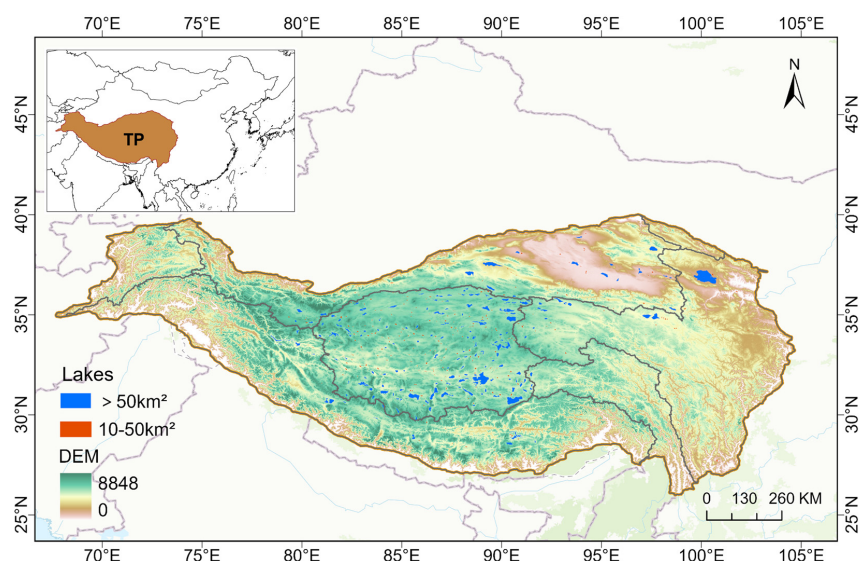


Figure 1: Overview map of the study area. TP denotes the Tibetan Plateau.



3 Method

Figure 2 summarised the monthly lake area data generation workflow for this study: First, the monthly baseline imagery is constructed by selecting the image with the lowest cloud cover from MOD09A1 (500 m) for each month from the GEE platform, followed by cloud masking and temporal gap-filling interpolation. Second, water bodies are classified with a random forest classifier using multispectral bands and spectral indices, and assessed accuracy with confusion matrices and multi-algorithm comparisons. Third, post-classification processing included noise reduction, missing pixel detection and filling, overlaying the JRC global surface water mask, morphological smoothing, boundary IoU constraints, and invalid geometry repair to optimize boundaries. Finally, we validated accuracy against multi-source data and existing products, and conducted spatial heterogeneity and time series analysis to reveal spatiotemporal evolution patterns of lake area.

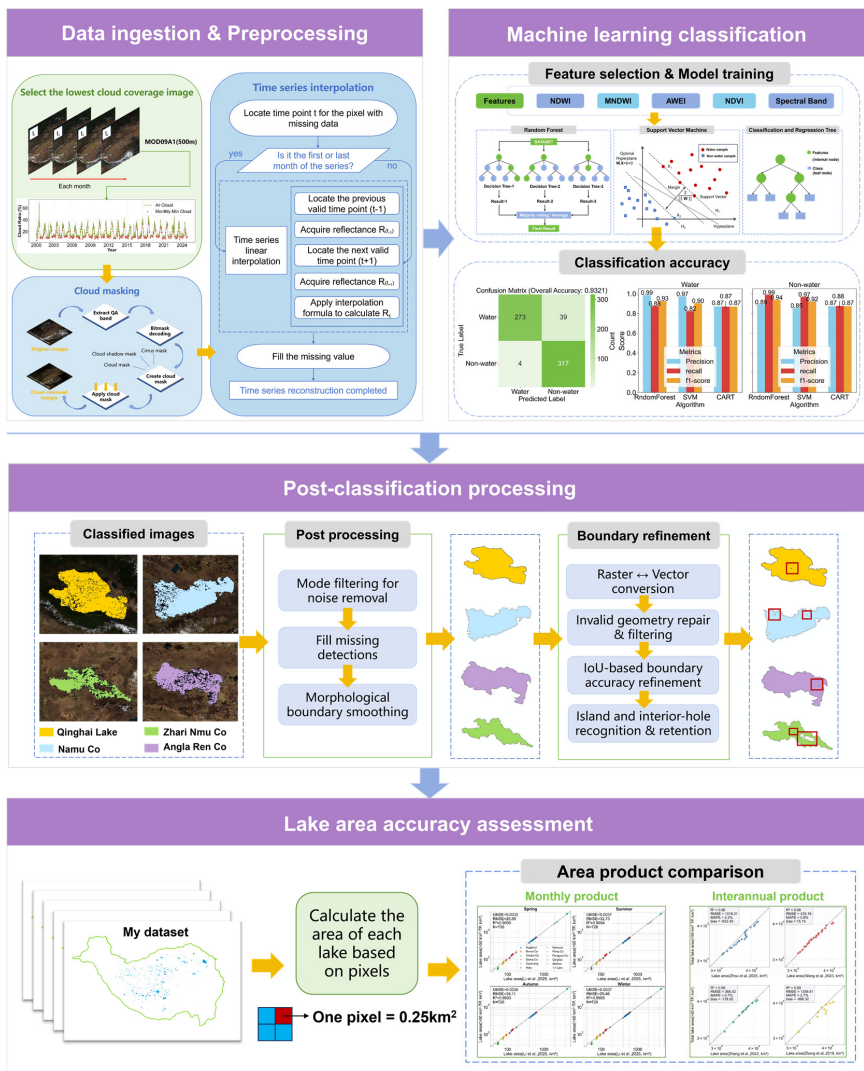


Figure 2: Framework of monthly lake data generation in this study. The process comprises four steps: data preprocessing to select



monthly imagery with the lowest cloud cover (Step 1), random forest classification to obtain preliminary water body extraction results (Step 2), post-classification lake morphological and boundary optimisation (Step 3), lake area calculation and comparative validation (Step 4).

3.1 Data preprocessing

We utilized the MODIS MOD09A1 surface reflectance (SR) product with a spatial resolution of 500 metres. The MOD09A1 product provides atmospheric-corrected surface reflectance data under favourable atmospheric conditions and is directly accessible via GEE. For each month, we selected the MOD09A1 scene with the lowest cloud cover as the monthly baseline image for subsequent lake area extraction. Before using the base image, we further processed it using the product's built-in StateQA band for cloud masking (by bitwise operation to detect the 10th cloud flag) to remove residual cloud and cloud shadow-affected pixels. To address data gaps remaining after cloud masking, we employed a time series linear interpolation method. This involves stacking the best monthly images in chronological order to generate a reflectance time series for each pixel, then interpolating missing values using the reflectance values from the available pixels in the preceding and following months, as shown in Eq. (1):

$$R_t = R_{t-1} + (t - t_{-1}) \times \frac{R_{t+1} - R_{t-1}}{t_{+1} - t_{-1}} \quad (1)$$

Where R_t is the reflectance of the pixel to be filled, and $t_{\pm 1}$ are the effective time points adjacent to the missing value. Missing data in the first and last months were filled using the nearest pixel values to ensure the spatio-temporal continuity of lake boundary extraction. Based on these monthly image data that had undergone rigorous screening, cloud masking, and interpolation processing, we further performed water body classification to extract lake area. Selecting the lowest cloud scene each month minimised cloud obstruction, yielding the clearest monthly surface observation and providing a robust basis for accurate lake boundary identification. Subsequent cloud masking ensured that reflectance utilised for classification primarily originates from true surface water or non-water, thereby reducing misclassification due to cloud contamination.

3.2 Machine learning classification

To accurately extract water bodies from MODIS imagery, we compared three machine learning algorithms—random forest (RF), support vector machine (SVM), and classification and regression tree (CART)—under identical input features and training sample conditions, and selected the optimal model for large scale water body extraction across the Tibetan Plateau. RF constructs an ensemble of decision trees and aggregates their predictions by majority voting (Breiman, 2001; Wen et al., 2023). SVM identifies the optimal separating hyperplane to maximize the classification margin, performing well on both linear and non-linear problems and particularly excelling in high dimensional sparse datasets (Zhou et al., 2009; Kuter, 2021). CART recursively splits data based on the Gini index without any distributional assumptions, effectively handling non-linear



relationships and evaluating the importance of multi-source features such as spectral, textural, and index-based variables (Chen et al., 2015).

