https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

Nonlinear hydro-climatic controls on an arid-region lake: Evidence from 40 years of remote sensing

Rui Zou[a], Xiaojun Wang[a,b*], Jianyun Zhang[a,b], Wentai Pang[c], Jianfeng Liu[d]

[a] State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing 210029, China

[b] Research Center for Climate Change, Ministry of Water Resources, Nanjing 210029, China

[c] Academy of Science and Technology of Inner Mongolia, Huhhot 010000, China

[d] Water Resources Research Institute of Inner Mongolia, Huhhot 010020, China

* **Corresponding Author (Xiaojun Wang) Email**: xjwang@nhri.cn.

Abstract: Accurate measurement of lake surface area is essential for understanding eco-hydrological processes in arid regions, yet long-term records are often limited by cloud contamination, seasonal ice cover, and data gaps. In this study, we developed an optimized extraction framework that integrates seasonal index selection, adaptive thresholding, maximum connectivity analysis, and mutual information－based gap filling to construct a continuous monthly lake area series for Bahannao Lake from 1984 to 2024. This method effectively addressed common challenges in remote sensing water extraction and provided reliable long-term lake dynamics in a data-scarce desert region. Based on the reconstructed time series, we examined the multi-factor drivers of lake evolution using an XGBoost model combined with climatic and energy-balance variables. Results reveal pronounced interannual and seasonal variability: precipitation dominates lake expansion in spring and summer, while shortwave radiation is the main driver of evaporation in autumn and winter, even under cold conditions. Long-term trends indicate a shift in controlling mechanisms—from humidity and precipitation decline (1984－1999), to increased radiation and humidity variability (2000－2014), and finally to intensified sensible heat flux and potential evapotranspiration (2015－2024).Our findings highlight the nonlinear and evolving interactions between hydro-climatic factors regulating arid-region lakes. The proposed framework provides a robust approach for generating long-term lake records, advancing understanding of eco-hydrological responses to climate change, and offering scientific support for water resources management and adaptation in arid regions.

Keywords: R remote sensing, lake area extraction, XGBoost, arid region, hydrology,

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

climate change

1 Introduction

Over the past century, with the intensification of global climate change and the increasing human ability to modify nature, the impact of climate change on lake systems and the surrounding water environment has become more pronounced. The formation and disappearance, expansion and contraction of lakes, as well as changes in water and ecological environments, are the result of interactions among global, regional, and local tectonic activities, climate events, and human activities. Within these systems, a series of complex interactions drive the evolution of lake systems (Ma et al., 2020).

Lakes are vital natural resources that are highly sensitive to climate change (Adrian et al., 2009; Schmid et al., 2014). Globally, there are over 100 million lakes, which store 87% of the Earth's liquid surface freshwater. Climate change is one of the most severe threats to global lake ecosystems. As observed in recent decades, lake surface conditions—such as ice cover, surface temperature, evaporation, and water levels—have responded significantly to this threat (Woolway et al., 2020; Tong et al., 2023). Approximately 53% of the world's lakes have experienced a decline in water storage, with a reduction of about 22 billion tons per year. Climate change and human water use have primarily driven the net decrease in water volume in approximately 100 large natural lakes worldwide. Lakes in both arid and humid regions are experiencing water loss, with drying trends being more widespread than previously understood.

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

Despite the shrinking of most lakes globally, 24% of lakes and reservoirs have shown a significant increase in water storage. These lakes and reservoirs are mostly located in sparsely populated regions, such as the Tibetan Plateau and the northern Great Plains of North America, as well as areas with newly constructed reservoirs, including the Yangtze River, Mekong River, and Nile River basins (Pickens et al., 2020).

