Earth Syst. Sci. Data, 18, 4219–4239, 2026
https://doi.org/10.5194/essd-18-4219-2026

Open Access Earth System Science Data

TPLake-MED: a monthly extent dataset for lakes on the Tibetan Plateau

Siyu Zhao[1,2], **Xiang Zhao**[1,2], **Jiacheng Zhao**[3], **Xin Zhang**[4], **Xingyu Liu**[1,2], **and Chengzhi Yao**[1,2]

[1]State Key Laboratory of Remote Sensing and Digital Earth, Faculty of Geographical Science,
Beijing Normal University, Beijing 100875, China
[2]Beijing Engineering Research Center for Global Land Remote Sensing Products,
Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
[3]School of Ecology and Applied Meteorology, Nanjing University of Information Science
and Technology, Nanjing, Jiangsu 210044, China
[4]Department of Computing and Mathematics, Manchester Metropolitan University, Manchester M15GD, UK

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

Received: 29 October 2025 – Discussion started: 13 November 2025
Revised: 4 June 2026 – Accepted: 5 June 2026 – Published: 22 June 2026

Abstract. Lakes on the Qinghai–Tibet Plateau have expanded markedly over recent decades, reflecting complex interactions between the regional water cycle and the cryosphere. Annual datasets capture long-term trends but often overlook short-term hydrological responses and seasonal transitions, which 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, making 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 TPLake-MED, a monthly vector boundary dataset for lakes larger than 10 km^2 across the Qinghai–Tibet Plateau for 2000–2024. Within this dataset 142 to 175, and the number of smaller lakes (10–50 km^2) ranges from 232 to 260 across the study period. A random forest classifier based on spectral indices was developed and validated using 633 balanced water/non-water sample pixels, achieving a precision of 0.986, recall of 0.875, and F_1-score of 0.927 for the water class. We also implemented a boundary optimisation workflow that integrates filtering, morphological operations, and geometric rectification to improve agreement between the extracted and actual lake extents. The total lake area on the plateau increased by 34.91 $\mathrm{km}^2\,\mathrm{yr}^{-1}$, 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 on the central plateau. 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. This is the first monthly resolution vector dataset of Qinghai–Tibet 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/10.5281/zenodo.19205532 (Zhao et al., 2026).

1 Introduction

Alpine and polar lakes are critical components of the global ecosystem that are currently undergoing rapid transformations driven by accelerating global warming. Across major mountain environments such as the Andes, the Alps, the Himalaya, and the Arctic, rising temperatures have accelerated glacier retreat and permafrost degradation. These cryospheric changes have fundamentally altered hydrological cycles and driven widespread shifts in the formation, expansion, and seasonal variability of glacial lakes (Huggel et al., 2020; Shugar et al., 2020; Zemp et al., 2019). Acting as sentinels of climate change, these lakes integrate signals from glacier melt and shifting water balances while simultaneously influencing downstream freshwater availability, ecosystem stability, and hazards like glacial lake outburst floods (Carrivick and Tweed, 2016; Emmer et al., 2022). Consequently, systematically characterising the long-term spatiotemporal dynamics of lake extent in these sensitive regions has become a central scientific objective for understanding the evolution of global lake systems and assessing broader ecological stability.

As the “Third Pole” and “Water Tower of Asia”, the unique geo-ecological status of the Qinghai–Tibet Plateau (QTP) has directed exceptional scientific attention toward the rapid transformations of its lake systems. Commonly known as the “Roof of the World,” this region stands as the world’s highest and largest plateau, characterised by distinctive topography, cryospheric conditions, and coupled land–atmosphere processes (Qiu, 2008; Bibi et al., 2018). As the headwater region of the Yangtze, the Yellow River, and several major transboundary rivers, it supports livelihoods and socio-economic development across South and South-East Asia (Immerzeel et al., 2010). Beyond its ecological significance and diverse alpine ecosystems, the plateau strongly influences the Asian monsoon and large-scale atmospheric circulation through its extreme elevation and snow-albedo feedbacks (Ding et al., 2020). Lakes are an integral component of the Qinghai–Tibet Plateau hydrological system, modulating regional water cycles and preserving climatic signals across multiple temporal scales (Li et al., 2018; Song et al., 2020). The plateau hosts the largest and most numerous cluster of high-elevation lakes globally (Zhao et al., 2022). Over the past few decades, extensive research has documented dramatic lake variations on the Qinghai–Tibet Plateau, identifying rising temperatures, increased precipitation, and accelerated glacier retreat as primary drivers of this widespread lake expansion (Liao et al., 2013; Li et al., 2014; Song et al., 2014; Yang et al., 2017; Zhang et al., 2021b). Simultaneously, variations in lake extent act as crucial indicators of regional hydrological dynamics and cryospheric change (Wan et al., 2014; Qiao et al., 2019). Since the late twentieth century, endorheic lakes have expanded more rapidly than exorheic lakes in line with an overall expansion trend (Zhao et al., 2021; Li et al., 2022a). This pronounced spatial heterogeneity introduces uncertainties in watershed-scale hydrological responses and directly affects regional water balance, ecosystem stability, and infrastructure safety (Deng et al., 2018; Chen et al., 2022; Xu et al., 2024; 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, providing robust solutions for the challenging alpine environments of the Qinghai–Tibet Plateau (Saha and Pal, 2024). Traditional water body index-based approaches, which rely on mathematical combinations of specific spectral bands to enhance water signals while suppressing background noise, have found application in large-scale dynamic monitoring (Pekel et al., 2016; Yan et al., 2018; Han et al., 2019; Miao et al., 2025). Common indices such as the Normalized Difference Water Index (NDWI; McFeeters, 1996), the Modified Normalized Difference Water Index (MNDWI; Xu, 2006), and the Automated Water Extraction Index (AWEI; Feyisa et al., 2014) 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 (Yan et al., 2020). In contrast, water body classification methods integrating machine learning (e.g., random forest, support vector machine) demonstrate favourable adaptability and generalisation capabilities by effectively weighing various spectral signatures to improve the accuracy of lake boundary identification in complex environments. To further mitigate shadow artefacts, subsequent studies have integrated these models with ancillary terrain data (Li et al., 2022b; Liu et al., 2023; Yang et al., 2025). For instance, Li et al. (2022b) integrated Landsat-8/9 multispectral data with DEM terrain features using a random forest algorithm to extract water bodies on the Qinghai–Tibet Plateau, achieving an overall accuracy of 95.84 %, significantly outperforming traditional methods like NDWI. Liu et al. (2023) integrated Landsat multispectral data, DEM terrain features, and textural features using a random forest algorithm to generate a water body distribution map for the Qinghai–Tibet Plateau, achieving an overall accuracy of 92.9 %, which is markedly superior to single-index methods. Yang et al. (2025) 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. However, while incorporating DEMs is effective for medium-scale imagery, relying on terrain features at a coarse spatial resolution (e.g., 500 m) introduces severe scale-dependent shoreline inconsistency and computational inefficiencies, and remains fundamentally ineffective against dynamic cloud shadows, thereby necessitating robust spatiotemporal post-processing to effectively eliminate these residual artefacts and classification noise.

While several vector datasets of lake area on the Qinghai–Tibet Plateau are currently available and widely utilised, they exhibit limitations in spatiotemporal continuity and coverage (Zhang et al., 2017, 2019, 2020; Pang et al., 2021; Ran et al., 2024; Wang et al., 2024). For instance, Zhang et al. (2017) constructed triennial area time series for 364 lakes larger than 10 km^2 between 1970 and 2013; Pang et al. (2021) generated continuous area time series for 20 lakes larger than 100 km^2 from 1976 to 2019; Wang et al. (2024) provided annual mean area data for 180 lakes from 1986 to 2020; Zhang et al. (2019) acquired lake area data for 16 periods from 1970 to 2022; Zhang et al. (2020) extracted monthly scale area for lakes larger than 50 km^2 from 2015 to 2018 based on Sentinel-1 SAR data; Ran et al. (2024) generated a dataset of monthly lake area changes for lakes larger than 30 km^2 from 2015 to 2020 using multi-source remote sensing imagery. However, most interannual products typically select only a single phase (e.g., a single summer scene per year) to represent annual lake extent, thereby 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. 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., 2025b). Furthermore, maintaining spatiotemporal consistency simultaneously is highly challenging; long-term high-frequency datasets struggle to robustly cover lakes below the 50 km^2 threshold over multi-decadal scales due to data gaps and methodological inconsistencies (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 conducted long-term, high-frequency, comprehensive monitoring of monthly lake area change across the Qinghai–Tibet 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 modelling of lake water volume dynamics, analysis of driving mechanisms, and research into responses to climate change (Khandelwal et al., 2022; Liu et al., 2024). Consequently, this study aims to: (1) develop an automated extraction method for Qinghai–Tibet 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 (TPLake-MED) for lakes exceeding 10 km^2 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 the spatiotemporal patterns of lake area change across the Qinghai–Tibet 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 Qinghai–Tibet Plateau (26°00′–39°47′ N, 73°19′–104°47′ E) exceeds 4000 m in average elevation and is the highest and largest plateau in the world (Long et al., 2022; Zhang et al., 2023). The Qinghai–Tibet Plateau is rich in lake resources, with over 1500 lakes of various sizes, including approximately 180 lakes larger than 50 km^2. The total lake area is approximately 44 993 km^2, accounting for about 50 % of China's total lake area (Huang et al., 2025) (Fig. 1). 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 the Qinghai–Tibet Plateau is influenced by both the altitude and geographical features. The larger lakes, those over 50 km^2, 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 km^2) are scattered more widely across areas with more complex terrain, including the southern and southeastern regions of the plateau. These areas are characterised by rugged mountain systems, such as the Himalayas and Hengduan Mountains, which shape the hydrological landscape and influence lake distribution.

3 Methods

The monthly lake area data generation workflow comprises four main stages (Fig. 2): First, the monthly baseline imagery is constructed by selecting the MOD09A1 (500 m) image with the lowest cloud cover for each month from the GEE platform, followed by cloud masking and temporal gap-filling interpolation. Second, water bodies are classified using a random forest classifier with multispectral bands and spectral indices, and accuracy is evaluated using confusion matrices and multi-algorithm comparisons. Third, post-classification processing includes noise reduction, missing pixel detection and filling, overlaying the Joint Research Centre (JRC) global surface water mask (Pekel et al., 2016; https://global-surface-water.appspot.com/, last access: 7 June 2026), morphological smoothing, temporal consistency checks to filter ephemeral noise, boundary intersection over union (IoU; Jaccard, 1912) constraints, and invalid geometry repair to optimise boundaries. Finally, we validated accuracy against multi-source data and existing products, and conducted spatial heterogeneity and time-series analyses to reveal spatiotemporal patterns in lake area.