3.2.1 Calculation of Input Features

We collected ten variables from the MOD09A1 imagery, including six original bands and four calculated indices (Zhang et al., 2021). These variables were utilised to discriminate surface water from non-water bodies in the Tibetan Plateau. The four indices included the Normalised Difference Vegetation Index (NDVI; Tucker, 1979), the Normalised Difference Water Index (NDWI; McFeeters, 1996), the Modified Normalised Difference Water Index (MNDWI; Xu, 2006), and the Automatic Water Extraction Index (AWEI; Feyisa, 2014), with the formulas as follows:

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}} \quad (2)$$

$$NDWI = \frac{\rho_{GREEN} - \rho_{NIR}}{\rho_{GREEN} + \rho_{NIR}} \quad (3)$$

$$MNDWI = \frac{\rho_{GREEN} - \rho_{SWIR1}}{\rho_{GREEN} + \rho_{SWIR1}} \quad (4)$$

$$AWEI = 4(\rho_{GREEN} - \rho_{SWIR1}) - (0.25\rho_{NIR} + 2.75\rho_{SWIR2}) \quad (5)$$

where ρ_{GREEN} , ρ_{RED} , ρ_{SWIR1} and ρ_{SWIR2} denote the surface reflectance in the green, red, near-infrared, and the first and second short-wave infrared bands of the MODIS Surface Reflectance product, respectively.

3.2.2 Model training and parameter optimization

Classifiers are highly sensitive to sampling design (Belgiu et al., 2016). Appropriate training samples are critical for the classification accuracy and stability of models (Xie, 2022). In the absence of suitable labelled sample data, we manually labelled 2,420 sample points on the GEE platform as the training dataset, including 1,275 points labelled as water bodies and 1,145 points labelled as non-water bodies (Fig. 7a), ensuring that the sample points were uniformly distributed across the study area. Additionally, we utilised the JRC dataset (Kibret et al., 2021) as auxiliary reference for classification. Finally, we randomly divided 70% of the samples into a training set and the remaining 30% into a validation set.

For the RF classifier, hyperparameters were tuned by grid search with cross validation to ensure generalisation capability (Zhang et al., 2021). The number of trees (numberOfTrees) was set to 150, the minimum number of samples per leaf node (minLeafPopulation) was set to 5, and the sampling ratio (bagFraction) was set to 0.5, meaning each tree was trained on a random 50% subset of the data with replacement (Liu et al., 2015). These settings increased ensemble diversity and mitigated overfitting. SVM and CART classifiers are trained using the same dataset and parameter settings on the GEE platform. SVM was implemented with a radial basis function (RBF) kernel to capture non-linear boundaries, while CART utilised Gini impurity as the splitting criterion for optimal threshold selection.



180 3.2.3 Model evaluation and selection

To assess and compare the performance of different models, a confusion matrix-based evaluation was conducted. The predicted labels from each model were compared with the true labels to quantify classification accuracy, with overall accuracy adopted as the primary metric (Congalton, 1991). To further compare the three algorithms, we computed precision (Eq. (6)), recall (Eq. (7)), and F1 score (Eq. (8)) from the confusion matrix. We then compared the three models under identical experimental conditions (Fawcett, 2006).

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

Where TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative) represent the numbers of correctly and incorrectly classified samples. TP represents the number of pixels correctly classified as water, while TN indicates non-water pixels correctly identified as non-water. FP refers to non-water pixels that were incorrectly classified as water, and FN represents water pixels that were misclassified as non-water. Based on these indices, the classification performance of the three models was quantitatively compared to determine the most accurate and stable approach for water body extraction.

195 3.3 Post-classification processing

Although the preliminary results generated by the random forest classifier effectively identified water bodies with significant spectral features, they still faced two key limitations: first, the absence of terrain constraint mechanisms led to false water bodies in mountain shadow areas and slope-related misclassification; second, spectral confusion caused salt-and-pepper noise, significantly reducing the geographical plausibility of boundaries (Fu et al., 2022). To overcome these limitations and meet the accuracy requirements for large scale lake area extraction, we constructed a complete automated post-processing workflow for lake classification, covering core modules such as multi-file batch processing, dynamic coordinate system unification, geometric topology repair, and fine-grained retention of internal islands (Huang et al., 2017). This workflow enabled precise correction of lake water body boundaries by efficiently processing classified outputs. Morphological optimisation and a boundary-control mechanism based on the intersection-over-union ratio (IoU) improved geometric integrity and spatial realism, effectively suppressing spurious water detections and classification noise (Chen et al., 2022; Li et al., 2020). An accompanying island screening step further improved the extraction of water features. Figure 3 illustrates the processing workflow and example results for selected lakes.

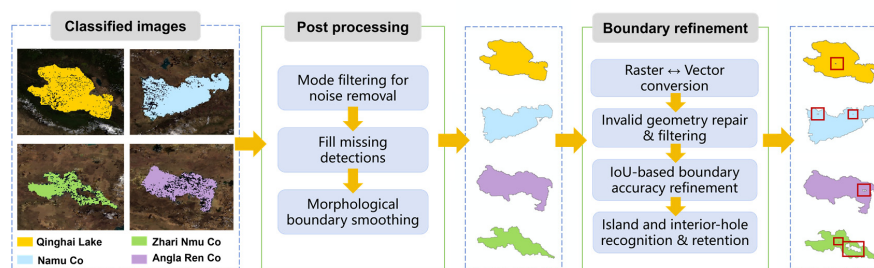


Figure 3: Schematic diagram of the lake remote sensing image classification and post-processing workflow. From left to right, the images represent the classification result, the morphological correction result, and the final result after precise hole retention.

3.3.1 Lake boundary extraction through filtering and morphological optimisation

Using a random forest classifier, we obtained preliminary classifications of water and non-water across the study area. Due to terrain constraints and spectral selection differences in water body distribution, such as slope and mountain shadow limitations, we did not apply terrain filtering. As a result, spectral confusion produced salt-and-pepper noise, that is, isolated misclassified pixels (Chi et al., 2017). To reduce such noise, we first filtered out classified water body pixels with fewer than nine neighbouring pixels. Additionally, to leverage high confidence water body prior knowledge, we integrated the JRC Global Surface Water Dataset (occurrence $\geq 80\%$) as a mask applied to the preliminary classification results. We then applied spatial smoothing and morphological operations for optimisation. A 3×3 majority filter was utilised to smooth the classifications, suppressing noise and aggregating neighbouring pixels. Subsequently, on the GEE platform, we applied a circular kernel with a radius of 1 pixel to the smoothed water body classification results, performing dilation (focal_max) and erosion (focal_min) operations once each. Dilation connected adjacent small water patches and promotes more complete inclusion of water edges, especially near wetlands or unclassified areas, whereas erosion removed small artefacts introduced by dilation, yielding smoother boundaries. Finally, all post-processing results were vectorised and output for subsequent analysis.

3.3.2 Boundary refinement techniques using IoU and geometric repair

Based on the vectorised water body boundaries generated through morphological optimisation, automated operations were implemented using a multi-file batch processing framework. During the data loading phase, dynamic coordinate system unification ensured that all input files were spatially referenced to the reference layer (Zhou et al., 2025), guaranteeing the reliability of subsequent overlay analyses. During processing, vector lake boundaries were first subjected to geometric repair, filling unclosed gaps within polygonal features and correcting topological anomalies such as self-intersections and duplicate vertices to ensure the geometric integrity of water body polygons. Second, a boundary control mechanism based on the intersection-over-union ratio (IoU) precisely constrained lake boundaries, and in conjunction with the reference layer, refined



the retention of internal holes and islands within lakes. Finally, the corrected global lake vector file was output via batch processing, with manual visual inspection performed to ensure the completeness and accuracy of lake boundaries.

3.4 Verification methods

235 To assess seasonal performance, we validated lake area extraction for spring, summer, autumn, and winter. We compared our results with the dataset of Li et al. (2025). Correlation analysis utilised two monthly scale lake datasets from 2001 to 2023. A total of 11 lakes were selected from the two sets, with 3, 4, and 4 lakes selected from the small, medium, and large lake types, respectively, yielding a total of 726 sample data points. The correlation analysis of the data was quantified using root mean square error (RMSE; Eq.(9)), unbiased RMSE (URMSE; Eq.(11)), and the coefficient of determination (R²; Eq.(12)).

240 Additionally, the extracted monthly average lake area data were aggregated to the annual scale and compared with the existing dataset for validation (Table 1).

Table 1: Information on the reference lake area products utilised for comparison.

Dataset description	Image used	Lake numbers	Time span	Source
Annual 30-m lake maps on the Tibetan Plateau	Landsat	6158	1991–2023	Zhou et al., 2025
Annual area dataset of lakes over 50 km ² on the Tibetan Plateau	Landsat	180	1986–2020	Wang et al., 2023
Time series dataset of lake area on Tibetan Plateau for the past 100 years	Landsat	1236	1920–2020	Zhang et al., 2022
The lakes larger than 1km ² in Tibetan Plateau (v3.1)	Landsat	1400	1970–2022	Zhang et al., 2019

Note: All the above data are from the National Tibetan Plateau Science Data Center.