China has a vast territory with an extensive network of rivers and lakes. There are 2693 lakes with an area greater than 1 km^2, among which 2557 lakes (95% of the total) have an area between 1 and 100 km^2. Additionally, there are 10 exceptionally large lakes with an area exceeding 1000 km^2. The total lake area in China has shown a significant increasing trend, expanding by approximately 7858.53 km^2 (11.41%) over the past 30 years (Ma et al., 2010; Ma et al.,2011). However, the spatial and temporal imbalance of water resources has intensified, with notable differences in trends across various lake regions. The lake areas in the Tibetan Plateau and Xinjiang regions have increased significantly, contributing 111.55% and 28.41% of the national lake area growth, respectively. In contrast, the lake areas in the Eastern Plain, Inner Mongolia Plateau, Northeast Plain and Mountainous Region, and Yunnan-Guizhou Plateau have declined significantly, with reductions of 24.53%, 9.30%, 6.06%, and 0.54%, respectively. Among these, the Mongolian-Xinjiang Plateau experienced the largest decline in lake numbers, with a loss of 111 lakes. Some lakes in this region have shown signs of shrinkage and salinization (Yang et al., 2010).

Scientists have discovered that the abrupt change timing of river and lake systems varies significantly across different latitudes and altitudes (Råman Vinnå., 2021; Zhou

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

et al., 2021). Mountain and polar lakes tend to experience abrupt changes earlier than temperate and tropical river-lake systems (Jeppesen et al, 2014). Additionally, under varying levels of human impact, the timing of abrupt changes in lakes also differs. Lakes in regions with low human impact generally experience abrupt changes earlier than those in areas with strong human influence (Preston et al., 2016). Analysis of the driving factors of lake abrupt changes indicates that the causes vary. Before the 1950s, climate change was the primary factor controlling abrupt changes in lake ecosystems. However, after the 1950s, both climate change and human disturbances became dominant factors. In temperate and tropical regions with strong human influence, lake changes are mainly driven by nutrient enrichment and pollution. In contrast, lakes located in high-altitude and high-latitude regions, which are less affected by human activities, are more vulnerable to climate change. Furthermore, the interaction of multiple drivers increases the likelihood of abrupt changes in lakes, with climate change being the most frequently interacting factor leading to transformations in river-lake ecosystems (Vincent et al., 2009.). Li et al. (2025) pointed out that seasonality is the dominant driver of lake-surface-extent variations globally

For example, Plug et al.(2008) investigated lake area changes in the Tuktoyaktuk Peninsula in northwest Canada. They found that from 1978 to 1992, the total lake area increased, while from 1992 to 2001, the total lake area decreased. Their study identified precipitation as the main factor driving these changes. Similarly, Carroll et al. (2011) studied the lake area changes in high-latitude northern Canada and discovered that lake areas showed a significant decline, exhibiting regional clustering characteristics, with

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

climate factors driving these changes. Labazhuoma et al. (2017) explored the expansion of Tangra Yum Co from 1977 to 2014. Their results indicated that, under the background of climate warming, the combined effects of glacier melt, precipitation increase, and evaporation changes contributed to the lake's expansion. Likewise, Li Meng et al. (2017) examined the changes in the water surface area and water storage of Nam Co from 1976 to 2015. Their findings showed that the water surface area and water storage of Nam Co continued to increase, with the fastest growth in water storage occurring between 1997 and 2009. The study concluded that the primary factor driving the increase in Nam Co's water volume was glacier melt, followed by increased precipitation and reduced evaporation.

However, the precise measurement of lake area remains a major constraint for analyzing lake changes. With advancements in science and technology, remote sensing has provided a unique and effective method for monitoring the spatiotemporal variations in surface water areas on broad geographic scales (Liu et al., 2020).

Currently, water extraction methods using optical sensors have been widely applied28-30 (McFeeters 1996; Yao et al., 2015; Donchyts et al., 2016). However, existing water body area products often fail to meet ideal spatial or temporal resolution requirements31-32 (Cooley et al., 2017; Huang et al., 2018). For example, the 2016 Global Climate Observing System (GCOS) Implementation Plan recommended a resolution of 20 meters and a daily monitoring frequency (Secretariat, 2009). High-temporal-resolution sensors, such as the Moderate Resolution Imaging Spectroradiometer (MODIS) onboard Terra and Aqua satellites, have been used to