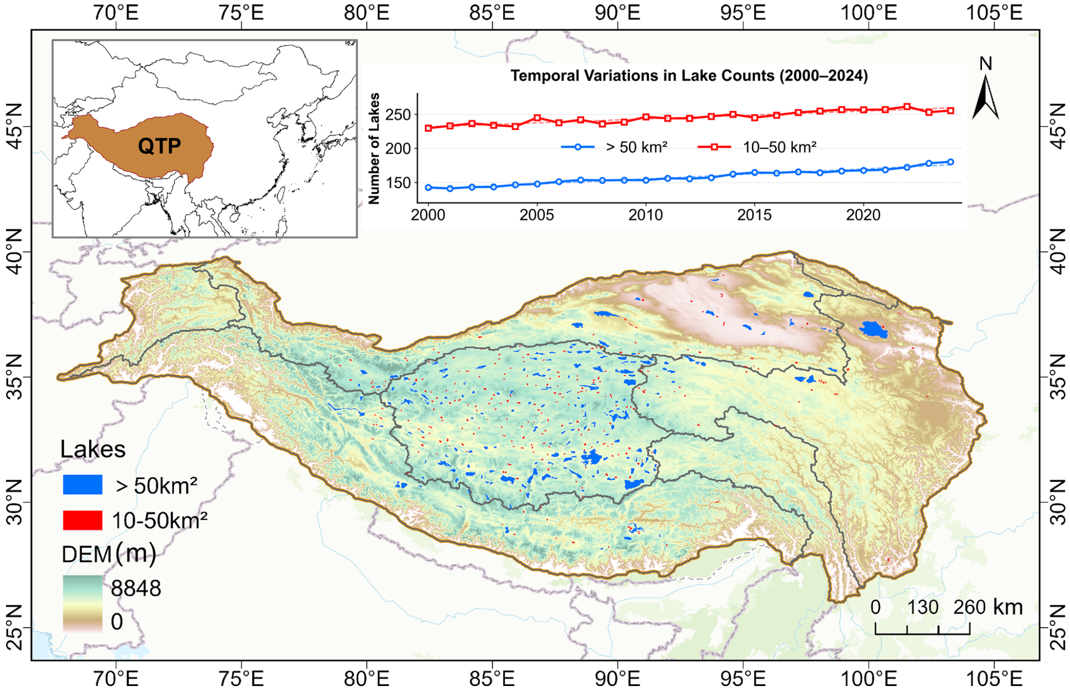

Figure 1. Overview map of the study area. The main map illustrates the spatial distribution of lakes categorised by size, overlaid on the topographic relief. Blue contours indicate large lakes ($> 50\,\mathrm{km}^2$), while red contours denote medium-sized lakes (10–$50\,\mathrm{km}^2$). The inset time-series plot displays the corresponding annual fluctuations in the total number of lakes for these two size categories from 2000 to 2024. QTP denotes the Qinghai–Tibet Plateau.

3.1 Data preprocessing

We utilised the MODIS MOD09A1 surface reflectance (SR) product with a spatial resolution of 500 m. The MOD09A1 product provides atmospherically 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 operations to detect the 10th cloud flag) to remove residual cloud and cloud shadow-affected pixels. To address remaining data gaps 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):

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

where R_t is the reflectance of the pixel to be filled at target month t, and $t_{\pm 1}$ are the effective time points adjacent to the missing value. Missing data in the first and last months were filled with the nearest pixel values to maintain the spatiotemporal continuity in the 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 the 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 the reflectance utilised for classification primarily originated 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 Qinghai–Tibet 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 (Kuter, 2021). CART recursively splits data based

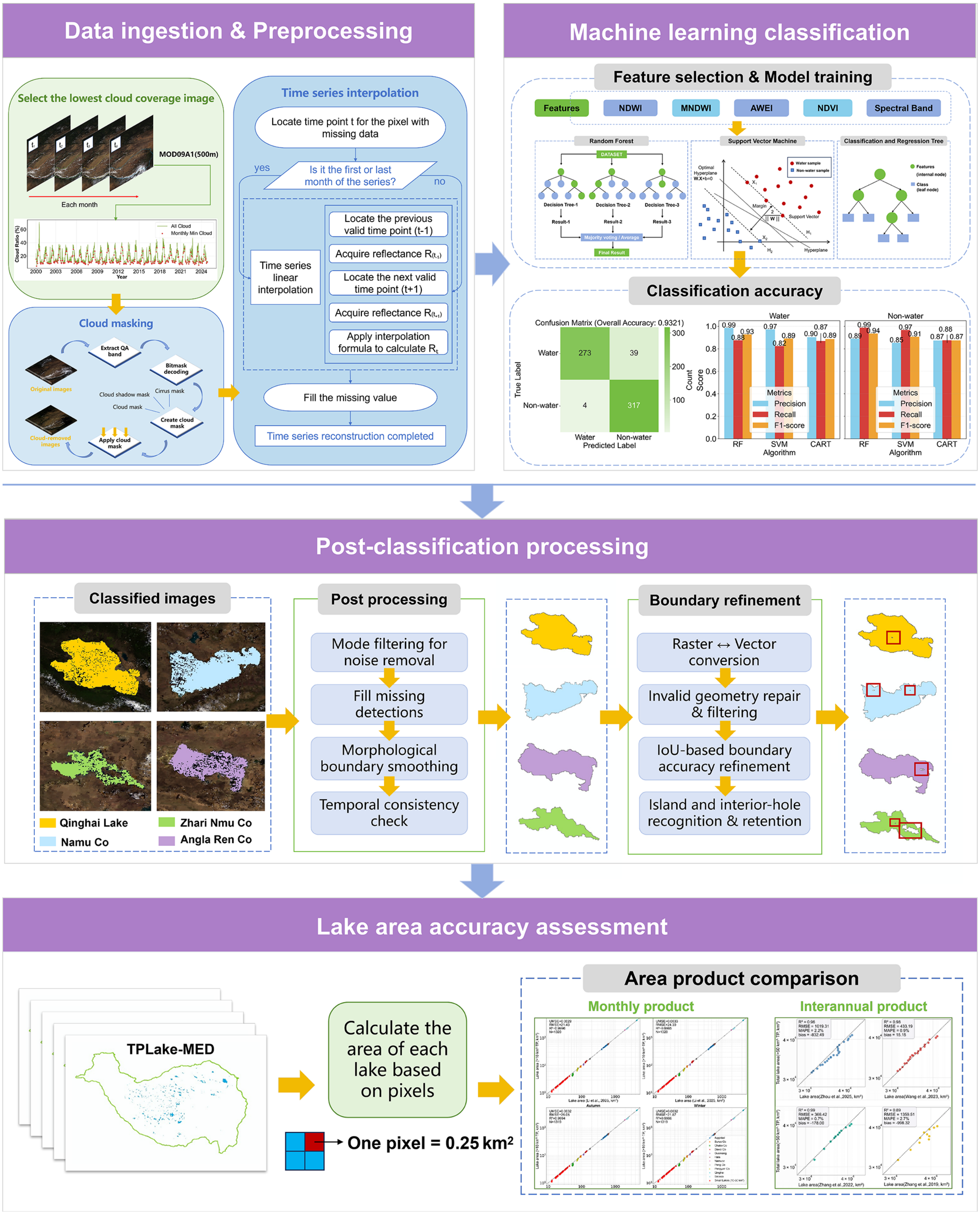

Figure 2. Framework of monthly lake data generation in this study. The process comprises four steps: data preprocessing to select monthly imagery with the lowest cloud cover (Step 1), random forest classification to obtain preliminary water body extraction results (Step 2), post-classification lake morphological and boundary optimisation (Step 3), lake area calculation, comparative validation and spatiotemporal analysis (Step 4). The image examples are based on MODIS/Terra MOD09A1 Collection 6.1 8 d surface reflectance imagery (500 m). Data source: NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC).

on the Gini index without any distributional assumptions, effectively handling non-linear relationships and evaluating the importance of multi-source features such as spectral and index-based variables (Loh, 2011).

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., 2021c). These variables were utilised to discriminate between surface water and non-water bodies on the Qinghai–Tibet Plateau. The four indices included the Normalized Difference Vegetation Index (NDVI; Tucker, 1979), the Normalized Difference Water Index (NDWI; McFeeters, 1996), the Modified Normalized Difference Water Index (MNDWI; Xu, 2006), and the Automated Water Extraction Index (AWEI; Feyisa et al., 2014), with the formulas as follows:

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

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

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

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

where ρ_{GREEN}, ρ_{RED}, ρ_{NIR}, 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 optimisation

Classifiers are highly sensitive to sampling design (Belgiu and Drăguţ, 2016). Appropriate training samples are critical for the classification accuracy and stability of models. In the absence of suitable labelled sample data, we manually labelled 2420 sample points on the GEE platform as the training dataset, including 1275 points labelled as water bodies and 1145 points labelled as non-water bodies (Fig. 3a), ensuring that the sample points were uniformly distributed across the study area. Additionally, we utilised the JRC dataset as an auxiliary reference for classification. Finally, we randomly divided 70 % of the samples into a training set and the remaining 30 % into a validation set. After excluding ambiguous or invalid samples, 633 validation samples were retained for accuracy assessment.

For the RF classifier, hyperparameters were tuned by grid search with cross-validation to ensure generalisation capability (Zhang and Hu, 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. These settings increased ensemble diversity and mitigated overfitting. For the comparative baseline models, we also applied analogous parameter optimisation strategies. The SVM classifier was implemented with a Radial Basis Function (RBF) kernel to effectively handle non-linear decision boundaries. Through empirical tuning using grid search, the cost parameter (Cost), which penalises misclassification, was set to 10, and the kernel coefficient (Gamma) to 0.5. For the CART benchmark, we utilised the Gini impurity index for node splitting. To prevent the excessive overfitting typical of single trees, we applied structural regularisation by capping the maximum number of nodes (maxNodes) at 128 and setting the minimum leaf population (minLeafPopulation) to 5. These constraints were introduced to suppress high-frequency noise and ensure robust classification rules (Table 1).

Table 1. Hyperparameter settings for the machine learning classifiers.

Model	Hyperparameter	Value
Random Forest	Number of Trees	150
	Min Leaf Population	5
	Variables per Split	$\sqrt{N}$
SVM	Kernel Type	RBF
	Cost (C)	10
	Gamma (γ)	0.5
CART	Split Criterion	Gini Index
	Max Nodes	128
	Min Leaf Population	5

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, recall, and F_1-score from the confusion matrix. We then compared the three models under identical experimental conditions (Fawcett, 2006).

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, mountain shadows and dynamic cloud shadows led to false water bodies and slope-related misclassifications; second, spectral confusion caused salt-and-pepper noise, significantly reducing the geographical plausibility of boundaries (Wang et al., 2023; Li et al., 2025a). To overcome these limitations and meet the accuracy requirements for large-

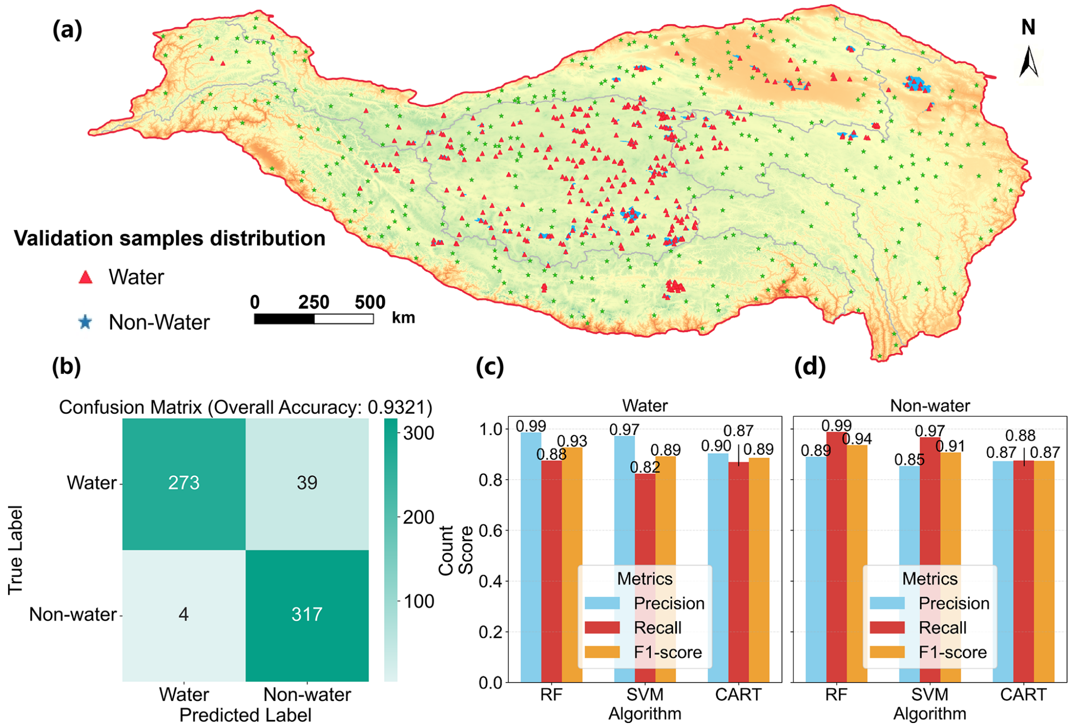

Figure 3. Validation sample distribution and classification accuracy of the Random Forest model for lake water and non-water body identification. **(a)** The distribution of validation sample points for the classification models. **(b)** The confusion matrix of the Random Forest classification model. **(c, d)** The bar chart compares the main evaluation metrics (Precision, Recall, and F_1-score) of different classification algorithms for the two categories of "water body" and "non-water body."

scale lake area extraction, we constructed a comprehensive automated post-processing workflow for lake classification, encompassing core modules such as multi-file batch processing, dynamic coordinate system unification, geometric topology repair, and fine-grained retention of internal islands. 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 (Li et al., 2020), while a temporal consistency check eliminated dynamic cloud shadows by identifying discontinuous single-month water patches as transient noise. An accompanying island screening step further improved the extraction of water features.