Interannual lake area datasets from Zhou et al. (2025), Wang et al. (2020), Zhang et al. (2022), and Zhang et al. (2019) were utilised as comparison baselines. Quantitative assessment utilised the coefficient of determination (R²), root-mean-square error (RMSE), mean absolute percentage error (MAPE; Eq. (13)), and bias (Eq. (10)) (Hui et al., 2025). By constructing comparison scatter plots, the consistency between the study data and the comparison datasets was analysed to validate the data quality and the effectiveness of the methods (Li et al., 2015).

245

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (9)$$

$$Bias = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \quad (10)$$

250 $URMSE = \sqrt{RMSE^2 - Bias^2} \quad (11)$



$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (12)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (13)$$

In the above formula, n represents the number of samples, *i.e.*, the total number of data points utilised for evaluation; i is utilised to iterate through each sample, from the first ($i = 1$) to the n th ($i = n$). y_i represents the true value of the i th sample; $\hat{y}_i$ represents the predicted value of the i th sample calculated according to the model. $\bar{y}$ represents the average of all true values; $\hat{y}_i - y_i$ represents the prediction error or residual of the i th sample, indicating the degree of deviation of the predicted value from the true value.

4 Results

4.1 Changes in Tibetan Plateau lakes (2000–2024)

260 4.1.1 Yearly lake expansion: patterns and responses

By combining random forest classifier with morphological methods, We generated a dataset (TPLake-MED) of lake boundary ranges for lakes larger than 10 km² on the Tibetan Plateau from 2000 to 2024, with a spatial resolution of 500 m. Among lakes larger than 50 km², most showed increasing area, particularly in the central and north-eastern Plateau. A minority decreased in size, mainly in the western and some marginal areas (Fig. 4a). The decrease in the western region may have been influenced by tectonic activity and human water abstraction, while changes in marginal areas may be associated with enhanced evaporation (temperature increase of 1.2°C per decade) (Wang et al., 2024). As of 2024, the fastest growing lakes by relative area change were Selin Co (2,369.5 km²; +21.6%), Aqikkol (570.5 km²; +17.7%), Ayagekumuli (1,077.75 km²; +20.3%), Duoersuo Co (1,031.5 km²; +13.5%), Hulu Lake (311.25 km²; +18.2%), and Ruola Co Lake (270.5 km²; +11.3%) (Fig. 4b). The total lake area shows a significant upward trend, with an average annual growth rate of 34.91 km² per year, and the correlation is significant ($p < 0.05$), overall expansion of high-elevation lakes is evident. In particular, the total area of lakes larger than 50 km² increased by 32.5% compared with that in 2000 (Zhu et al., 2019) (Fig. 4c). The six fastest growing lakes exhibited marked interannual fluctuations but an overall increase. Annual area for these lakes fluctuated substantially, with the smallest lake area occurring in March and April, when precipitation, temperature, and evapotranspiration were all relatively low. The largest lake area occurred in September and October (Fig. 4d), when precipitation was relatively abundant, while temperature and evapotranspiration were relatively low. This pattern indicates joint control of lake area by precipitation and evapotranspiration (Li et al., 2022).

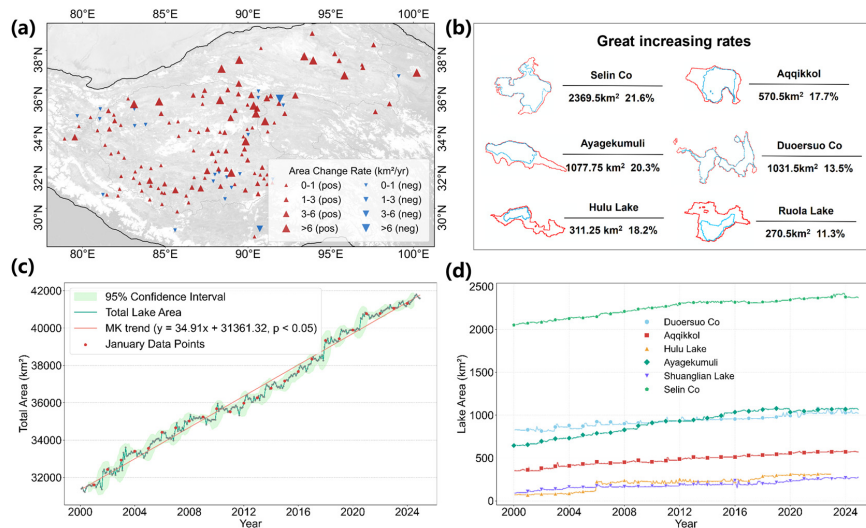


Figure 4: Spatial-temporal characteristics of lake area changes and trends in typical lakes from 2000 to 2024. (a) The spatial distribution of interannual lake area change rates. Triangular symbols indicate the interannual area change rates (km² per year) of lakes on the Tibetan Plateau. Red triangles indicate area increase, and blue triangles indicate area decrease. The size of the triangles represents the magnitude of the change rate. (b) The fastest growing lakes by relative area change on the Tibetan Plateau (2024). Below the horizontal line are the lake area in 2024 and the percentage expansion from 2000 to 2024. Blue contours indicate the boundary extent in 2000, while red contours represent the boundary extent for the same month in 2024. (c) The monthly trend of total lake area. (d) The monthly trend of typical lake areas.

4.1.2 Monthly area change: heterogeneity and scale

The monthly rate of lake area change on the Tibetan Plateau exhibited marked spatial variation and scale dependence. By comparing intra-annual relative change rates across years, we found that 2005 showed the largest variability and therefore selected it as a representative year for detailed analysis. Within that year, the maximum monthly relative change rate observed for an individual lake reached 28.43%. Longitudinally, lakes in the western plateau region between 80–85° E showed higher rates of change with pronounced fluctuations. As longitude increased eastward, the rate of change gradually decreased, stabilizing notably east of 95° E. Latitudinally, change rates were generally higher between 30–34° N, with the strongest variability between 32–34° N. North of 36° N, rates diminished markedly (Fig. 5a). This spatial variation was closely linked to regional climatic conditions, the western region experienced significant seasonal precipitation variations and intense evaporation, leading to pronounced lake water fluctuations. In contrast, the eastern region, influenced by stable monsoons and supplemented by glacial runoff, exhibited relatively smoother changes.

Approximately 70% of lakes had change rates below 0.05, with a right skewed distribution peaking at 0–0.05. The number of lakes decreased sharply as change rates increased, consistent with an inverse J-shaped pattern, indicating overall stability in the plateau lake system (Fig. 5b). The lake change rate was significantly negatively correlated with lake area: larger lakes exhibit lower monthly change rates and reduced data dispersion. Specifically, lakes smaller than 150 km² exhibited the highest variation rates and greatest variability, with multiple outliers present. Lakes between 150–500 km² showed intermediate



variation rates and variability. Lakes between 500–1000 km² demonstrated a marked decrease in variation rates with a more concentrated distribution. Lakes larger than 1000 km² had the lowest variation rates and minimal variability, with only a few isolated (Fig. 5c). This scale dependence largely reflected the greater storage and stronger buffering capacity of large lakes, which enabled them to withstand short-term climatic fluctuations. Conversely, small lakes exhibited heightened sensitivity to environmental factors such as precipitation, evaporation, and runoff. Our data (TPLake-MED) support the “area-stability” hypothesis, providing an empirical basis for lake classification, dynamic monitoring, and climate change impact assessment on the Tibetan Plateau (Zhu et al., 2025).

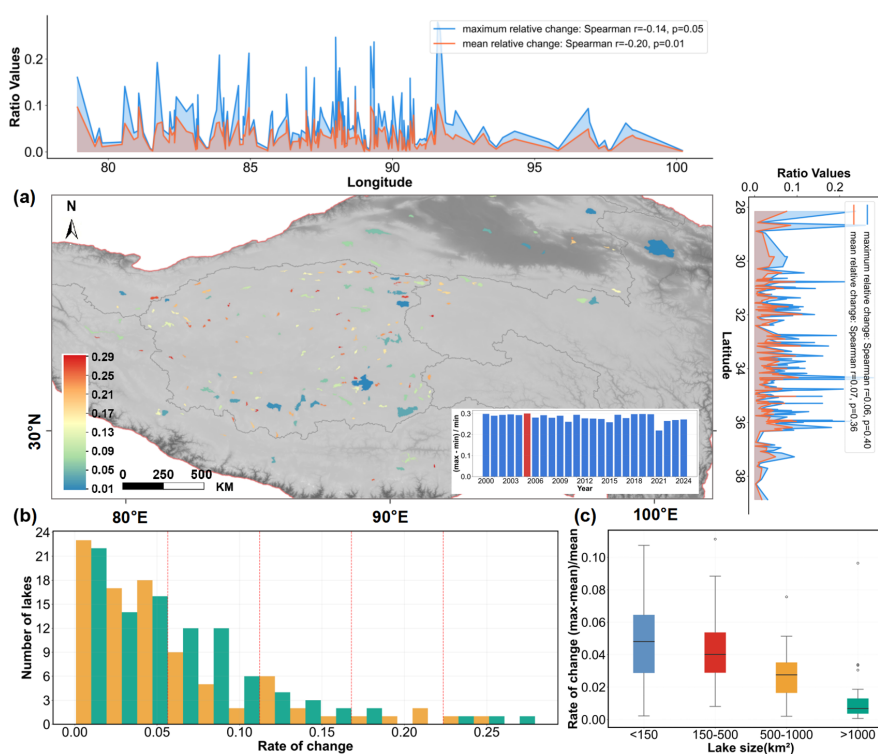


Figure 5: Spatial distribution and statistical characteristics of inter-monthly lake area change rates on the Tibetan Plateau. (a) Temporal variation and spatial distribution of inter-monthly change rates, with the orange and blue lines representing mean and maximum relative changes, the red bars indicating the year of maximum rate, and the coloured background showing lake specific change rates in that year. (b) The frequency distribution histogram of change rates. (c) The box plots of change rates for lakes of varying sizes.