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

assess water body areas at time scales ranging from daily to 16-day intervals (Bergé-Nguyen et al., 2015; Wang et al., 2018). However, many small water bodies (e.g., 10–50 km^2 or smaller) and irregularly shaped larger water bodies may not be accurately distinguished using coarse-resolution MODIS images (250–500 meters in the visible and near-infrared bands) (Tao et al., 2015). Compared with MODIS, Landsat images (e.g., Landsat 5 Thematic Mapper (TM), Landsat 7 Enhanced Thematic Mapper Plus (ETM+), and Landsat 8 Operational Land Imager (OLI)) offer higher spatial resolution (30 meters) and a temporal resolution of 16 days (or better when combining multiple Landsat sensors). However, due to cloud contamination (Rossow et al., 1999), the actual temporal frequency of water body mapping based on Landsat is often much lower than the nominal resolution and may extend to a year for lakes with persistent ice cover (Yao et al., 2018). The recently launched Sentinel-2A and 2B satellites, equipped with Multispectral Instruments (MSI), provide a resolution of 10 meters in the visible and near-infrared bands, with a revisit period of 5–10 days. However, their observations currently cover only the past few years (since 2015) and are not yet suitable for long-term decadal monitoring.

Beyond the trade-offs between spatial and temporal resolution, several other factors challenge high-resolution monitoring of long-term global surface water area changes (Klein et al., 2017). These include the inherent spectral heterogeneity of water, atmospheric influences (clouds and aerosols), topographic shadows, aquatic vegetation, and spectral contamination from ice/snow cover. In such complex conditions, integrating multiple techniques is often necessary to achieve robust water body

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

extraction.

Recently, Pekel et al. (Pekel et al., 2014) utilized a large training dataset, combined with expert systems and visual analysis, to identify the presence or absence of water on a monthly basis for each pixel in archival Landsat images from 1984 to 2015. This product was named the Joint Research Centre (JRC) Global Surface Water dataset (hereinafter referred to as GSW). Despite its significant achievements, GSW is based on cloud-free pixels, meaning that the mapped extent of specific water bodies is only complete when monthly composite images have minimal cloud cover. A follow-up study by Busker et al. (Buske et al., 2019) used a subset of the GSW dataset, selecting images with cloud cover below 5%, to extract the monthly area of 137 lakes/reservoirs. For nearly half of these lakes/reservoirs, the correlation between area and radar altimetry-measured water levels exceeded 0.8. However, the temporal frequency of the resulting area time series was still constrained by the availability of cloud-free images, and due to the current availability of GSW, the time series was interrupted after October 2015.One potential method to increase the temporal frequency of lake mapping based on Landsat data is to estimate water surface area from contaminated images (e.g., those affected by clouds or observation gaps). Although these images are of relatively lower quality, the exposed portions of lakes within them may provide useful information for inferring the complete extent. For instance, Zhao and Gao (2018) 41 applied the monthly water mapping data from the GSW dataset to generate area time series for 6,817 reservoirs worldwide from 1984 to 2015. Their method involved recovering complete reservoir extents from cloud-contaminated images by segmenting pixels

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

based on the water occurrence probability provided in the GSW dataset. Compared to the results of Busker et al., their generated area time series increased the number of observations by approximately 80%. However, the reliance on the existing GSW dataset restricted their reservoir area records to the 1984–2015 period, and the validation of their recovery method was limited to only nine reservoirs with significant water level variations.

Bahanao Lake is located in a semi-arid region and has long lacked systematic observational data. There are no complete records of its surface area, yet its changes are crucial to the stability of the regional ecosystem. With the intensification of global climate change, the lake's surface area has significantly shrunk, experiencing multiple abrupt shifts and exhibiting a continuous declining trend. The driving mechanisms behind these changes are complex and diverse. However, its dynamic variation characteristics and driving forces remain insufficiently studied.