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. However, these preliminary results inevitably contained misclassifications induced by mountain shadows, dynamic cloud shadows, and spectral confusion, manifesting as false water patches and isolated salt-and-pepper noise. To mitigate these issues and ensure boundary reliability without relying on explicit terrain features, we implemented a robust spatial filtering strategy. We first filtered out classified water-body pixels with fewer than nine neighbouring pixels. Additionally, to leverage high-confidence prior water information and effectively restrict out-of-basin topographic shadow anomalies, 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 promoted more complete inclusion of water edges, especially near wetlands or unclassified areas, whereas erosion removed small artefacts introduced by dilation, yielding smoother boundaries. To address the remaining dynamic cloud shadows, we implemented a temporal consistency check across the monthly time series. Recognising that true lake expansion is a continuous process, we identified and removed isolated water patches that appeared in only a single month, treating them as transient cloud-induced noise.

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 vectorised water body boundaries generated through morphological optimisation, automated operations were implemented using a multi-file batch-processing framework. During data loading, dynamic coordinate system unification ensured that all input files were spatially referenced to the reference layer (Zhou et al., 2025), thereby 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 (IoU) ratio 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

To comprehensively evaluate the reliability of the TPLake-MED dataset, we designed a multi-tiered verification framework encompassing spatial accuracy evaluation using independent high-resolution ground truth, followed by temporal consistency and cross-product comparisons at both regional and global scales.

To evaluate the classification method across different spatial scales, particularly for small lakes that are highly sensitive to spatial resolution limitations and mixed pixels, we conducted a stratified accuracy assessment using independent high-resolution ground truth. We selected a representative sample of 22 lakes stratified into four size categories: 10–50, 50–100, 100–500, and $> 500\,\mathrm{km}^2$. For each lake, reference masks were generated from 10 m Sentinel-2 imagery acquired in October 2020. Subsequently, dense stratified random sampling at a 30 m spatial resolution was performed to evaluate sub-pixel errors along the 500 m MODIS boundaries. To prevent statistical distortion from varying sample sizes across lakes, we aggregated the absolute counts of True Positives (TP), False Positives (FP), and False Negatives (FN) across all sample points within each specific size tier to calculate micro-averaged metrics.

Beyond spatial accuracy, we assessed the temporal consistency of our dataset. To evaluate intra-annual dynamics and seasonal performance, we validated our lake-area extractions against the dataset from Li et al. (2025c). This correlation analysis selected 20 representative lakes, comprising 10 small, 7 medium, and 3 large lakes, yielding a total of 1319 monthly sample points from 2001 to 2023 (Li et al., 2025c). The consistency of these seasonal dynamics was quantified using root-mean-square error (RMSE), unbiased RMSE (URMSE), and the coefficient of determination (R^2).

Furthermore, we evaluated the long-term interannual stability of our dataset. Specifically, the extracted monthly-average lake areas were aggregated to the annual scale and benchmarked against four established interannual lake datasets, as presented in Table 2. Using the datasets from Zhang (2019), Zhang and Ran (2022), Wang and Jin (2023), and Zhou et al. (2025) as comparative baselines, we constructed scatter plots to analyse the long-term consistency between our observations and existing records (Zhang et al., 2014, 2017, 2021a; Wang et al., 2024; Zhou et al., 2024). This quantitative assessment employed R^2, RMSE, mean absolute percentage error (MAPE), and bias (Hui et al., 2025) to validate data quality and method effectiveness.

To provide a broader global outlook and evaluate TPLake-MED within the context of established international frameworks, we introduced the JRC Global Surface Water product and the HydroLAKES dataset (Messager et al., 2016; https://www.hydrosheds.org/products/hydrolakes, last access: 7 June 2026) as additional baselines. By comparing our dataset with these widely utilised global products, we aimed to comprehensively assess spatial completeness and temporal continuity under complex local atmospheric conditions.

3.5 Uncertainty analysis

To address the uncertainty induced by the 500 m spatial resolution of the MODIS dataset, we assessed boundary-shift uncertainties and evaluated the signal-to-noise ratio (SNR). These analyses were stratified by lake size (10–50, 50–100, 100–500, and $> 500\,\mathrm{km}^2$) and utilised the entire 25-year dataset (2000–2024). We simulated a “worst-case scenario” by assuming a continuous one-pixel (500 m) and half-pixel (250 m) shift along the extracted boundaries for all lakes. To quantify this impact, the relative area error was estimated geometrically: we multiplied the extracted lake mean perimeter by the assumed boundary shift distance (250 or 500 m) to estimate the maximum potential area of misclassification. This theoretical error value was then divided by the originally extracted lake area to determine the relative area error percentage. To determine whether the observed long-term lake area variations are genuine physical signals or merely artefacts of spatial resolution noise, we calculated the SNR for each lake over the 25-year period. SNR was defined as the ratio of historical absolute area fluctuation (Signal) to the one-pixel boundary uncertainty (Noise), calculated as follows:

$$\mathrm{SNR} = \frac{\mathrm{Area}_{\mathrm{max}} - \mathrm{Area}_{\mathrm{min}}}{\mathrm{Perimeter}_{\mathrm{mean}} \times 0.5} \tag{6}$$

Table 2. Information on the reference lake area products utilised for comparison.

Dataset description	Image used	Number of lakes	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 and Jin (2023)
Time series dataset of lake area on Tibetan Plateau for the past 100 years	Landsat	1236	1920–2020	Zhang and Ran (2022)
The lakes larger than 1 km^2 in Tibetan Plateau (v3.1)	Landsat	1400	1970–2022	Zhang (2019)

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

where $\mathrm{Area}_{\mathrm{max}}$ and $\mathrm{Area}_{\mathrm{min}}$ are the maximum and minimum lake areas (km^2), $\mathrm{Perimeter}_{\mathrm{mean}}$ is the mean perimeter of the lake across all valid observations (km), and 0.5 represents the 500 m spatial resolution to express the 500 m spatial resolution in kilometres. The spatial uncertainty of the extracted water area is inherently proportional to its perimeter (Wang et al., 2012; Reza et al., 2023). Furthermore, to eliminate the potential exaggeration of area fluctuations caused by transient observational anomalies, we conservatively defined the genuine physical variation using the 95th and 5th percentiles of the historical area time series, rather than absolute extremes.

4 Results

4.1 Validation of lake extraction accuracy

4.1.1 Classification accuracy and boundary validation

The accuracy assessment of binary classifiers is crucial in lake classification practices (Olofsson et al., 2014; Stehman and Foody, 2019). A 30 % subset of the labelled samples was reserved as a validation set to evaluate the three algorithms' binary classification of water and non-water. 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. 3a), 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, achieving an overall classification accuracy of 93.21 % (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. 3b). Overall, the random forest showed higher accuracy and stability than SVM and CART in distinguishing water from non-water.

Table 3. Comparison of accuracy between different classification algorithms for the "water bodies" and "non-water bodies" categories.

Algorithm	Class	Precision	Recall	F_1-score
Random Forest	Water	0.986	0.875	0.927
SVM	Water	0.972	0.824	0.892
CART	Water	0.903	0.869	0.886
Random Forest	Non-water	0.890	0.988	0.936
SVM	Non-water	0.854	0.967	0.907
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 F_1-score (Table 3). For the water class, RF performed best across all metrics, with precision, recall, and F_1-score of 0.986, 0.875, and 0.927, respectively, exceeding SVM (0.972, 0.824, 0.892) and CART (0.903, 0.869, 0.886) (Fig. 3c). For the non-water class, RF also maintained its leading advantage, with precision, recall, and F_1-score of 0.890, 0.988, and 0.936, all higher than the other two algorithms (Fig. 3d). These results indicate that RF offers strong performance and practical value for remote sensing classification of lake water on the Qinghai–Tibet Plateau, providing a useful reference for subsequent monitoring of lake area dynamics and spatiotemporal analyses.

Based on the sub-pixel evaluation against the Sentinel-2 ground truth, the quantitative results demonstrate consistently high precision (> 0.95) across all size categories. This indicates that our algorithm and morphological post-processing effectively minimise false positives, regardless of lake size (Fig. 4a). For the smallest tier (10–50 km^2), the F_1-score is 0.947. This slight reduction is primarily driven by a lower recall (0.927). The lower recall reflects the inherent physical limitations of the 500 m sensor. Coarse spatial resolution results in minor omission errors along the highly convoluted, fragmented shorelines of small lakes. Visually, the extracted boundaries for representative lakes of each size tier align closely with the water bodies in the corresponding seasonal satellite imagery (Fig. 4b). We selected Rena Co

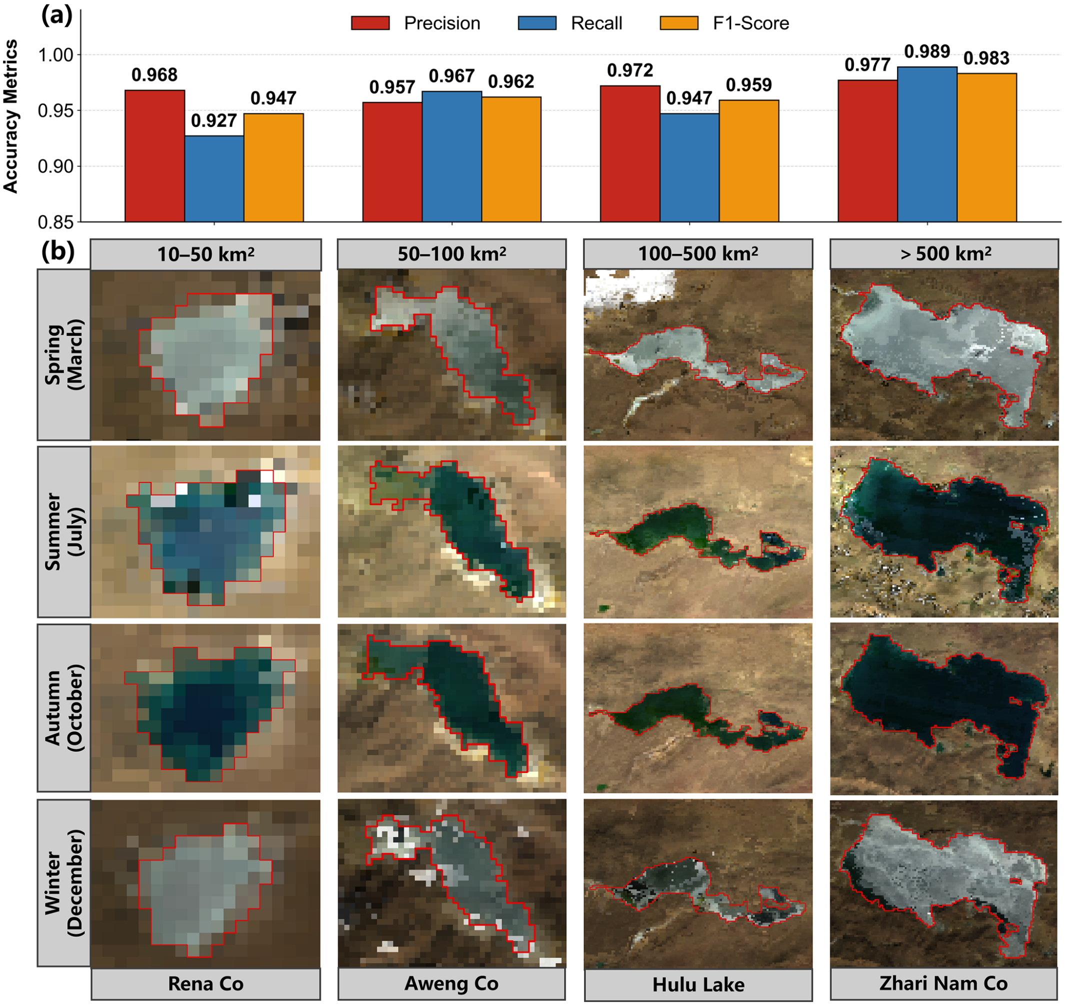

Figure 4. Comprehensive visual and quantitative validation of lake boundary extraction stratified by size categories. **(a)** Overall spatial accuracy metrics (Precision, Recall, and F_1-score) were evaluated across four distinct lake-size tiers derived from the comprehensive sub-pixel validation dataset. **(b)** Seasonal visual extraction results for four representative lakes serving as visual examples for each respective size tier: Rena Co (10–50 km^2), Aweng Co (50–100 km^2), Hulu Lake (100–500 km^2), and Zhari Nam Co (> 500 km^2). Red outlines indicate the extracted continuous boundaries at the 500 m spatial resolution. The background imagery is based on MODIS/Terra MOD09A1 Collection 6.1 8 d surface reflectance imagery (500 m). Data source: NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC).