Inter-monthly lake area variation across the Tibetan Plateau exhibited spatial heterogeneity, with many lakes showing pronounced fluctuations in surface area. Monthly area differences of 6.50–26.38 km² occurred predominantly in the Plateau interior, indicating heightened sensitivity to climatic shifts. Differences of 3.91–6.50 km² were widely distributed across the interior, indicating moderate variability; and differences of 0.25–3.91 km² occurred mainly along Plateau margins and transition zones, representing relatively stable lake systems (Fig. 6a). Figures (b–i) show the intra-annual maximum and minimum boundary differences of the lakes exhibiting the most significant variations. Overlay analysis of satellite imagery and vector



boundaries further validated the extracted outlines and substantiated intra-annual fluctuations, whereby observed area changes directly reflected the magnitude of seasonal expansion and contraction. This multi-scale presentation revealed macro-spatial patterns of Plateau lake dynamics and supported data quality and analytical reliability through detailed validation of representative lakes. It provided evidence to deepen understanding of the stability and vulnerability of Plateau lake ecosystems.

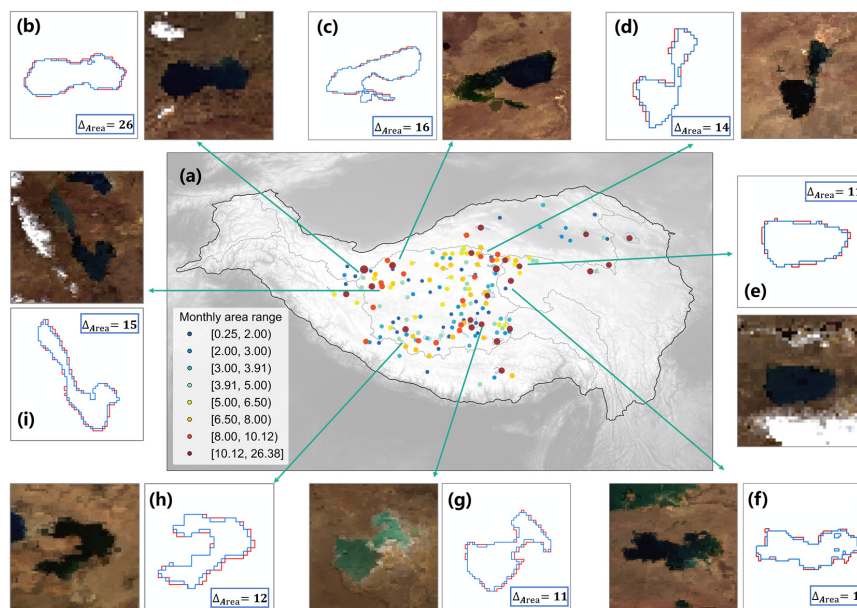


Figure 6: Spatial distribution of inter-monthly lake area variation on the Tibetan Plateau and verification of boundary changes for representative lakes. (a) The spatial distribution pattern of inter-monthly lake area variation (max-min) across the Tibetan Plateau. (b–i) The comparative boundary change diagrams for eight lakes with the most pronounced intra-annual fluctuations. The red contours indicate the maximum intra-annual boundary extent, the blue contours denote the minimum intra-annual boundary extent, and the numerical labels represent the corresponding expanded lake area (km²). The left subfigures use MODIS/Terra MOD09A1 Collection 6.1 8-day (500 m) imagery composited to monthly products and processed in Google Earth Engine; data © NASA EOSDIS/LP DAAC.

4.2 Validation of lake extraction accuracy

The accuracy assessment of binary classifiers is crucial in lake classification practices (Olofsson et al., 2014; Stehman et al., 2019). A 30% subset of the labelled samples was reserved as a validation set to evaluate binary classification of water and non-water by the three algorithms. We present a comprehensive assessment based on this set, including the spatial distribution of validation samples, confusion matrix analysis, and cross algorithm performance comparison. The validation samples totalled 633, exhibiting a relatively uniform spatial distribution and effectively covering the main geographical units and different terrain conditions of the study area (Fig. 7a), providing a representative sample basis for classification accuracy assessment. The confusion matrix results indicated that the random forest algorithm performed excellently in water body classification, with an overall classification accuracy of 93.21%, with 273 correctly classified water body samples and only 39



misclassified as non-water bodies, while the SVM and CART algorithms achieved overall accuracies of 84.98% and 85.75%, respectively. For non-water bodies, 317 samples were correctly classified by the random forest, with only 4 misclassified as water bodies (Fig. 7b). Overall, random forest showed higher accuracy and stability than SVM and CART in distinguishing
345 water from non-water.

Table 2: Comparison of accuracy between different classification algorithms for the “water bodies” and “non-water bodies” categories.

Algorithm	Class	Precision	Recall	F1-score
Random Forest	Water	0.986	0.875	0.927
SVM	Water	0.972	0.871	0.869
CART	Water	0.903	0.824	0.870
Random Forest	Non-water	0.988	0.890	0.936
SVM	Non-water	0.967	0.921	0.854
CART	Non-water	0.873	0.875	0.874

Further quantitative comparison analysis was conducted on three machine learning algorithms—random forest (RF), support vector machine (SVM), and classification and regression tree (CART), using three metrics: precision, recall, and F1-score
350 (Table 2). For the water class, RF performed best across all metrics, with precision, recall, and F1 score of 0.90, 0.93, and 0.97, respectively, exceeding SVM (0.88, 0.82, 0.90) and CART (0.87, 0.87, 0.87) (Fig. 7c). For the non-water class, RF also maintained its leading advantage, with precision, recall, and F1 score of 0.99, 0.94, and 0.97, all higher than the other two algorithms (Fig. 7d). These results indicate that RF offers strong performance and practical value for remote sensing classification of lake water on the Tibetan Plateau, providing a useful reference for subsequent monitoring of lake area
355 dynamics and spatiotemporal analyses (Wang, 2023).

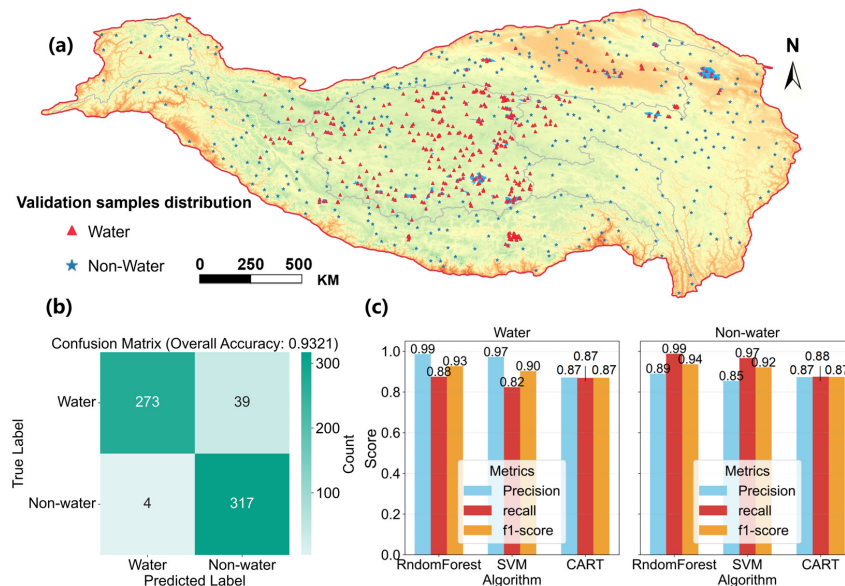


Figure 7: Validation sample distribution and classification accuracy of the Random Forest model for lake water and non-water body identification. (a) The distribution of validation sample points for the classification models. (b) The confusion matrix of the Random Forest classification model. (c–d) The bar chart compares the main evaluation metrics (Precision, Recall, and F1-score) of different classification algorithms for the two categories of “water body” and “non-water body.”