Despite substantial progress in global lake monitoring, significant gaps remain for lakes in arid regions. First, long-term and continuous lake area records are scarce, as existing products are often interrupted by cloud contamination, seasonal ice cover, and striping artifacts. Second, the role of hydro-climatic drivers in regulating lake dynamics remains insufficiently understood, particularly regarding nonlinear interactions between precipitation, humidity, radiation, and energy fluxes in arid environments. Third, current extraction methods lack robustness across different seasons and fail to ensure accuracy in data-scarce regions.

To address these challenges, this study develops an optimized lake area extraction

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

framework that integrates seasonal index selection, adaptive thresholding, connectivity analysis, and mutual information - based gap filling to construct a continuous monthly record of Bahannao Lake from 1984 to 2024. By coupling the reconstructed time series with multi-factor analysis using the XGBoost model, we reveal the temporal shifts and nonlinear controls of hydro-climatic drivers on lake dynamics. This framework not only improves the reliability of long-term lake monitoring under complex conditions but also advances understanding of eco-hydrological responses to climate change and provides implications for water resource management in arid regions.

2 Data and Methods

2.1 Dataset Selection

This study utilizes remote sensing imagery from the Landsat 5 TM, Landsat 7 TM, and Landsat 8 OLI sensors, specifically using atmospherically corrected reflectance data (Tier 1 TOA Reflectance). Tier 1 data is selected due to its highest quality, making it suitable for time-series analysis and studies on global surface water extent and dynamics. The Landsat 5 TM imagery covers the period from 1984 to 2011, while Landsat 8 imagery spans from 2013 to 2023. Since imagery for 2012 is missing in both datasets, Landsat 7 TM is used as a supplement. However, Landsat 7 TM imagery exhibits significant striping artifacts, which were avoided as much as possible during data selection.

For meteorological data, this study employs the fifth-generation atmospheric reanalysis dataset from ECMWF (European Centre for Medium-Range Weather Forecasts), covering global climate data from January 1950 to the present. The dataset has a temporal resolution of daily and a spatial resolution of 0.1°×0.1°.

2.2 Methods

(1) Optimized Lake Area Extraction Method

This study employs 30-meter full-atmosphere imagery from the Landsat 5 Thematic Mapper (TM), Landsat 7 Enhanced Thematic Mapper Plus (ETM+), and Landsat 8 Operational Land Imager (OLI) satellites to derive monthly lake area estimates for the study region from January 1984 to December 2024.

Different lake remote sensing indices were selected for non-freezing and freezing periods, respectively. For non-freezing periods, remote sensing indices were processed to remove cloud and snow interference. Images were filtered based on cloud cover percentage (C), and monthly composite images were generated. The Otsu thresholding method was then applied to automatically determine segmentation thresholds. To distinguish between lakes and mountainous areas, a digital elevation model (DEM) was used, setting the slope (θ) and aspect (ϕ) thresholds to 0.

Considering that most lakes exhibit connectivity, this study adopts the maximum connected component analysis algorithm from the OpenCV computer vision library to delineate lake boundaries. Images were categorized based on cloud cover information ('CLOUD_COVER'): those with cloud cover ≤30% were classified as cloud-free images, while the remaining images were considered cloudy. For cloudy images, the MI (Mutual Information) algorithm was used to match them with the most similar cloud-free images. The most similar image was then merged with the original cloudy image to generate a filled version.

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

For images with striping artifacts, the same filling method was applied as for cloudy images. Clear lake boundaries from historical cloud-free images were used, and the MI algorithm was employed to find the most similar historical cloud-free images for filling missing water pixels in striped areas, ultimately obtaining the final lake water extent. The specific process is shown in Figure 1.