(10–50 km^2), Aweng Co (50–100 km^2), Hulu Lake (100–500 km^2), and Zhari Nam Co (> 500 km^2). The method accurately captures shoreline inflections, peninsulas, and bay entrances. The extraction demonstrates robust performance under various challenging surface conditions, including partial snow cover, cloud shadows, and topographic shadows. Seasonally, these lakes exhibit typical highland lake variation. Smaller lakes (10–50 km^2) reach their peak in summer (July), whereas larger lakes (> 5 km^2) reach their maximum area in autumn (October). Generally, their areas are smallest in spring and gradually contract in winter. This confirms that our TPLake-MED dataset preserves robust boundary delineation and sub-pixel accuracy, even near the theoretical spatial resolution limits of the MODIS sensor.

4.1.2 Uncertainty assessment

For large lakes (> 500 km^2), the median relative area error under a 500 m pixel shift is minimal (< 10 %). However, for small lakes (10–50 km^2), the median relative uncertainty reaches approximately 40 %–50 % due to their high perimeter-to-area ratio (Fig. 5a). The SNR evaluation of our dataset yields the following statistics: for the > 500 km^2 group, the median SNR is 1.41, with 88.2 % of lakes exhibiting genuine variations exceeding the noise threshold (SNR > 1). For the 100–500 km^2 group, the median SNR is 1.29, with 76.4 % valid variations. For the 50–100 km^2 group, the median SNR is 1.23, with 63.5 % valid variations. Finally, for the 10–50 km^2 group, the median SNR is 0.73, with 40.4 % valid variations. While the 500 m resolution noise may mask subtle hydrological dynamics for a portion of small lakes, over 40 % of small lakes (115 lakes) successfully breached the strict noise threshold (SNR $>$ 1) (Fig. 5b). Even under coarse spatial resolution constraints, a significant portion of small lakes exhibits undeniable area fluctuations.

4.2 Comparison with other products

4.2.1 Seasonal product comparison

To validate the accuracy of our lake area estimates, we compared results for 20 lakes with the dataset of Li et al. (2025c). 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 0.9996, URMSE was 0.0029, and RMSE was 21.43; summer R^2 was 0.9995, URMSE was 0.0033, and RMSE was 24.30; autumn R^2 was 0.9994, URMSE was 0.0032, and RMSE was 26.05; winter R^2 was 0.9996, URMSE was 0.0032, and RMSE was 21.87 (Fig. 6). 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.

4.2.2 Inter-annual product comparison

Given the use of MOD09A1 and the available mapping period, we analysed annual and monthly variations in lake area from 2000 to 2024 and compared them with four interannual products from the National Tibetan Plateau Data Center. Using Qinghai Lake as a demonstration case, the lake exhibited a clear long-term expansion trend, with its area increasing from approximately 4300 km^2 in the early 2000s to more than 4600 km^2 after 2020 (Fig. 7a). 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 the 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 exhibited an R^2 of 0.96, RMSE of 1019.31 km^2, MAPE of 2.2 %, and a bias of -832.49 km^2 relative to Zhou et al. (2025) (Fig. 7b); 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 and Jin (2023) (Fig. 7c); 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 and Ran (2022) (Fig. 7d); 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 (2019) (Fig. 7e). Scatter points aligned closely along the 1 : 1 line, indicating high agreement with existing interannual datasets. Agreement was particularly high with Zhang and Ran (2022) and Wang and Jin (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 provides a sound basis for long-term analyses of lake change.

4.2.3 Global products comparison

To provide a broader global outlook and evaluate TPLake-MED within the context of established international frameworks, we conducted a comprehensive comparison with the JRC Global Surface Water (GSW) and HydroLAKES datasets. During clear-sky conditions, typically in summer, TPLake-MED demonstrates high spatial consistency with both JRC and HydroLAKES, which validates that our MODIS-based random forest algorithm captures the baseline lake extent with reliability comparable to high-resolution global products when atmospheric interference is minimal. However, as shown in the time-series comparison for Nam Co (Fig. 8a), the JRC dataset exhibits frequent and significant data gaps, with area estimates often dropping to zero, primarily due to the strict filtering of cloud-contaminated pixels in Landsat-based global products. In contrast, TPLake-MED maintains high temporal continuity across the 2000–2024 period, successfully capturing the intra-annual seasonal fluctuations that are often lost in global observations (Fig. 8b). The comparative advantage of TPLake-MED is most evident under challenging atmospheric conditions; during months with heavy cloud cover (e.g., June 2006, Fig. 8c–e) or extensive snow/ice interference (e.g., January 2005, Fig. 8f–h), the JRC product often results in fragmented water extents (Fig. 8d and g), whereas TPLake-MED leverages high-frequency MODIS observations and our morphological optimisation workflow to effectively distinguish water from snow/ice and reconstruct complete lake boundaries (Fig. 8e and h) where global products struggle. Furthermore, while HydroLAKES provides a robust static baseline for global hydrological modelling, it cannot reflect the rapid expansion of TP lakes, such as the ~ 300 km^2 expansion of Nam Co shown in Fig. 8b. TPLake-MED fills this gap by providing a dynamic, monthly-resolved vector dataset that couples long-term trends with seasonal precision.

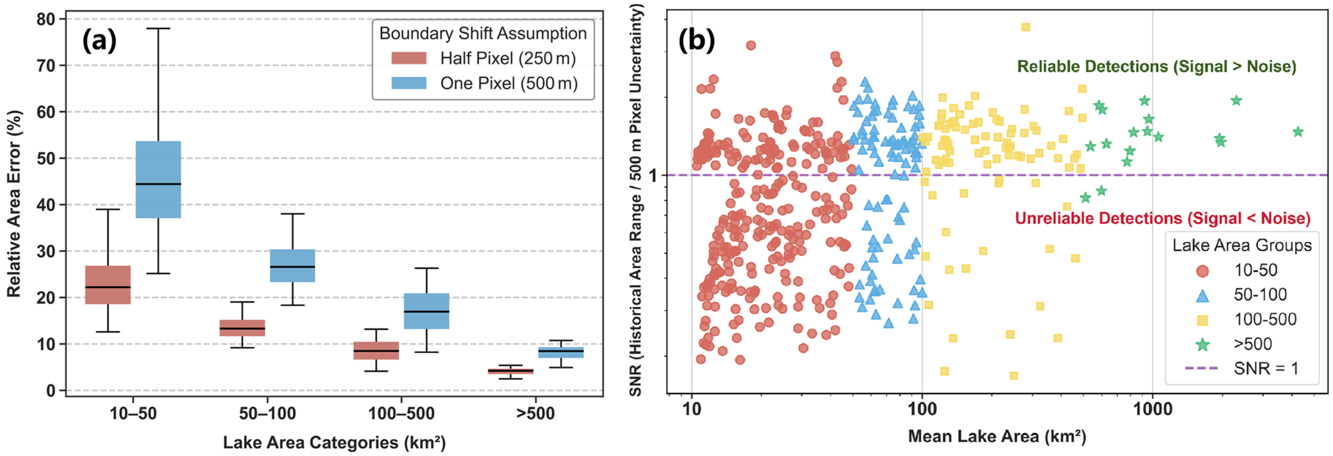

Figure 5. Quantitative assessment of lake area uncertainty and signal-to-noise ratio (SNR) analysis. **(a)** Simulated relative area error (%) under hypothetical worst-case boundary shifts of a half-pixel (250 m, red) and one-pixel (500 m, blue), stratified by four distinct lake size categories. **(b)** SNR evaluation of historical lake area variations over the 25-year period. The horizontal purple dashed line indicates the detection threshold of SNR = 1, separating reliable (Signal > Noise) and unreliable (Signal < Noise) area variation detections.

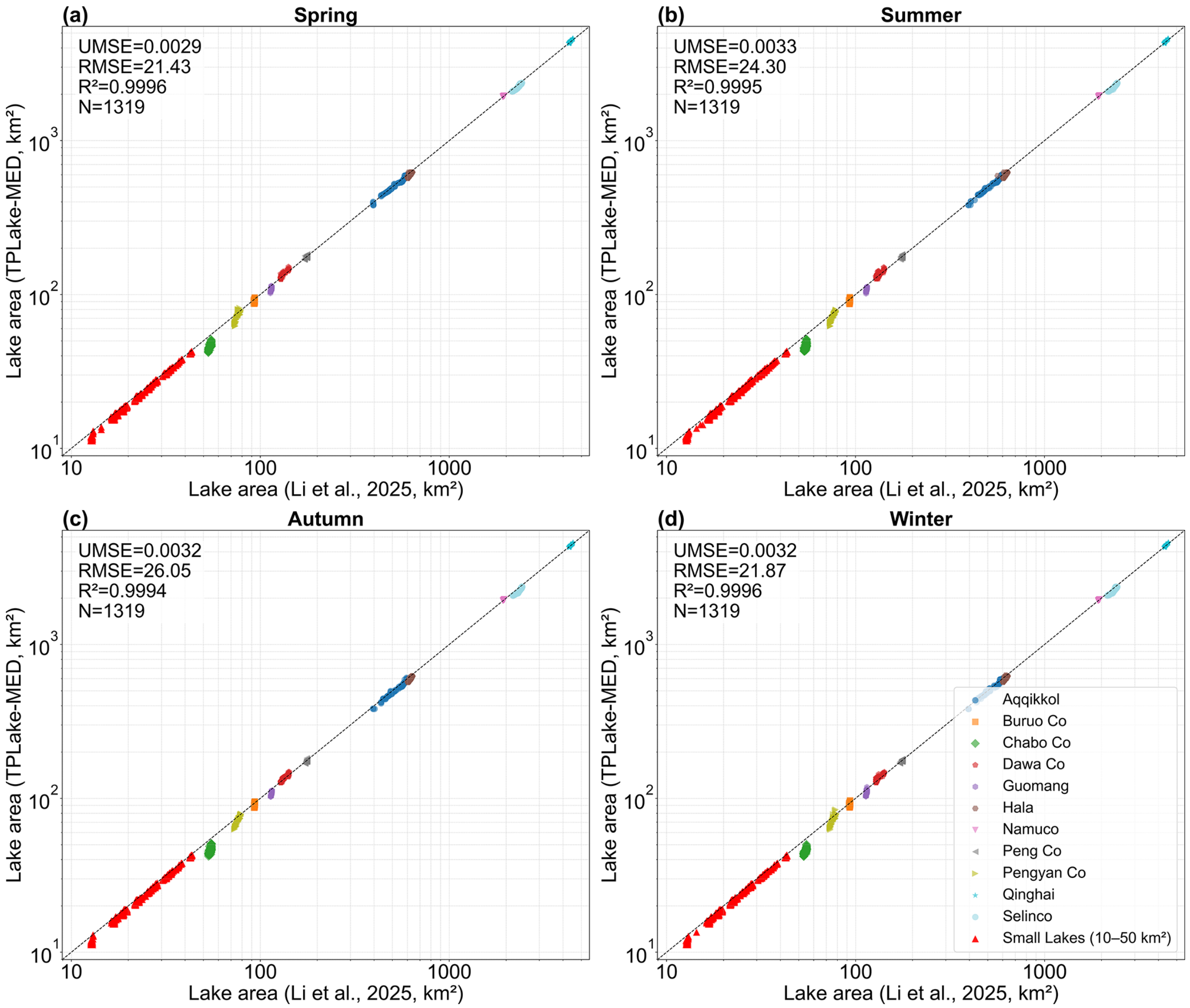

Figure 6. Scatter plot comparing different lake area extraction methods in spring, summer, autumn, and winter.