Figure 8 presents the water body boundary extraction results for five representative lakes—Qinghai Lake, Nam Co, Selin Co, Zhari Namu Co, and Angla Ren Co—across spring, summer, autumn, and winter. The extracted red boundaries aligned closely with the water bodies depicted in corresponding seasonal satellite imagery, accurately capturing shoreline inflections, peninsulas, and bay entrances. This indicates robust performance of the seasonal extraction under varying surface conditions, including snow cover, cloud shadow, and low solar elevation. From the seasonal variation patterns, most lakes exhibited the typical seasonal variation pattern of highland lakes, with the largest area in summer and relatively smaller area in winter. Among them, Qinghai Lake, as the largest lake in the study area, showed relatively stable seasonal variation, with similar water body boundaries in spring and winter, and a significant expansion of the water body range in summer and autumn, reflecting the lake’s strong water body stability (Li et al., 2012). Nam Co exhibited significant fluctuations in area across the four seasons, with relatively smaller water areas in spring, reaching a maximum in summer, and gradually contracting in autumn and winter, reflecting the lake’s sensitive response to seasonal climate changes. Medium lakes such as Selin Co, Zhari Nam Co, and Angla Ren Co also followed the basic pattern of expansion in summer and contraction in winter, but the specific magnitude of changes varied among lakes. Such variability likely reflected differences in location, elevation, catchment characteristics, and local climate. Overall, the seasonal comparison showed high consistency between extracted boundaries and the imagery under varying conditions, supporting the method’s efficacy and stability. These seasonal boundary variations provided spatial



evidence for understanding lake dynamics and hydrological processes on the Plateau, and offered a basis for elucidating lake responses to climate change.

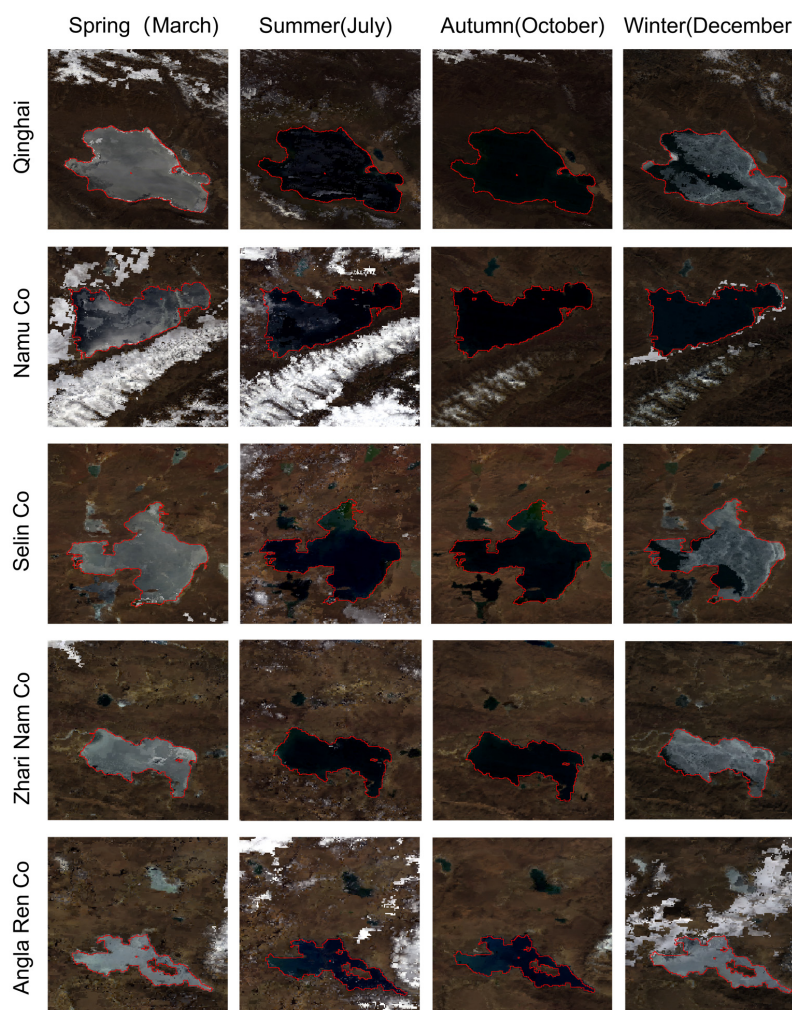


Figure 8: Typical lake boundary extraction results for spring, summer, autumn, and winter. The background uses MODIS/Terra MOD09A1 Collection 6.1 8-day (500 m) imagery composited to monthly products and processed in Google Earth Engine; data © NASA EOSDIS/LP DAAC; overlays © Authors.

4.3 Comparison with other products

4.3.1 Seasonal product comparison

To validate the accuracy of our lake area estimates, we compared results for 11 lakes with the dataset of Li et al. (2025). Using two sets of lake data from 2001 to 2023 for comparison analysis, the results showed that the area extraction accuracy for all four seasons was excellent: the spring correlation coefficient R^2 reached 0.9995, URMSE was 0.0032, and RMSE was 28.88; summer R^2 was 0.9994, URMSE was 0.0032, and RMSE was 32.75; autumn R^2 was 0.9993, URMSE was 0.0036, and RMSE



was 28.88; winter R2 was 0.9995, URMSE was 0.0037, and RMSE was 29.46 (Fig. 9). Scatter points clustered tightly around the 1:1 line, indicating high reliability and seasonal stability of the extraction. Results for both large lakes (e.g. Qinghai Lake, Nam Co, Selin Co) and numerous small and medium lakes were highly consistent with the reference data. These findings supported high overall accuracy, especially for large lakes with clear boundaries, where the extraction results were almost identical. Therefore, it can be concluded that the lake area extraction results obtained using the proposed automated extraction algorithm based on random forest classification and morphological post-processing are accurate and reliable.

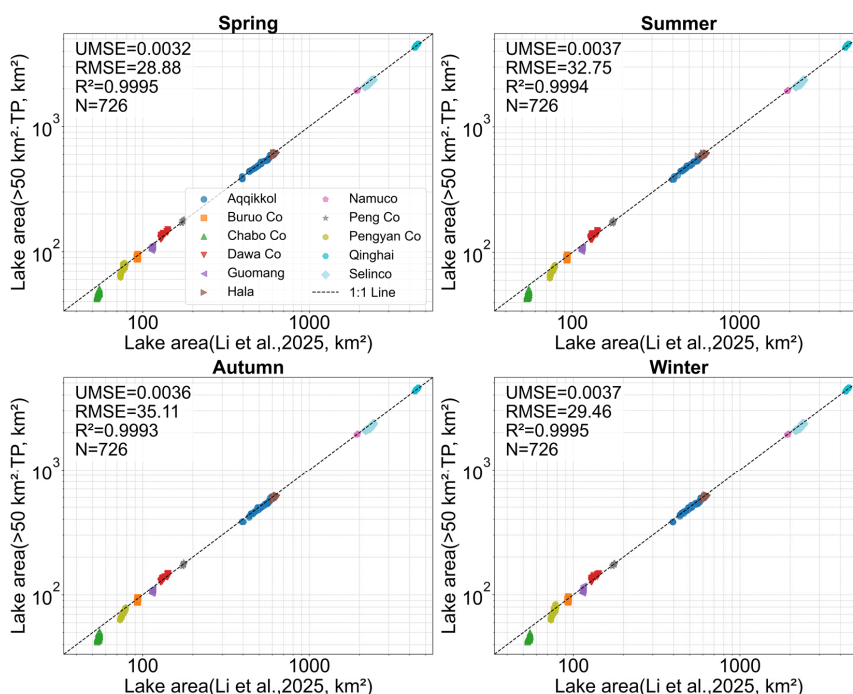


Figure 9: Scatter plot comparing different lake area extraction methods in spring, summer, autumn, and winter.

4.3.2 Inter-annual product comparison

Given the use of MOD09A1 and the available mapping period, we analysed annual and monthly lake area variations from 2000 to 2024 and compared them with four interannual products from the National Tibetan Plateau Data Centre. Using Qinghai Lake as a demonstration case, the lake exhibited a clear long-term expansion trend, with its area increasing from approximately 4300 km² in the early 2000s to more than 4600 km² after 2020 (Fig. 10a). The lake area typically reached its minimum in January and gradually expanded with increasing snowmelt and runoff, peaking in September before slightly declining in late autumn and early winter. This pattern was consistent with combined influences of seasonal precipitation, glacier and snowmelt, and evaporation, demonstrating the seasonal regulation of the lake's hydrological balance. Comparisons with the four reference datasets showed strong overall consistency. The comparison results indicated that the lake area data extracted in this study exhibit an R2 of 0.96, RMSE of 1019.31 km², MAPE of 2.2%, and a bias of -832.49 km² relative to Zhou et al. (2025) (Fig.