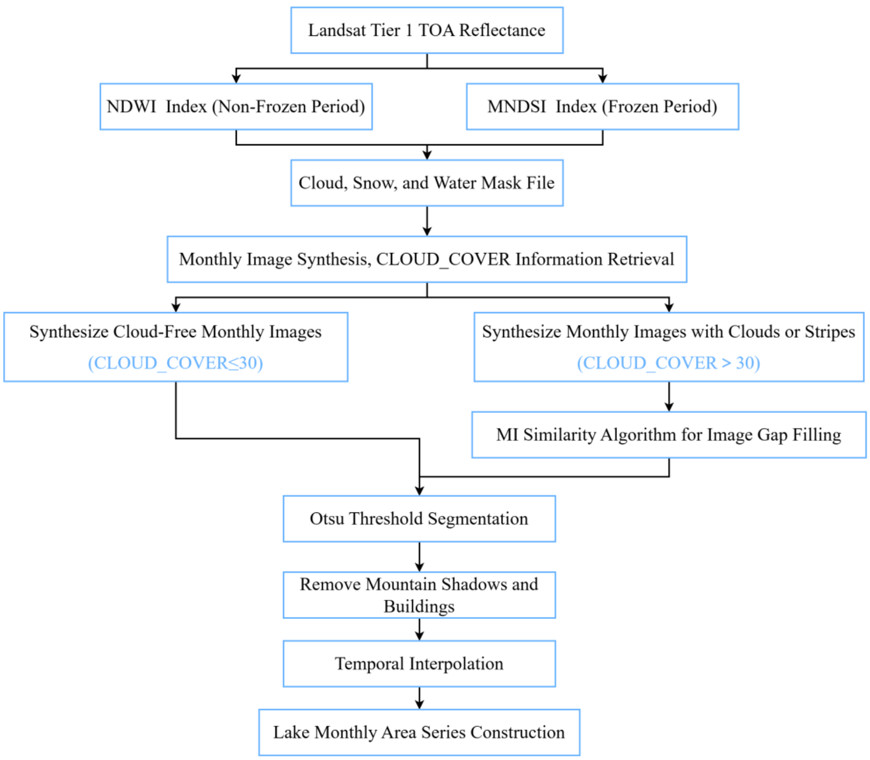


Figure 1. Flowchart of Lake Area Extraction Process.

(2) XGBoost Model

The objective function of the XGBoost model is:

$$L(\theta) = \sum_{i=1}^{n} l(y_i, f(x_i)) + \sum_{i=1}^{n} \Omega(f_k)$$

Where $L(\theta)$ represents the objective function, which measures the model's performance in prediction and consists of two parts: $l(y_i, f(x_i))$ is the loss function, indicating the difference between the true value y_i and the predicted value $f(x_i)$, while $\Omega(f_k)$ is the regularization term used to control the model complexity.

The input factors $x_i = \{x_1, x_2, ..., x_n\}$ include various environmental variables such as temperature, precipitation, humidity, and radiation.

$$FI(x_j) = \frac{1}{T}\sum_{t=1}^{T} I(t, x_j)$$

Here, $FI(x_j)$ represents the feature importance of factor x_j, while $I(t, x_j)$ denotes the contribution of factor x_j when used as a splitting point in tree t, with T being the total number of trees. The generated feature importance ranking chart illustrates the contribution of various input factors (such as temperature, precipitation, and humidity) to lake area changes. This ranking chart provides an intuitive way to identify the most influential factors.

To improve model performance, hyperparameters can be optimized using Grid Search or Random Search. Common hyperparameters include Learning rate, Max depth of trees and Number of trees. Adjusting these parameters affects the model's fitting ability and generalization performance.

Data Splitting: Divide the dataset into a training set and a test set (e.g., 80% for training, 20% for testing).

Train the XGBoost model on the training set. XGBoost uses the Gradient Boosting Algorithm, which iteratively improves the model by building multiple weak learners to reduce prediction errors. Each iteration refines the model by fitting the residuals (i.e.,

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

prediction errors).

Model Validation: Evaluate model performance using metrics such as Mean Squared Error (MSE) and Coefficient of Determination (R^2) to assess accuracy and stability.

The formula for Mean Squared Error (MSE) is:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2$$

The formula for the coefficient of determination R^2 is:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - f(x_i))^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

Where $\overline{y}$ represents the mean of the samples.