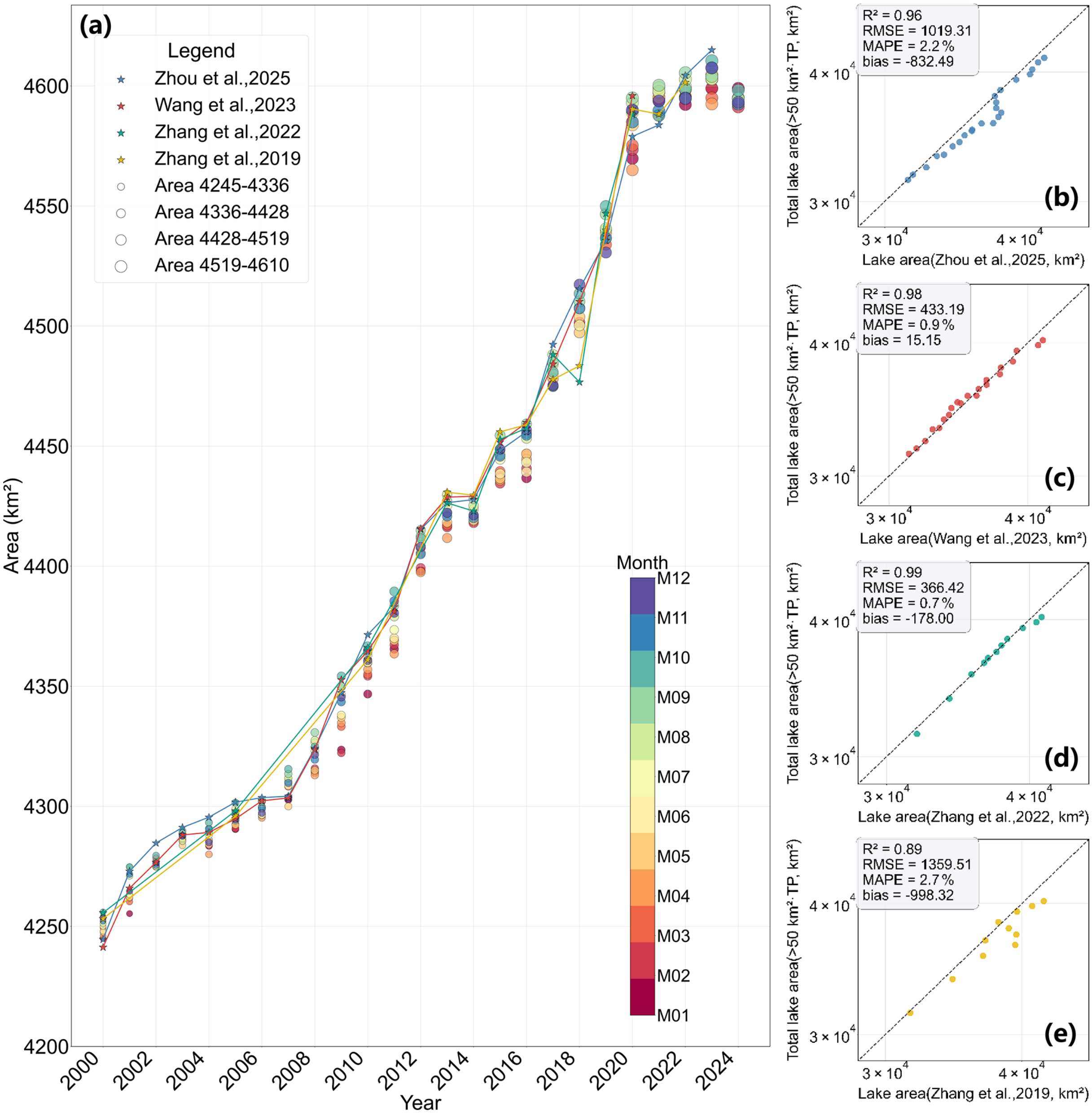

Figure 7. Temporal variations of Qinghai Lake area from 2000 to 2024 and consistency comparison with existing interannual datasets. **(a)** Annual variations in the Qinghai Lake area derived from this study (2000–2024). Monthly lake areas are shown as coloured scatter points (M01–M12) using a spectral gradient, and the annual results from four existing datasets are marked with star symbols. **(b–e)** The comparison of total lake area ($> 50\,\mathrm{km}^2$) from this study with four reference datasets.

4.3 Changes in Qinghai–Tibet Plateau lakes (2000–2024)

4.3.1 Yearly lake expansion: patterns and responses

By combining a random forest classifier with morphological methods, we generated a dataset (TPLake-MED) of lake boundary ranges for lakes larger than $10\,\mathrm{km}^2$ on the Qinghai–Tibet Plateau from 2000 to 2024, with a spatial resolution of 500 m. Among lakes larger than $50\,\mathrm{km}^2$, 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. 9a). 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) (Tao et al., 2015; Du et al., 2023). As of 2024, the fastest growing lakes by relative area change were Selin Co ($2369.5\,\mathrm{km}^2$; +21.6 %),

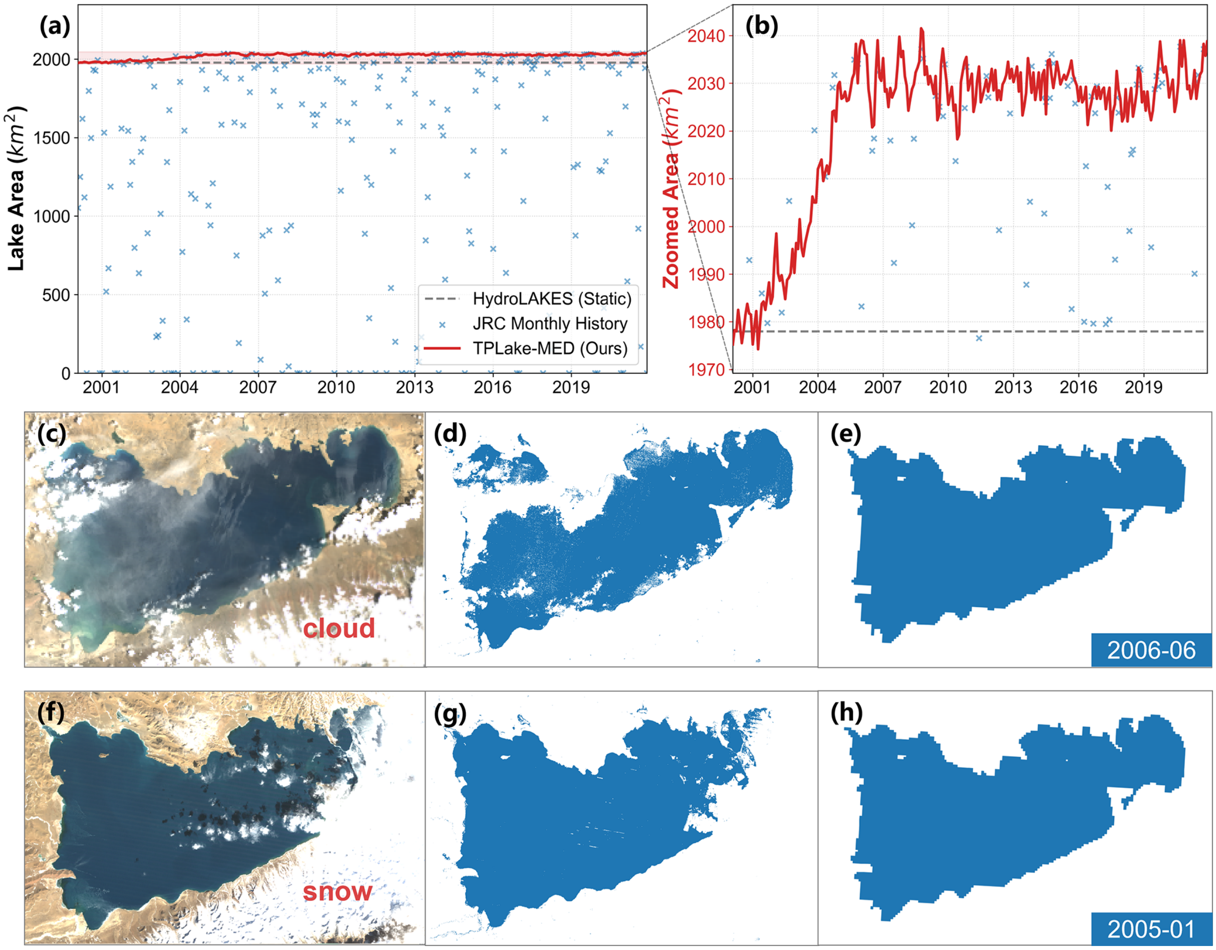

Figure 8. Comparison of temporal consistency and spatial completeness between the TPLake-MED dataset and the JRC Global Surface Water product under complex atmospheric conditions. **(a)** The long-term time-series comparison (2000–2024) of the lake area for Nam Co. The overview highlights frequent data gaps and near-zero values in the JRC dataset, compared to TPLake-MED's continuous monitoring capability. **(b)** The zoomed-in panel details the seasonal dynamics captured by TPLake-MED, demonstrating high consistency with valid JRC observations during clear-sky periods and reflecting the significant decadal expansion compared to the static HydroLAKES baseline. **(c–e)** Visual verification of spatial completeness during a cloud-contaminated month (June 2006). **(c)** Reference Landsat imagery obscured by clouds; **(d)** fragmented JRC water extent; **(e)** complete lake boundary extracted by TPLake-MED. **(f–h)** Comparison during a snow-covered month (January 2005). **(f)** Reference imagery; **(g)** JRC result with snow interference; **(h)** optimised TPLake-MED boundary. (JRC data © European Commission Joint Research Centre; Landsat imagery courtesy of USGS.)

Aqikkol (570.5 km^2; +17.7 %), Ayagekumuli (1077.75 km^2; +20.3 %), Duoersuo Co (1031.5 km^2; +13.5 %), Hulu Lake (311.25 km^2; +18.2 %), and Shuanglian Lake (270.5 km^2; +11.3 %) (Fig. 9b). The total lake area shows a significant upward trend, with an average annual growth rate of 34.91 $\mathrm{km}^2\,\mathrm{yr}^{-1}$, and the correlation is significant ($p < 0.05$), and the overall expansion of high-elevation lakes is evident. In particular, the total area of lakes larger than 50 km^2 increased by 32.5 % compared with 2000 (Fig. 9c). 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. 9d), when precipitation was relatively abundant, while temperature and evapotranspiration were relatively low. This pattern indicates joint control of the lake area by precipitation and evapotranspiration (Li et al., 2022b).