10b); an R^2 of 0.98, RMSE of 433.19 km^2 , MAPE of 0.9%, and a bias of 15.15 km^2 relative to Wang et al. (2023) (Fig. 10c); an R^2 of 0.99, RMSE of 366.42 km^2 , MAPE of 0.7%, and a bias of -178.00 km^2 relative to Zhang et al. (2022) (Fig. 10d); and an R^2 of 0.89, RMSE of 1359.51 km^2 , MAPE of 2.7%, and a bias of -998.32 km^2 relative to Zhang et al. (2019) (Fig. 10e). Scatter points aligned closely along the 1:1 line, indicating high agreement with existing interannual datasets. Agreement was particularly high with Zhang et al. (2022) and Wang et al. (2023). Our series also resolved intra-annual variations, reflecting seasonal expansion and contraction. Together, these comparisons suggest that the MOD09A1-based method captures both interannual and intra-annual lake area dynamics across the Plateau and offers a sound basis for long-term lake change analyses.

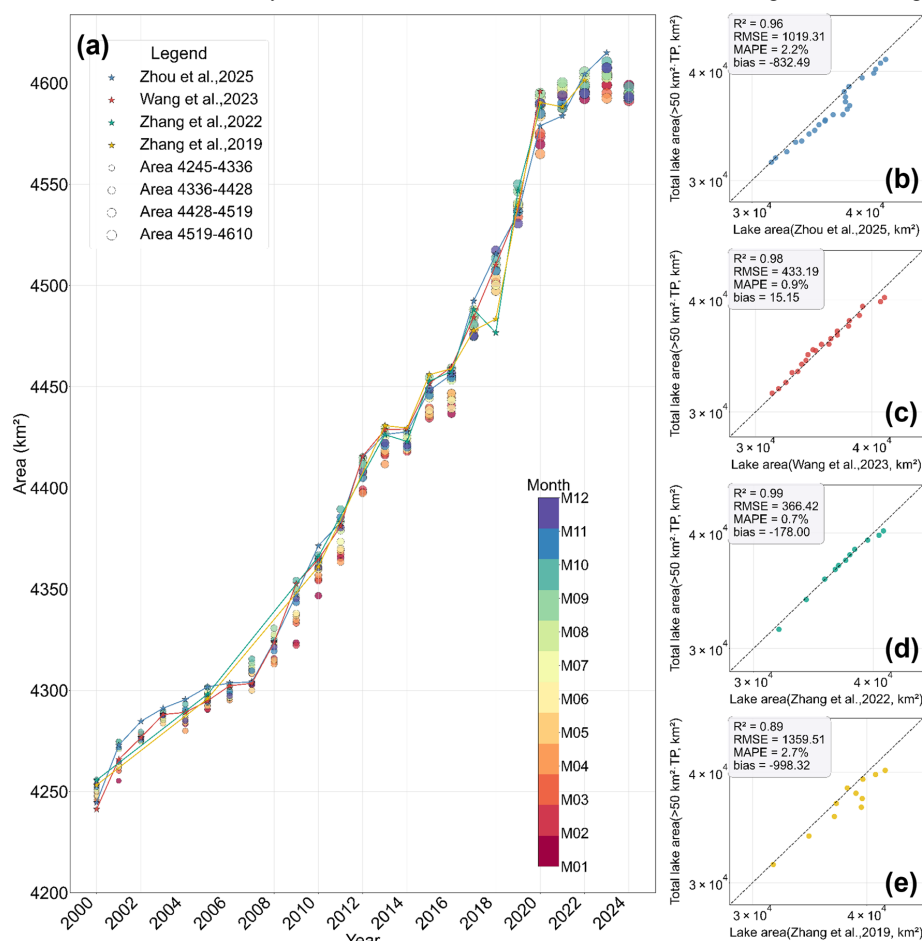


Figure 10: Temporal variations of Qinghai Lake area from 2000 to 2024 and consistency comparison with existing interannual datasets. (a) Annual variations in Qinghai Lake area derived from this study (2000–2024). Monthly lake areas are shown as coloured scatter points (M01–M12) using a spectral gradient, and the annual results from four existing datasets are marked with star symbols. Tibetan Plateau. (b–e) The comparison of total lake area (>50 km^2) from this study with four reference datasets.



5 Data availability

The monthly scale vector boundary dataset of lakes larger than 10 km² on the Tibetan Plateau from 2000 to 2024 (420 TPLake-MED) constructed in this study has been publicly shared on the National Tibetan Plateau Scientific Data Centre, accessible at: <https://doi.org/10.12443/BNU.RSEC.TPLake-MED20251028> (Zhao et al., 2025). The dataset includes monthly lake boundary vector files, area statistics tables, and related metadata, supporting Shapefile and GeoJSON formats, and can provide standardised data support for related research.

6 Code availability

425 The codes are available from the first and corresponding authors upon request.

7 Conclusions

Tibetan Plateau lakes are changing quickly, but most existing datasets only describe year-to-year trends and cannot capture seasonal behaviour. In this study, we utilised Google Earth Engine (GEE) and multi-temporal MODIS data to construct a monthly scale vector boundary dataset for lakes larger than 10 km² across the Tibetan Plateau for 2000–2024. Using a random (430 forest model and morphological optimisation, we improved water classification accuracy (overall accuracy 93.21%, F1 score 0.927) and boundary precision. The dataset reveals that total lake area has increased steadily (~34.91 km² per year); lake area typically reaches its annual maximum in September or October; the maximum monthly relative change rate for an individual lake can reach 28.43%; western lakes show stronger monthly growth and larger intra-annual fluctuations than eastern lakes, and the central plateau exhibits strongest variability; and smaller lakes are more sensitive to environmental change than larger (435 lakes. Compared with existing interannual products, our monthly scale data provide higher temporal resolution, filling a key gap and offering baseline information for studies of water-ice interactions and climate-change-related hydrological responses in high-altitude cold regions.

Author contributions

SZ and XZ: conceptualization; SZ: methodology and software; SZ: writing–original draft; SZ: writing–review and editing; XZ: (440 funding acquisition and project administration; XZ and JZ: writing–review and editing, data curation, and supervision; SZ, XL, and CY: data curation.

Competing interests

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



Disclaimer

445 Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements

We would like to express our sincere gratitude to all those who provided data and suggestions for this study.

450 Financial support

This research has been supported by the National Natural Science Foundation of China (grant nos. 42330205).

References

- Belgiu, M. and Drăguț, L.: Random forest in remote sensing: A review of applications and future directions, *ISPRS J. Photogramm. Remote Sens.*, 114, 24–31, <https://doi.org/10.1016/j.isprsjprs.2016.01.011>, 2016.
- 455 Bibi, S., Wang, L., Li, X. P., Zhou, J., Chen, D. L., and Yao, T. D.: Climatic and associated cryospheric, biospheric, and hydrological changes on the Tibetan Plateau: A review, *Int. J. Climatol.*, 38, e1–e17, <https://doi.org/10.1002/joc.5411>, 2018.
- Breiman, L.: Random forest, *Mach. Learn.*, 45, 5–32, <https://doi.org/10.1023/A:1010933404324>, 2001.
- Chen, J., Liu, R., Yang, X., Li, J., and Zhang, Y.: Water body information extraction combining improved Otsu algorithm and morphology, *Remote Sens. Inf.*, 37, 101–109, <https://doi.org/10.3969/j.issn.1000-3177.2022.01.015>, 2022.
- 460 Chen, P. and Xu, X.: Study on influencing factors of tensile strength of strip steel based on CART algorithm, *Control Eng. China*, 22, 276–281, <https://doi.org/10.14107/j.cnki.kzgc.131910>, 2015.
- Chen, W., Liu, Y., Zhang, G., Yang, K., Zhou, T., Wang, J., and Shum, C. K.: What controls lake contraction and then expansion in Tibetan Plateau's endorheic basin over the past half century?, *Geophys. Res. Lett.*, 49, e2022GL101200, <https://doi.org/10.1029/2022GL101200>, 2022.
- 465 Chi, Y., Lai, R., Yan, Q., and Li, M.: Detection and extraction of mountain shadow information based on Landsat 8 OLI data, *J. Mt. Sci.*, 35, 580–589, <https://doi.org/10.16089/j.cnki.1008-2786.000256>, 2017.
- Congalton, R. G.: A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sens. Environ.*, 37, 35–46, 1991.
- Deng, H., Pepin, N. C., Liu, Q., and Chen, Y.: Understanding the spatial differences in terrestrial water storage variations in the
- 470 Tibetan Plateau from 2002 to 2016, *Climatic Change*, 151, 379–393, <https://doi.org/10.1007/s10584-018-2325-9>, 2018.