The lake area model is based on model training, the predicted lake area $\hat{y}$ can be expressed as a nonlinear combination of input factors x_i:

$$\hat{y} = f(x_i) = \sum_{k=1}^{K}\omega_k h_k(x_i)$$

Where: ω_k is the weight of the k tree, and $h_k(x_i)$ is the prediction function of the tree, represented as a set of decision rules.

3. Lake Area Time Series Construction

3.1 Remote Sensing Interpretation and Monthly Lake Image Synthesis

(1) Selection of Remote Sensing Indices

The study area is located in a high-altitude region, where lake surfaces freeze between November and March. Since the NDWI index is less effective for frozen lakes, different indices are used for different seasons. During the non-freezing period (May–

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

November), the NDWI index is applied for conventional water body extraction. During the freezing period (December–April), the Modified Normalized Difference Snow Index (MNDSI) is used to evaluate water surface area.

The NDWI index utilizes the strong absorption of water bodies in the near-infrared band and their high reflectance in the green band to enhance the distinction between water and other land cover types. However, this index may misidentify bright white buildings, clouds, snow, and mountain shadows as water bodies. Therefore, additional data quality bands and methods are integrated to remove these interferences and improve the accuracy of water body extraction.

$$NDWI = \frac{(Green - NIR)}{(Green + NIR)}$$

Where: Green band typically refers to the green portion of the visible spectrum, generally ranging from 500–570 nm. NIR band refers to the near-infrared spectrum, generally ranging from 800–900 nm.

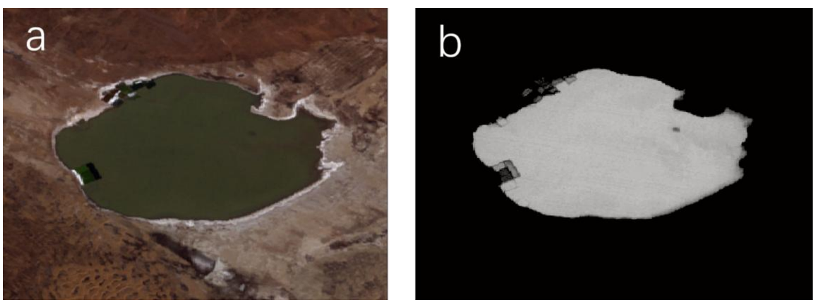


Figure 2 Original lake image during the non-freezing period (a) and NDWI-identified image (b). Source: Landsat imagery courtesy of the U.S. Geological Survey (USGS), processed and interpreted by the authors.

The Modified Normalized Difference Snow Index (MNDSI) is an index calculated

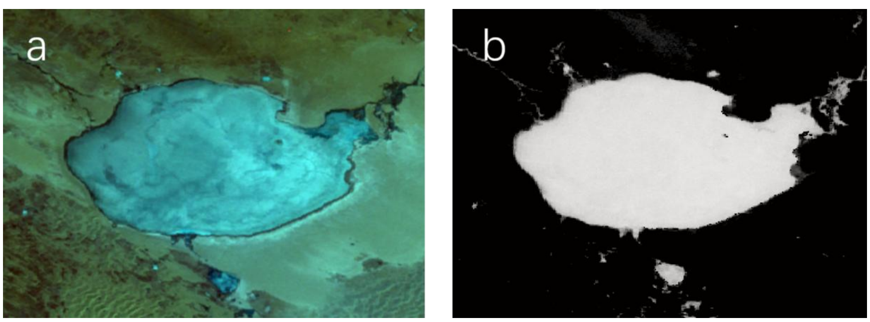
https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

using the reflectance of the near-infrared (NIR) and short-wave infrared (SWIR) bands. It is an effective method for distinguishing ice surfaces from water bodies. This index is particularly suitable for regions with frozen water surfaces, such as lakes and rivers, where seasonal changes are significant. Ice surfaces and water bodies have different reflectance characteristics in various bands. Ice has higher reflectance in the SWIR band, while water has lower reflectance. By calculating the difference between the NIR and SWIR bands, MNDSI can effectively distinguish between ice surfaces and water bodies, thus improving the accuracy of ice extraction. By combining these two bands, MNDSI highlights the differences between water bodies and ice surfaces, making it easier to differentiate between them. Similar to NDWI, MNDSI enhances the contrast between ice and water by utilizing reflectance values from different bands.