4.3.2 Monthly area change: heterogeneity and scale

The monthly rate of lake area change on the Qinghai–Tibet 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

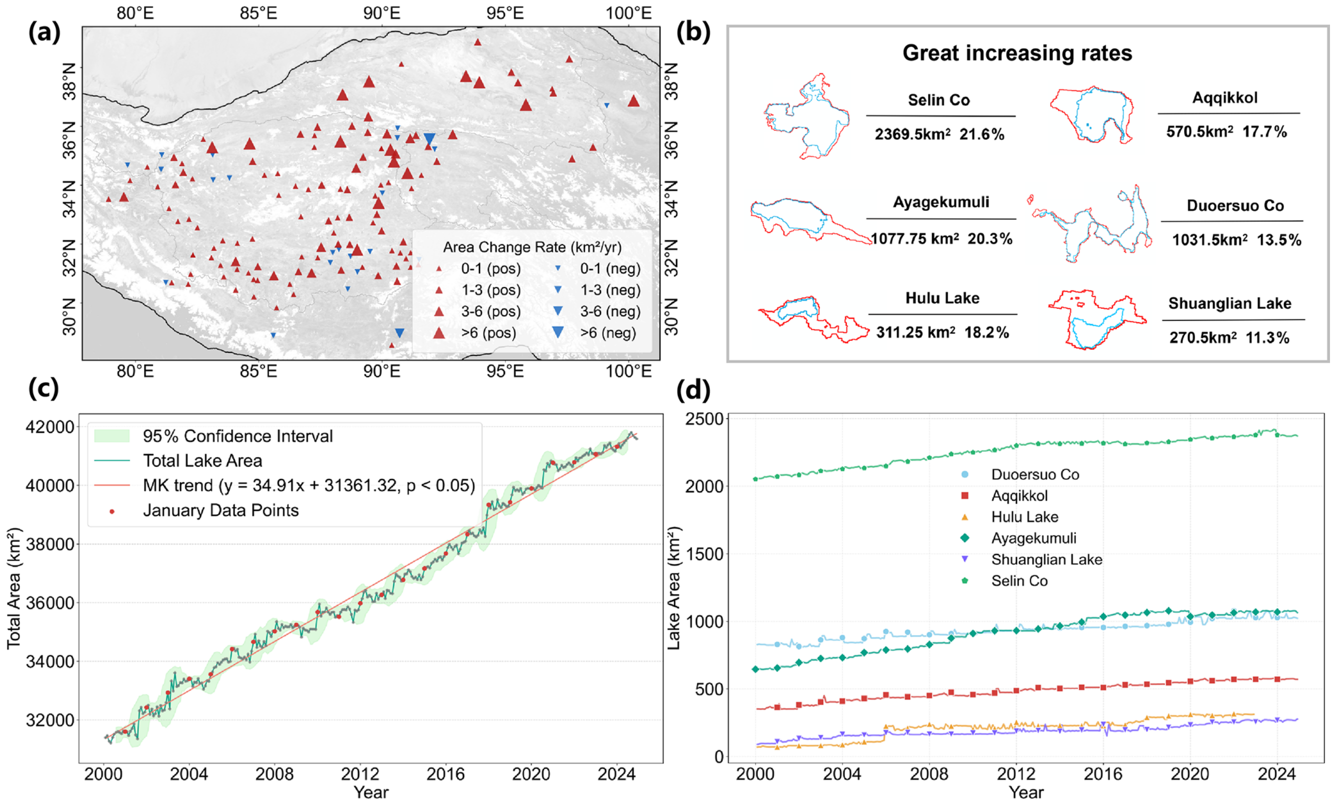

Figure 9. Spatial-temporal characteristics of lake area changes and trends in typical lakes from 2000 to 2024. **(a)** The spatial distribution of interannual changes in lake area. Triangular symbols indicate the interannual area change rates ($\mathrm{km}^2\ \mathrm{yr}^{-1}$) of lakes on the Qinghai–Tibet 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 Qinghai–Tibet 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.

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, stabilising 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. 10a). This spatial variation was closely linked to regional climatic conditions. To explicitly quantify these driving mechanisms across the longitudinal gradient, we conducted a Pearson correlation analysis, coupled with a standardised multiple linear regression, for all studied lakes across the western, central, and eastern zones (Fig. S1 in the Supplement). The statistical results reveal that in the western region, lake dynamics are jointly dominated by seasonal precipitation pulses (39.8 % contribution, $r = 0.83$) and intense evaporation (34.1 %), leading to pronounced lake water fluctuations. Conversely, moving eastward, temperature emerges as the primary driver (38.8 % contribution, $r = 0.74$), while the impact of evaporation drops significantly to 27.2 %. Because temperature is the primary driver of cryospheric ablation, this confirms that the eastern lakes are largely supplied by glacial runoff and experience relatively smoother changes under a more stable climate regime. The central region exhibits a remarkably balanced transition among precipitation (35.1 %), temperature (32.7 %), and evaporation (32.2 %).

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. 10b). 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 km^2 exhibited the highest variation rates and the greatest variability, with multiple outliers. Lakes between 150 and 500 km^2 showed intermediate variation rates and variability. Lakes between 500–1000 km^2 demonstrated a marked decrease in variation rates with a more concentrated distribu-

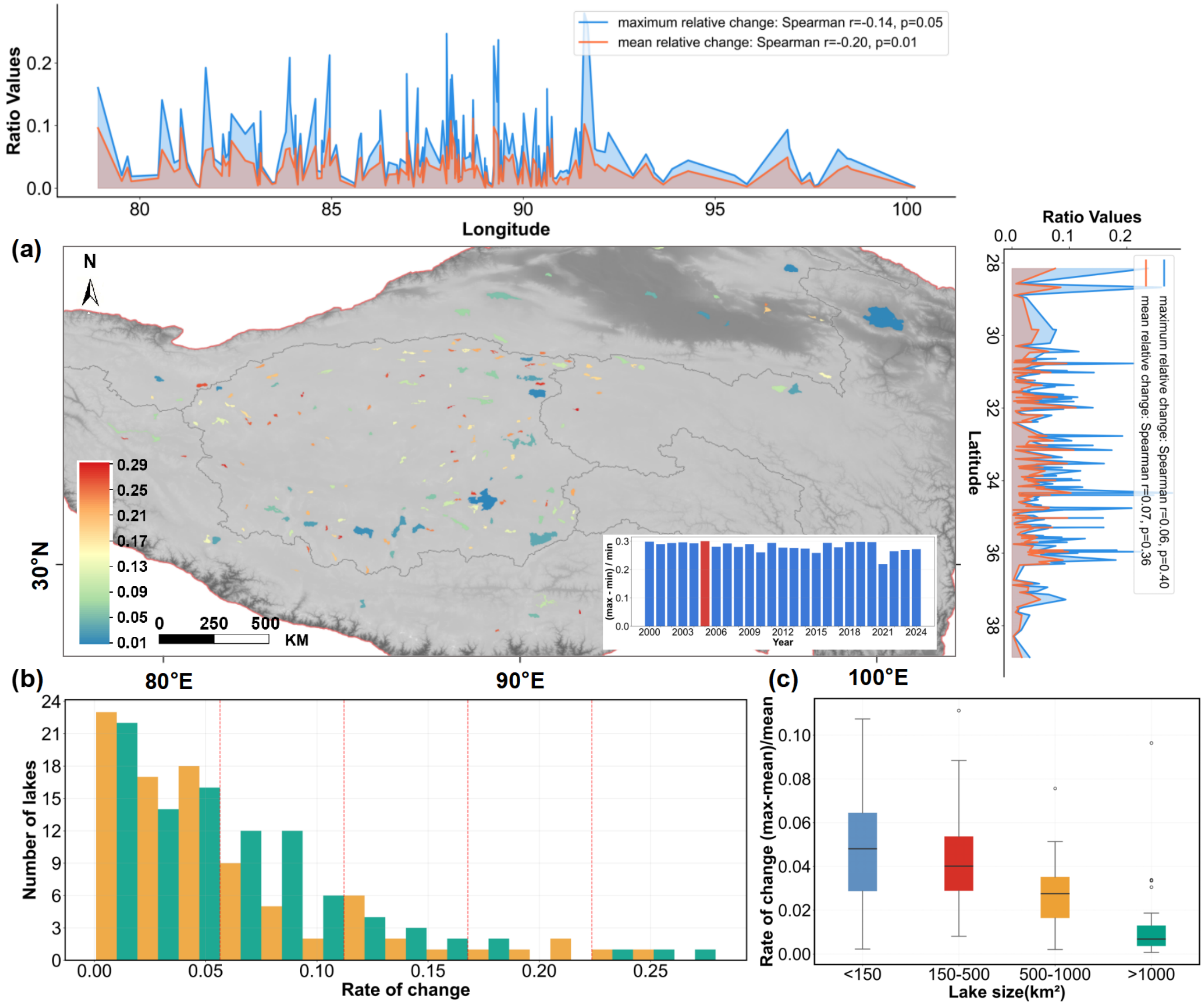

Figure 10. Spatial distribution and statistical characteristics of inter-monthly lake area change rates on the Qinghai–Tibet Plateau. **(a)** Temporal variation and spatial distribution of inter-monthly change rates, with the orange and blue lines representing mean and maximum relative changes, the red 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.

tion. Lakes larger than 1000 km^2 had the lowest variation rates and minimal variability, with only a few isolated outliers (Fig. 10c). 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 Qinghai–Tibet Plateau (Zhu et al., 2025).

Inter-monthly lake-area variation across the Qinghai–Tibet Plateau showed spatial heterogeneity, with many lakes exhibiting pronounced fluctuations in surface area. Monthly area differences of 6.50–26.38 km^2 occurred predominantly in the plateau interior, indicating heightened sensitivity to climatic shifts. Differences of 3.91–6.50 km^2 were widely distributed across the broader plateau region, indicating moderate variability; and differences of 0.25–3.91 km^2 occurred mainly along plateau margins and transition zones, representing relatively stable lake systems (Fig. 11a). Analysis of lakes with significant variability reveals distinct intra-annual maximum and minimum boundary differences, highlighting the intensity of seasonal fluctuations (Fig. 11b–i). Over-