- Ding, W., Bee, B. H., Spider, R. A., and Xing, Y.-W.: Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora, *Science*, 369, 579–581, <https://doi.org/10.1126/science.a004444>, 2020.
- Fawcett, T.: An introduction to ROC analysis, *Pattern Recogn. Lett.*, 27, 861–874, <https://doi.org/10.1016/j.patrec.2005.10.010>, 2006.
- 475 Feyisa, G. L., Fensholt, R., and Proud, S. R.: Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery, *Remote Sens. Environ.*, 140, 23–35, <https://doi.org/10.1016/j.rse.2013.08.029>, 2014.
- Fu, J., Song, L., Lei, L., Li, Y., and Zhang, Y.: Remote sensing monitoring of Panjin coastal zone based on optimal scale and random forest algorithm, *Remote Sens. Inf.*, 37, 7–15, <https://doi.org/10.3969/j.issn.1000-3177.2022.02.002>, 2022.
- Gu, C., Zhang, Y., Liu, L., Wei, B., Cui, B., and Gong, D.: Evaluation of consistency among four NDVI datasets applied to
480 Three River Source Region, Qinghai province, China, *Geogr. Res.*, 42, 1378–1392, <https://doi.org/10.11821/dlyj020220703>, 2023.
- Han, W., Huang, C., Wang, Y., and Gu, J.: Study on the area variation of Qinghai Lake based on long-term Landsat 5/8 multi-band remote sensing imagery, *Adv. Earth Sci.*, 34, 346–355, <https://doi.org/10.11867/j.issn.1001-8166.2019.04.0346>, 2019.
- 485 Huang, D., Zhang, J., Liu, J., Hu, X., and Jin, J.: Hydrologic regime dynamics and driving factors of Anglaren Co on the Tibetan Plateau based on multi-source remote sensing data, *J. Hydrol.*, 661, 133645, <https://doi.org/10.1016/j.jhydrol.2025.133645>, 2025.
- Huang, T., Liang, D., Jia, L., Li, X., and Wang, X.: Automatic extraction of lake areas on the Tibetan Plateau using interpolation-iteration method, *Remote Sens. Technol. Appl.*, 32, 289–298,
490 <https://doi.org/10.11873/j.issn.1004-0323.2017.2.0289>, 2017.
- Huang, Y., Deng, K., Ren, C., Yu, Z., and Pan, Y.: New water index and its stability study, *Prog. Geophys.*, 35, 829–835, <https://doi.org/10.6038/pg2020DD0311>, 2020.
- Hui, Y., Hu, H., Xiang, J., and Du, X.: Comparative analysis of machine learning algorithms for predicting tibial intramedullary nail length from patient characteristics, *J. Orthop. Traumatol.*, 26, 56,
495 <https://doi.org/10.1186/s10195-025-00869-4>, 2025.
- Immerzeel, W. W., van Beek, L. P. H., and Bierkens, M. F. P.: Climate change will affect the Asian water towers, *Science*, 328, 1382–1385, <https://doi.org/10.1126/science.1183188>, 2010.
- Khandelwal, A., Karpatne, A., Ravirathinam, P., Ghosh, R., Wei, Z., Dugan, H. A., Hanson, P. C., and Kumar, V.: RealSAT, a global dataset of reservoir and lake surface area variations, *Sci. Data*, 9, 356, <https://doi.org/10.1038/s41597-022-01449-5>,
500 2022.
- Kibret, S., McCartney, M., Lautze, J., Nhamo, L., and Yan, G.: The impact of large and small dams on malaria transmission in four basins in Africa, *Sci. Rep.*, 11, 13355, <https://doi.org/10.1038/s41598-021-92924-3>, 2021.



- Kuter, S.: Completing the machine learning saga in fractional snow cover estimation from MODIS Terra reflectance data: Random forests versus support vector regression, *Remote Sens. Environ.*, 255, 112294, 505 <https://doi.org/10.1016/j.rse.2021.112294>, 2021.
- Lei, D., Zhang, P., Wang, X., and Wang, C.: Study on water index construction based on Landsat 8 data, *Adv. Appl. Math.*, 11, 1–9, <https://doi.org/10.12677/AMI.2022.113127>, 2022.
- Li, G., Li, Z., Zhang, B., and Li, Z.: Changes in runoff from major alpine watersheds on the Qinghai-Tibetan Plateau: A review, *J. Hydrol.-Reg. Stud.*, 60, 102514, <https://doi.org/10.1016/j.ejrh.2025.102514>, 2025.
- 510 Li, L., Long, D., Wang, Y., and Woolway, R. I.: Global dominance of seasonality in shaping lake-surface-extent dynamics, *Nature*, 642, 361–368, <https://doi.org/10.1038/s41586-025-09046-3>, 2025.
- Li, M., Meng, B., Yan, D., Bi, W., and Wang, H.: Variation trends and attribution analysis of lakes in the Qiangtang Plateau, the Endorheic Basin of the Tibetan Plateau, *Sci. Total Environ.*, 837, 155595, <https://doi.org/10.1016/j.scitotenv.2022.155595>, 2022.
- 515 Li, X., Yao, X., Sun, M., Gong, P., An, L., Qi, M., and Gao, Y.: Spatial-temporal variations in lakes in northwest China from 2000 to 2014, *Acta Ecol. Sin.*, 38, 96–104, <https://doi.org/10.5846/stxb201612262677>, 2018.
- Li, X., Zhang, D., Jiang, C., Zhao, Y., Li, H., Lu, D., Qin, K., Chen, D., Liu, Y., Sun, Y., and Liu, S.: Comparison of lake area extraction algorithms in Tibetan Plateau leveraging Google Earth Engine and Landsat-9 data, *Remote Sens.*, 14, 4612, <https://doi.org/10.3390/rs14184612>, 2022.
- 520 Li, Y., Ma, J., and Chen, X.: Selection of methods for evaluating consistency of test results, *Chin. J. Clin. Lab. Sci.*, 33, 622–625, <https://doi.org/10.13602/j.cnki.jcls.2015.08.17>, 2015.
- Li, Y., Zhang, Q., Zhou, X., and Wang, L.: Watershed image segmentation algorithm based on morphology and region merging, *Comput. Eng. Appl.*, 56, 190–195, <https://doi.org/10.3778/j.issn.1002-8331.1903-0348>, 2020.
- Liu, J., He, G., Peng, L., Wang, J., and Yin, X.: A dataset of land surface water with a spatial resolution of 30 meters on the 525 Tibetan Plateau in 2022, *China Sci. Data*, 8, 179–187, <https://doi.org/10.57760/sciencedb.06937>, 2023.
- Liu, M., Lang, R., and Cao, Y.: The number of trees in random forests, *Comput. Eng. Appl.*, 51, 126–131, <https://doi.org/10.3778/j.issn.1002-8331.1401-0264>, 2015.
- Liu, Z., Chao, N., Chen, G., Zhang, G., Wang, Z., Li, F., and Ouyang, G.: Changes in monthly surface area, water level, and storage of 194 lakes and reservoirs in the Yangtze River Basin during 1990–2021 using multisource remote sensing data, *Sci. Total Environ.*, 944, 173840, <https://doi.org/10.1016/j.scitotenv.2024.173840>, 2024.
- 530 Long, D., Li, X., Li, X., Han, P., Zhao, F., Hong, Z., Wang, Y., and Tian, F.: Remote sensing retrieval of water storage changes and underlying climatic mechanisms over the Tibetan Plateau during 2000–2020, *Adv. Water Sci.*, 33, 375–389, <https://doi.org/10.14042/j.cnki.32.1309.2022.03.003>, 2022.