MNDSI (Modified Normalized Difference Snow Index) is calculated by combining the reflectance of the near-infrared (NIR) and short-wave infrared (SWIR) bands. The typical formula for MNDSI is as follows:

$$MNDSI = \frac{NIR - SWIR}{NIR + SWIR}$$

Where NIR is the reflectance in the near-infrared band (typically 800–900 nm), SWIR is the reflectance in the short-wave infrared band (typically 1500–1700 nm).



https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

Figure 3 Original image of the lake during the freezing period (a) and MNDSI recognition image (b). Source: Landsat imagery courtesy of the U.S. Geological Survey (USGS), processed and interpreted by the authors.

(3) Cloud and Snow Interference Removal

The cloud and snow interference removal is only applied to the NDWI of the non-freezing period from May to November. The Landsat series satellites provide their own pixel-scale data quality band (QA_PIXEL), which can be used to eliminate noise pixels in the image.

The QA_PIXEL band in the Landsat dataset provides information on various quality types, where different bits (Bit) correspond to different types of quality information. For example, Bit 3 corresponds to clouds, Bit 5 corresponds to snow, and Bit 7 corresponds to water bodies. Within the same bit, values of 0 and 1 represent different data qualities. For example, a 0 in Bit 7 indicates that the pixel has poor water body information, being land or covered by clouds, while a 1 indicates that the pixel represents water.

Using this pixel quality information, we selected Bit 3 (cloud), Bit 5 (snow), and Bit 7 (water body). By performing bitwise AND and OR operations, we generated a water body mask file with good data quality after cloud and snow removal. This mask file is then overlaid with the actual image to remove pixels affected by cloud or snow interference. The effect of cloud and snow removal is shown in the image below:

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

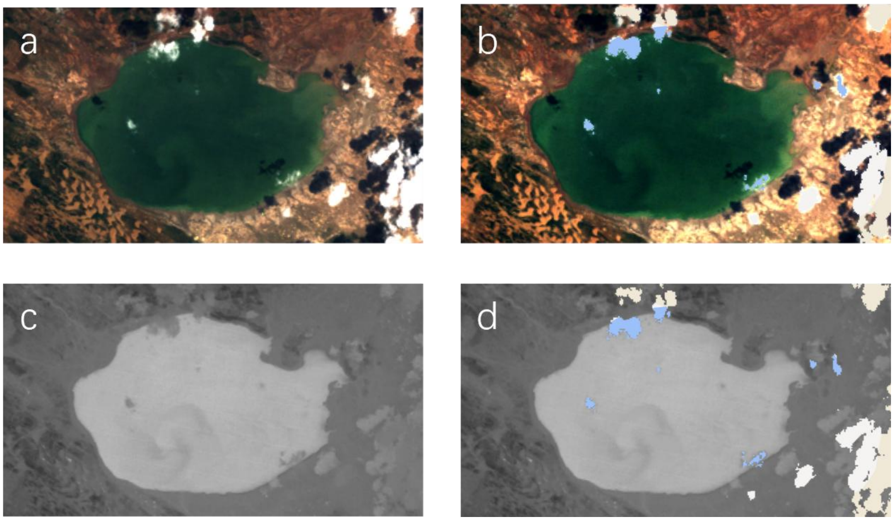


Figure 4: True color original image (a), true color image after cloud and snow removal (b), NDWI water body index calculation result of the true color original image (c), NDWI water body index calculation result of the true color image after cloud and snow removal (d). Source: Landsat imagery courtesy of the U.S. Geological Survey (USGS), processed and interpreted by the authors.