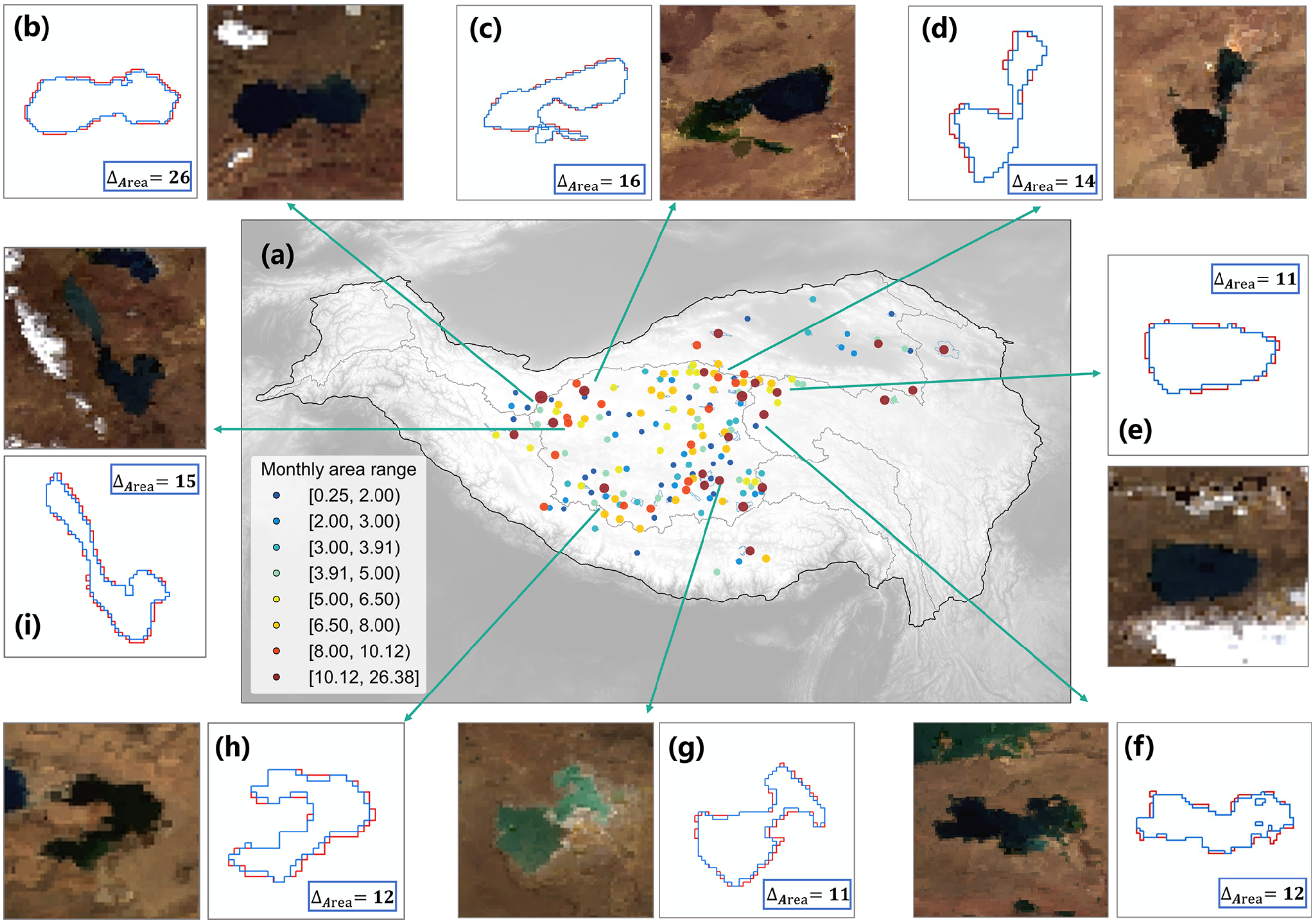

Figure 11. Spatial distribution of inter-monthly lake area variation on the Qinghai–Tibet Plateau and verification of boundary changes for representative lakes. **(a)** The spatial distribution pattern of inter-monthly lake area variation (maximum–minimum) across the Qinghai–Tibet Plateau. **(b–i)** 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 (km^2). The representative lake image panels are based on MODIS/Terra MOD09A1 Collection 6.1 8 d surface reflectance imagery (500 m). Data source: NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC).

lay 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 multiscale presentation revealed macro-spatial patterns in plateau lake dynamics and supported data quality and analytical reliability through detailed validation of representative lakes. It provided evidence to deepen the understanding of the stability and vulnerability of plateau lake ecosystems.

5 Data availability

The monthly scale vector boundary dataset of lakes larger than 10 km^2 on the Qinghai–Tibet Plateau from 2000 to 2024 (TPLake-MED) constructed in this study has been publicly shared on Zenodo, accessible at: https://doi.org/10.5281/zenodo.19205532 (Zhao et al., 2026). 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

The code is available from the first and corresponding authors upon request.

7 Conclusions

Qinghai–Tibet Plateau lakes are changing rapidly, but most existing datasets describe only 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 km^2 across the Qinghai–Tibet Plateau for 2000–2024. Using a random forest model and morphological optimisation, we improved water classification accuracy (precision of 0.986, recall of 0.875, and an F_1-score of 0.927) and boundary precision. The dataset reveals that total lake area has increased steadily ($\sim 34.91\,\mathrm{km^2\,yr^{-1}}$); 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 the strongest variability; and smaller lakes are more sensitive to environmental change than larger 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.

Supplement. The supplement related to this article is available online at https://doi.org/10.5194/essd-18-4219-2026-supplement.

Author contributions. SZ and XZ: conceptualization; SZ: methodology and software; SZ: writing – original draft; SZ: writing – review and editing; XZ: funding acquisition and project administration; XZ and JZ: writing – review and editing, data curation, and supervision; SZ, XL, and CY: data curation. All of the authors participated in the review and editing of the paper.

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

Acknowledgements. We gratefully acknowledge the National Tibetan Plateau Data Center for providing four interannual lake area datasets: Annual 30 m lake maps on the Tibetan Plateau (1991–2023) (V1.0) (https://doi.org/10.5281/zenodo.10686952) provided by Zhou et al. (2025); Time series dataset of lake area on the Tibetan Plateau for the past 100 years (1920–2020) (https://doi.org/10.11888/Terre.tpdc.272891) provided by Zhang and Ran (2022); The lakes larger than 1 km^2 in Tibetan Plateau (v3.1) (1970s–2022) (https://doi.org/10.1016/j.scib.2019.07.018) provided by Zhang (2019); and Annual area dataset of lakes over 50 km^2 on the Tibetan Plateau (1986–2020) (https://doi.org/10.11888/Terre.tpdc.300883) provided by Wang and Jin (2023). We thank the European Commission Joint Research Centre for providing the JRC Global Surface Water dataset (https://global-surface-water.appspot.com/, last access: 7 June 2026). We appreciate the HydroSHEDS team for providing the HydroLAKES dataset (https://www.hydrosheds.org/products/hydrolakes, last access: 7 June 2026). We also thank the National Earth System Science Data Center, National Science & Technology Infrastructure of China (http://www.geodata.cn, last access: 8 June 2026), and the National Aeronautics and Space Administration (NASA) Earth Observing System Data and Information System (EOSDIS) Land Processes Distributed Active Archive Center (LP DAAC) for providing the MODIS/Terra MOD09A1 surface reflectance product (https://doi.org/10.5067/MODIS/MOD09A1.061, Vermote, 2021). We extend our gratitude to Li et al. (2025c) for providing the seasonal comparative lake dataset (https://doi.org/10.1038/s41586-025-09046-3). Finally, we thank the editor and the anonymous reviewers for their constructive comments and suggestions.

Financial support. This research has been supported by the National Natural Science Foundation of China (grant no. 42330205) and the Open Fund of the State Key Laboratory of Remote Sensing and Digital Earth, Beijing Engineering Research Center for Global Land Remote Sensing Products (grant nos. OF202506 and OF202507).

Review statement. This paper was edited by Attila Demény and reviewed by Yingkui Li and one anonymous referee.

References

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

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 Forests, Mach. Learn., 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Carrivick, J. L. and Tweed, F. S.: A global assessment of the societal impacts of glacier outburst floods, Global Planet. Change, 144, 1–16, https://doi.org/10.1016/j.gloplacha.2016.07.001, 2016.

Chen, W. F., Liu, Y., Zhang, G. Q., 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.

Congalton, R. G.: A review of assessing the accuracy of classifications of remotely sensed data, Remote Sens. Environ., 37, 35–46, https://doi.org/10.1016/0034-4257(91)90048-B, 1991.

Deng, H. J., Pepin, N. C., Liu, Q., and Chen, Y. N.: Understanding the spatial differences in terrestrial water storage variations in the Tibetan Plateau from 2002 to 2016, Climatic Change, 151, 379–393, https://doi.org/10.1007/s10584-018-2325-9, 2018.

Ding, W.-N., Ree, R. H., Spicer, R. A., and Xing, Y.-W.: Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora, Science, 369, 578–581, https://doi.org/10.1126/science.abb4484, 2020.

Du, B. L., Zhu, L. P., Ju, J. T., Wang, J. B., Ma, Q. F., and Kou, Q. Q.: A Quantification of Heat Storage Change-Based Evaporation Behavior in Middle–Large-Sized Lakes in the Inland of the Tibetan Plateau and Their Temporal and Spatial Variations, Remote Sens., 15, 3460, https://doi.org/10.3390/rs15143460, 2023.

Emmer, A., Allen, S. K., Carey, M., Frey, H., Huggel, C., Korup, O., Mergili, M., Sattar, A., Veh, G., Chen, T. Y., Cook, S. J., Correas-Gonzalez, M., Das, S., Moreno, A. D., Drenkhan, F., Fischer, M., Immerzeel, W. W., Izagirre, E., Joshi, R. C., Kougkoulos, I., Knapp, R. K., Li, D., Majeed, U., Matti, S., Moulton, H., Nick, F., Piroton, V., Rashid, I., Reza, M., de Figueiredo, A. R., Riveros, C., Shrestha, F., Shrestha, M., Steiner, J., Walker-Crawford, N., Wood, J. L., and Yde, J. C.: Progress and challenges in glacial lake outburst flood research (2017–2021): a research community perspective, Nat. Hazards Earth Syst. Sci., 22, 3041–3061, https://doi.org/10.5194/nhess-22-3041-2022, 2022.

Fawcett, T.: An introduction to ROC analysis, Pattern Recogn. Lett., 27, 861–874, https://doi.org/10.1016/j.patrec.2005.10.010, 2006.

Feyisa, G. L., Meilby, H., 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.

Gu, C. J., Zhang, Y. L., Liu, L. S., Wei, B. H., Cui, B. Q., and Gong, D. G.: Evaluation of consistency among four NDVI datasets applied to Three River Source Region, Qinghai Province, China, Geogr. Res., 42, 1378–1392, https://doi.org/10.11821/dlyj020220703, 2023.

Han, W. X., Huang, C. L., Wang, Y. C., 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.1001-8166.2019.4.0346, 2019.

Huang, D., Zhang, J. Y., Liu, J. F., Hu, X. G., and Jin, J. L.: 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.

Huggel, C., Carey, M., Emmer, A., Frey, H., Walker-Crawford, N., and Wallimann-Helmer, I.: Anthropogenic climate change and glacier lake outburst flood risk: local and global drivers and responsibilities for the case of lake Palcacocha, Peru, Nat. Hazards Earth Syst. Sci., 20, 2175–2193, https://doi.org/10.5194/nhess-20-2175-2020, 2020.

Hui, Y. J., Hu, H. D., Xiang, J. H., and Du, X. Y.: Comparative analysis of machine learning algorithms for predicting tibial intramedullary nail length from patient characteristics, J. Orthop. Traumatol., 26, 56, 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.

Jaccard, P.: The distribution of the flora in the Alpine zone. 1, New Phytol., 11, 37–50, https://doi.org/10.1111/j.1469-8137.1912.tb05611.x, 1912.

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, 2022.

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, https://doi.org/10.1016/j.rse.2021.112294, 2021.

Li, B. H., Liu, K., Wang, M., Wang, Y. F., Zhuang, L. M., Zhu, W. H., Li, C. X., Zhang, L. H., and Chen, Y. N.: Evaluating the performance of pixel-based and object-based multidimensional clustering algorithms for automated surface water mapping, Geo-spatial Inf. Sci., 0, 1–21, https://doi.org/10.1080/10095020.2025.2523993, 2025a.

Li, G. H., Li, Z. X., Zhang, B. Q., and Li, Z. J.: 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, 2025b.

Li, L. Q., Long, D., Wang, Y. M., 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, 2025c.

Li, M., Weng, B., Yan, D. H., Bi, W. X., 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, 2022a.

Li, X. F., Yao, X. J., Sun, M. P., Gong, P., An, L. N., Qi, M. M., and Gao, Y. P.: 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. S., Zhang, D. H., Jiang, C. C., Zhao, Y. J., Li, H., Lu, D. H., Qin, K., Chen, D. H., Liu, Y. F., Sun, Y., and Liu, S. S.: Comparison of Lake Area Extraction Algorithms in Qinghai Tibet Plateau Leveraging Google Earth Engine and Landsat-9 Data, Remote Sens., 14, 4612, https://doi.org/10.3390/rs14184612, 2022b.

Li, Y. H., Zhang, Q. M., Zhou, X. J., and Jia, K. 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.

Li, Y. K., Liao, J. J., Guo, H. D., Liu, Z. W., and Shen, G. Z.: Patterns and Potential Drivers of Dramatic Changes in Tibetan Lakes, 1972–2010, PLOS ONE, 9, e111890, https://doi.org/10.1371/journal.pone.0111890, 2014.

Liao, J. J., Shen, G. Z., and Li, Y. K.: Lake variations in response to climate change in the Tibetan Plateau in the past 40 years, Int. J. Digit. Earth, 6, 534–549, https://doi.org/10.1080/17538947.2012.656290, 2013.

Liu, H. C., He, G. J., Peng, Y., Wang, G. Z., and Yin, R. Y.: A dataset of land surface water with a spatial resolution of 30 meters on the Qinghai–Tibet Plateau in 2022, China Sci. Data, 8, https://doi.org/10.11922/11-6035.csd.2023.0040.zh, 2023.