- Luo, K. L. and Yang, Z. H.: Traditional Tibetan nomadic practices and the security of the Three-River-Source "China Water Tower", *J. Jishou Univ. (Soc. Sci. Ed.)*, 32, 37–42, <https://doi.org/10.3969/j.issn.1007-4074.2011.01.008>, 2011.
- Ma, S., Gan, F., Wu, H., and Yan, B.: ICESat-2 data-based monitoring of 2018–2021 variations in the water levels of lakes in the Qinghai–Tibet Plateau, *Remote Sens. Nat. Resour.*, 34, 164–172, <https://doi.org/10.6046/zrzyyg.2021329>, 2022.
- McFeeters, S. K.: The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features, *Int. J. Remote Sens.*, 17, 1425–1432, <https://doi.org/10.1080/01431169608948714>, 1996.
- Miao, Q., Liu, X., Shi, H., Wei, Z., Luo, Y., Wang, Y., Gonçalves, J. M., and Feng, W.: Lake-area shrinkage driven by the combined effects of climate change and human activities, *Ecol. Indic.*, 175, 113606, <https://doi.org/10.1016/j.ecolind.2025.113606>, 2025.
- Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A.: Good practices for estimating area and assessing accuracy of land change, *Remote Sens. Environ.*, 148, 42–57, <https://doi.org/10.1016/j.rse.2014.02.015>, 2014.
- Pang, S.: Interannual variation in the area and water volume of lakes in different regions of the Tibet Plateau (1976–2019), National Tibetan Plateau / Third Pole Environ. Data Center [data set], <https://doi.org/10.11888/Hydro.tpd.271599>, 2021.
- Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A. S.: High-resolution mapping of global surface water and its long-term changes, *Nature*, 540, 418–422, <https://doi.org/10.1038/nature20584>, 2016.
- Qiao, B. J., Zhu, L. P., and Yang, R. M.: Temporal-spatial differences in lake water storage changes and their links to climate change throughout the Tibetan Plateau, *Remote Sens. Environ.*, 222, 232–243, <https://doi.org/10.1016/j.rse.2018.12.037>, 2019.
- Qiu, J.: China: The third pole, *Nature*, 454, 393–396, <https://doi.org/10.1038/454393a>, 2008.
- Ran, J., Qiu, J., and Hu, R.: Monthly area changes of lakes greater than 30 km² on the Tibetan Plateau over 2015–2020, National Tibetan Plateau / Third Pole Environment Data Center [data set], <https://doi.org/10.11888/Terre.tpd.300869>, 2023.
- Saha, A. and Pal, S. C.: Application of machine learning and emerging remote sensing techniques in hydrology: A state-of-the-art review and current research trends, *J. Hydrol.*, 632, 130907, <https://doi.org/10.1016/j.jhydrol.2024.130907>, 2024.
- Song, C., Zhan, P., and Ma, R.: Progress in remote sensing study on lake hydrologic regime, *J. Lake Sci.*, 32, 1406–1420, <https://doi.org/10.18307/2020.0514>, 2020.
- Stehman, S. V. and Foody, G. M.: Key issues in rigorous accuracy assessment of land cover products, *Remote Sens. Environ.*, 231, 111199, <https://doi.org/10.1016/j.rse.2019.05.018>, 2019.
- Tucker, C. J.: Red and photographic infrared linear combinations for monitoring vegetation, *Remote Sens. Environ.*, 8, 127–150, [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0), 1979.



- 565 Wan, W., Zhao, L., Xie, H., Liu, B., Li, H., Cui, L., and Ma, Y.: Monitoring lake changes of Qinghai–Tibetan Plateau over the past 30 years using satellite remote sensing data, *Chin. Sci. Bull.*, 59, 1021–1035, <https://doi.org/10.1007/s11434-014-0128-6>, 2014.
- Wang, G.: Study on the differences in seasonal variations of rivers and their impacts on lake changes in the northwestern Tibetan Plateau, Ph.D. thesis, Zhengzhou University, 2023.
- 570 Wang, M. Z.: Spatiotemporal variation of snow cover and its relationship with climatic factors over the Tibetan Plateau in recent 15 years, M.S. thesis, East China Normal University, 2016.
- Wang, S., Xu, T., Zhang, H., Li, J., and Zhao, Y.: Evolution patterns and driving factors of major lakes in Ningxia during 1980–2016, *J. Anhui Agric. Sci.*, 52, 57–64, <https://doi.org/10.3969/j.issn.0517-6611.2024.03.014>, 2024.
- Xu, F., Zhang, G., Woolway, R. I., Yang, K., Wada, Y., Wang, J., and Crétau, J.-F.: Widespread societal and ecological
 575 impacts from projected Tibetan Plateau lake expansion, *Nat. Geosci.*, 17, 516–523, <https://doi.org/10.1038/s41561-024-01446-w>, 2024.
- Xu, H.: Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery, *Int. J. Remote Sens.*, 27, 3025–3033, <https://doi.org/10.1080/01431160600589179>, 2006.
- Yan, F., Sillanpää, M., Kang, S., Aho, K. S., Qu, B., Wei, D., Li, X., Li, C., and Raymond, P. A.: Lakes on the Tibetan Plateau
 580 as conduits of greenhouse gases to the atmosphere, *J. Geophys. Res.-Biogeosci.*, 123, 2091–2103, <https://doi.org/10.1029/2017JG004379>, 2018.
- Yan, L., Zheng, X., and Qi, Y.: Surface area variations of lakes in the Tibetan Plateau and their influencing factors, *Sci. Technol. Rev.*, 35, 83–88, <https://doi.org/10.3981/j.issn.1000-7857.2017.06.010>, 2017.
- Yang, J., Jin, X., Jin, Y., Chen, K., Xie, H., Li, Z., and Fu, D.: Changes in wetland area and their influencing factors in alpine
 585 inland river basins, *Acta Ecol. Sin.*, 40, 3684–3699, <https://doi.org/10.20103/j.stxb.202409122208>, 2025.
- Yang, K.: Dynamic monitoring of lake area changes on the Tibetan Plateau using multi-source and multi-temporal satellite imagery, Ph.D. thesis, University of Chinese Academy of Sciences, 2017.
- Zhang, B. and Min, H.: Customer satisfaction prediction using grid search optimized random forest, *J. Beijing Inf. Sci. Technol. Univ. (Nat. Sci. Ed.)*, 36, 50–58, <https://doi.org/10.16508/j.cnki.11-5866/n.2021.04.010>, 2021.
- 590 Zhang, G.: Long-term sequence dataset of lake area on the Tibetan Plateau (1970–2013), National Tibetan Plateau / Third Pole Environment Data Center [data set], <https://doi.org/10.11888/Lake.tpe.249466.file>, 2018.
- Zhang, G.: The lakes larger than 1 km² in Tibetan Plateau (v3.1) (1970s–2022), National Tibetan Plateau / Third Pole Environ. Data Center [data set], <https://doi.org/10.1016/j.scib.2019.07.018>, 2019.
- Zhang, G. and Ran, Y.: Time series dataset of lake area on the Qinghai–Tibet Plateau for the past 100 years (1920–2020),
 595 National Tibetan Plateau / Third Pole Environ. Data Center [data set], <https://doi.org/10.11888/Terre.tpdc.272891>, 2022.



- Zhang, Q., Shen, Z., Pokhrel, Y., Farinotti, D., Singh, V. P., Xu, C.-Y., Wu, W., and Wang, G.: Oceanic climate changes threaten the sustainability of Asia's water tower, *Nature*, 615, 87–93, <https://doi.org/10.1038/s41586-022-05643-8>, 2023.
- Zhang, X., Zeraatpisheh, M., Rahman, M. M., Wang, S., and Xu, M.: Texture is important in improving the accuracy of mapping photovoltaic power plants: A case study of Ningxia Autonomous Region, China, *Remote Sens.*, 13, 3909, <https://doi.org/10.3390/rs13193909>, 2021.
- 600 Zhao, R., Fu, P., Zhou, Y., Xiao, X., Grebby, S., Zhang, G., and Dong, J.: Annual 30-m big lake maps of the Tibetan Plateau in 1991–2018, *Sci. Data*, 9, 164, <https://doi.org/10.1038/s41597-022-01270-5>, 2022.
- Zhao, S., Zhao, X., Zhao, J., Zhang, X., Liu, X., and Yao, C.: Monthly lake area changes larger than 10 km² on the Tibetan Plateau (2000–2024), Beijing Engineering Research Center for Global Land Surface Remote Sensing [data set], <https://doi.org/10.12443/BNU.RSEC.TPLake-MED20251028>, 2025.
- 605 Zhao, Z., Zhang, Y., Hu, Z., and Nie, X.: Contrasting evolution patterns of endorheic and exorheic lakes on the Central Tibetan Plateau and climate cause analysis during 1988–2017, *Water*, 13, 1962, <https://doi.org/10.3390/w13141962>, 2021.
- Zhou, Y., Liu, B., Cui, Y., and Dong, J.: Annual 30-m lake maps on the Tibetan Plateau (1991–2023) (V1.0), National Tibetan Plateau / Third Pole Environ. Data Center [data set], <https://doi.org/10.5281/zenodo.10686952>, 2025.
- 610 Zhu, L., Ju, J., Qiao, B., Liu, C., Wang, J., Yang, R., Ma, Q., Guo, L., and Pang, S.: Physical and biogeochemical responses of Tibetan Plateau lakes to climate change, *Nat. Rev. Earth Environ.*, 6, 284–298, <https://doi.org/10.1038/s43017-025-00650-5>, 2025.
- Zhu, L., Zhang, G., Yang, R., Wang, J., and Ju, J.: Lake variations on Tibetan Plateau of recent 40 years and future changing tendency, *Bull. Chin. Acad. Sci.*, 34, 1254–1263, <https://doi.org/10.16418/j.issn.1000-3045.2019.11.008>, 2019.