Liu, Z., Chao, N. F., Chen, G., Zhang, G. Q., Wang, Z. T., Li, F. P., and Ouyang, G. C.: Changes in monthly surface area, water level, and storage of 194 lakes and reservoirs in the Yangtze River Basin during 1990–2021 using multi-source remote sensing data, Sci. Total Environ., 944, 173840, https://doi.org/10.1016/j.scitotenv.2024.173840, 2024.

Loh, W.-Y.: Classification and regression trees, Wiley Interdisciplin. Rev.-Data Min. Knowl. Discov., 1, 14–23, https://doi.org/10.1002/widm.8, 2011.

Long, D., Li, X. Y., Li, X. D., Han, P. F., Zhao, F. Y., Hong, Z. K., Wang, Y. M., and Tian, F. Q.: 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.

Ma, S. M., Gan, F. P., Wu, H. C., and Yan, B. K.: 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.

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., and Schmitt, O.: Estimating the volume and age of water stored in global lakes using a geo-statistical approach, Nat. Commun., 7, 13603, https://doi.org/10.1038/ncomms13603, 2016.

Miao, Q. F., Liu, X. Y., Shi, H. B., Wei, Z. M., Luo, Y. L., Wang, Y. H., Gonçalves, J. M., and Feng, W. Y.: 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. L., Zhu, L. P., and Yang, R. M.: Interannual Variation in the Area and Water Volume of Lakes in Different Regions of the Tibet Plateau and Their Responses to Climate Change, Front. Earth Sci., 9, 738018, https://doi.org/10.3389/feart.2021.738018, 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. J., Liu, L., Zhang, G. Q., Shum, C. K., Qiu, J. H., Hu, R. G., Li, J. P., Peng, J. H., Hwang, C., Luan, Y., Sun, Y., Xu, M., Chen, D., Ding, J., and Zhong, Y. L.: Contrasting lake changes in Tibet revealed by recent multimodal satellite observations, Sci. Total Environ., 908, 168342, https://doi.org/10.1016/j.scitotenv.2023.168342, 2024.

Reza, M., Joshi, R. C., Nabiyal, C., and Negi, M.: Glacier Retreat and Pro Glacial Lake dynamics of Darma Valley, Central Himalaya, India, Bull. Geogr. Phys. Geogr. Ser., 51–62, https://doi.org/10.12775/bgeo-2023-0009, 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.

Shugar, D. H., Burr, A., Haritashya, U. K., Kargel, J. S., Watson, C. S., Kennedy, M. C., Bevington, A. R., Betts, R. A., Harrison, S., and Strattman, K.: Rapid worldwide growth of glacial lakes since 1990, Nat. Clim. Change, 10, 939–945, https://doi.org/10.1038/s41558-020-0855-4, 2020.

Song, C. Q., Huang, B., Richards, K., Ke, L. H., and Hien Phan, V.: Accelerated lake expansion on the Tibetan Plateau in the 2000s: Induced by glacial melting or other processes?, Water Resour. Res., 50, 3170–3186, https://doi.org/10.1002/2013WR014724, 2014.

Song, C. Q., Zhan, P. F., and Ma, R. H.: 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.

Tao, S. L., Fang, J. Y., Zhao, X., Zhao, S. Q., Shen, H. H., Hu, H. F., Tang, Z. Y., Wang, Z. H., and Guo, Q. H.: Rapid loss of lakes on the Mongolian Plateau, P. Natl. Acad. Sci. USA, 112, 2281–2286, https://doi.org/10.1073/pnas.1411748112, 2015.

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, 1979.

Vermote, E.: MODIS/Terra Surface Reflectance 8-Day L3 Global 500 m SIN Grid V061, NASA Land Processes Distributed Active Archive Center [data set], https://doi.org/10.5067/MODIS/MOD09A1.061, 2021.

Wan, W., Xiao, P. F., Feng, X. Z., Li, H., Ma, R. H., Duan, H. T., and Zhao, L. M.: 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, X., Liu, S. Y., Guo, W. Q., Yao, X. J., Jiang, Z. L., and Han, Y. S.: Using Remote Sensing Data to Quantify Changes in Glacial Lakes in the Chinese Himalaya, Mt. Res. Dev., 32, 203–212, https://doi.org/10.1659/MRD-JOURNAL-D-11-00044.1, 2012.

Wang, X. L., Zhou, G. S., Lv, X. M., Zhou, L., Hu, M. C., He, X. H., and Tian, Z. H.: Comparison of Lake Extraction and Classification Methods for the Tibetan Plateau Based on Topographic-Spectral Information, Remote Sens., 15, 267, https://doi.org/10.3390/rs15010267, 2023.

Wang, X. R. and Jin, R.: Annual area dataset of lakes over $50\,\mathrm{km}^2$ on the Tibetan Plateau (1986–2020), National Tibetan Plateau/Third Pole Environ. Data Center [data set], https://doi.org/10.11888/Terre.tpdc.300883, 2023.

Wang, X. R., Jin, R., Wang, W. Z., Xu, F. N., Geng, L. Y., and Shao, D. H.: Interannual variation in lake areas over $50\,\mathrm{km}^2$ on the Tibetan Plateau from 1986 to 2020 based on remote sensing big data, Int. J. Digit. Earth, 17, 2300308, https://doi.org/10.1080/17538947.2023.2300308, 2024.

Wen, J. X., Wang, M., and Liu, J.: Time Series Multi-Step Prediction Algorithm Based on Time Series Decomposition and Random Forest, J. East China Univ. Sci. Technol. (Nat. Sci. Ed.), 49, 873–881, https://doi.org/10.14135/j.cnki.1006-3080.20220810001, 2023.

Xu, F. L., Zhang, G. Q., Woolway, R. I., Yang, K., Wada, Y., Wang, J. D., and Crétaux, J.-F.: Widespread societal and ecological impacts from projected Tibetan Plateau lake expansion, Nat. Geosci., 17, 516–523, https://doi.org/10.1038/s41561-024-01446-w, 2024.

Xu, H. Q.: 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, D. J., Huang, C., Ma, N., and Zhang, Y. S.: Improved Landsat-Based Water and Snow Indices for Extracting Lake and Snow Cover/Glacier in the Tibetan Plateau, Water, 12, 1339, https://doi.org/10.3390/w12051339, 2020.

Yan, F. P., Sillanpää, M., Kang, S. C., Aho, K. S., Qu, B., Wei, D., Li, X. F., Li, C. L., and Raymond, P. A.: Lakes on the Tibetan Plateau as Conduits of Greenhouse Gases to the Atmosphere, J. Geophys. Res.-Biogeo., 123, 2091–2103, https://doi.org/10.1029/2017JG004379, 2018.

Yan, L. J., Zheng, M. P., and Qi, L. J.: Surface area variations of lakes in the Tibetan Plateau and their influencing factors, Sci. Technol. Rev., 35, 83–88, 2017.

Yang, J. Y., Jin, X., Jin, Y. X., Chen, K. L., Xie, C. H., Li, Z., and Fu, D.: Wetland area changes and influencing factors in Alpine Inland River Basin, Acta Ecol. Sin., 45, 3684–3699, https://doi.org/10.20103/j.stxb.202409122208, 2025.

Yang, R. M., Zhu, L. P., Wang, J. B., Ju, J. T., Ma, Q. F., Turner, F., and Guo, Y.: Spatiotemporal variations in volume of closed lakes on the Tibetan Plateau and their climatic responses from 1976 to 2013, Climatic Change, 140, 621–633, https://doi.org/10.1007/s10584-016-1877-9, 2017.

Zemp, M., Huss, M., Thibert, E., Eckert, N., McNabb, R., Huber, J., Barandun, M., Machguth, H., Nussbaumer, S. U., Gärtner-Roer, I., Thomson, L., Paul, F., Maussion, F., Kutuzov, S., and Cogley, J. G.: Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016, Nature, 568, 382–386, https://doi.org/10.1038/s41586-019-1071-0, 2019.

Zhang, B. B. and Hu, M.: Customer satisfaction prediction based on improved grid-search random forest, 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.

Zhang, G. Q.: The lakes larger than 1 km^2 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. Q. and Ran, Y. H.: Time series dataset of lake area on the Tibetan Plateau for the past 100 years (1920–2020), National Tibetan Plateau/Third Pole Environ. Data Center, [data set], https://doi.org/10.11888/Terre.tpdc.272891, 2022.

Zhang, G. Q., Yao, T. D., Xie, H. J., Zhang, K. X., and Zhu, F. J.: Lakes' state and abundance across the Tibetan Plateau, Chin. Sci. Bull., 59, 3010–3021, https://doi.org/10.1007/s11434-014-0258-x, 2014.

Zhang, G. Q., Yao, T. D., Piao, S. L., Bolch, T., Xie, H. J., Chen, D. L., Gao, Y., O'Reilly, C. M., Shum, C. K., Yang, K., Yi, S., Lei, Y. B., Wang, W. C., He, Y., Shang, K., Yang, X. K., and Zhang, H. B.: Extensive and drastically different alpine lake changes on Asia's high plateaus during the past four decades, Geophys. Res. Lett., 44, 252–260, https://doi.org/10.1002/2016GL072033, 2017.

Zhang, G. Q., Luo, W., Chen, W. F., and Zheng, G. X.: 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, G. Q., Ran, Y. H., Wan, W., Luo, W., Chen, W. F., Xu, F. N., and Li, X.: 100 years of lake evolution over the Qinghai–Tibet Plateau, Earth Syst. Sci. Data, 13, 3951–3966, https://doi.org/10.5194/essd-13-3951-2021, 2021a.

Zhang, J., Hu, Q. W., Li, Y. K., Li, H. D., and Li, J. Y.: Area, lake-level and volume variations of typical lakes on the Tibetan Plateau and their response to climate change, 1972–2019, Geo-spatial Inf. Sci., 24, 458–473, https://doi.org/10.1080/10095020.2021.1940318, 2021b.

Zhang, Q., Shen, Z. X., Pokhrel, Y., Farinotti, D., Singh, V. P., Xu, C.-Y., Wu, W. H., 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. H., Zeraatpisheh, M., Rahman, M. M., Wang, S. J., 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, 2021c.

Zhang, Y., Zhang, G. Q., and Zhu, T. T.: Seasonal cycles of lakes on the Tibetan Plateau detected by Sentinel-1 SAR data, Sci. Total Environ., 703, 135563, https://doi.org/10.1016/j.scitotenv.2019.135563, 2020.

Zhao, R., Fu, P., Zhou, Y., Xiao, X. M., Grebby, S., Zhang, G. Q., and Dong, J. W.: Annual 30-m big Lake Maps of the Tibetan Plateau in 1991–2018, Sci. Data, 9, 164, https://doi.org/10.1038/s41597-022-01275-9, 2022.

Zhao, S. Y., Zhao, X., Zhao, J. C., Zhang, X., Liu, X. Y., and Yao, C. Z.: Monthly lake area changes larger than 10 km^2 on the Tibetan Plateau (2000–2024), Zenodo [data set], https://doi.org/10.5281/zenodo.19205532, 2026.

Zhao, Z. L., Zhang, Y., Hu, Z. Z., and Nie, X. H.: Contrasting Evolution Patterns of Endorheic and Exorheic Lakes on the Central Tibetan Plateau and Climate Cause Analysis during 1988–2017, Water, 13, https://doi.org/10.3390/w13141962, 2021.

Zhou, Y., Liu, B. L., Cui, Y. P., Wang, X. X., Cao, M. M., Zhang, S., Xiao, X. M., and Dong, J. W.: Annual improved maps to understand the complete evolution of 9 thousand lakes on the Tibetan plateau in 1991–2023, ISPRS J. Photogramm. Remote Sens., 217, 134–148, https://doi.org/10.1016/j.isprsjprs.2024.08.012, 2024.

Zhou, Y., Liu, B. L., Cui, Y. P., and Dong, J. W.: 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.

Zhu, L. P., Ju, J. T., Qiao, B. J., Liu, C., Wang, J. B., Yang, R. M., Ma, Q. F., Guo, L. N., and Pang, S. Y.: 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.