(4) Monthly Image Download

The NDWI, MNDSI index calculation, and cloud/snow interference removal are performed directly on the GEE platform, followed by monthly composite image downloads. Based on the cloud cover information ('CLOUD_COVER'), which represents the cloud amount (range from 0 to 100, with larger values indicating more cloud coverage), the data is classified into three levels: 0-30, 30-60, and 60-100. If data is available in Level 1, Level 2 is not executed, and if Level 2 contains data, Level 3 is processed. All images from each year and month within the cloud cover level are selected, and the median pixel value is calculated to generate the composite monthly

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

NDWI (for 5-11 months) and MNDSI (for December to the following April) grayscale images.

Data is filtered based on the cloud cover proportion C, where C ∈ [0,100].

$$\text{Composite image} = \text{Med}(\text{S}(\text{C})), \quad \text{where C} = \text{CLOUD}_{\text{COVER}}$$

$$S(C) = \begin{cases} I(C) & \text{if} \quad 0 \le C \le 30 \\ I(C) & \text{else if} \quad 30 < C < 60 \\ I(C) & \text{else} \quad 60 < C \le 100 \end{cases}$$

Where I(C) is a set of image data filtered by cloud cover.

(5) Threshold Segmentation

This step applies the Otsu threshold algorithm to the downloaded NDWI and MNDSI monthly composite grayscale images, automatically generating a segmentation threshold. Pixels below the threshold are classified as water, and those above the threshold are classified as other areas.

The core of the Otsu thresholding method is to divide the image into two classes (foreground and background) by maximizing the between-class variance, thereby achieving the optimal threshold segmentation. Specifically, it involves iterating through all possible thresholds, and the optimal threshold is determined when the between-class variance is maximized while the variance within both the foreground and background is minimized. Compared to other methods, this algorithm maximizes the inclusion of the target feature while excluding other interfering factors.

The Otsu thresholding method is used to automatically generate the segmentation threshold, dividing the image into water and other regions:

$$T = \arg\max \max_{\tau}(\sigma_B^2(\tau))$$

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

Where, $\sigma_B^2(\tau)$ is the between-class variance, defined as:

$$\sigma_B^2(\tau) = \omega_1(\tau)\omega_2(\tau)(\mu_1(\tau) - \mu_2(\tau))^2$$

Where $\omega_1(\tau)$ and $\omega_2(\tau)$ are the weights of the foreground and background at the threshold τ, and $\mu_1(\tau)$ and $\mu_2(\tau)$ are the mean gray values of the foreground and background, respectively.

The portion smaller than the threshold T is classified as water, symbolized as water pixels, while the portion greater than the threshold is classified as other categories.

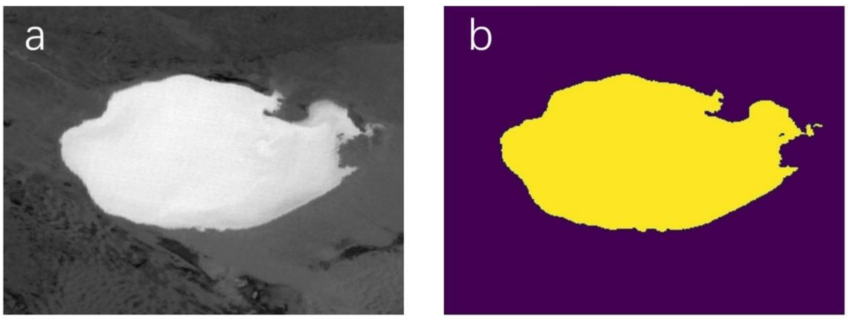


Figure 5 NDWI water index recognition result (a), and the effect of Otsu threshold method automatically separating water (yellow area) and background (purple area) based on NDWI recognition result (b). Source: Landsat imagery courtesy of the U.S. Geological Survey (USGS), processed and interpreted by the authors.

(6) Mountain Shadow Removal

Since the lake surface typically exhibits a flat state without significant slope and aspect features, digital elevation models (DEM) can be used to distinguish lakes from mountainous regions by utilizing slope and aspect information. By setting threshold values of 0 for slope and aspect, the distinction between lakes and mountainous areas can be made. However, the current frequency of elevation data updates does not align with real-time imagery, leading to an inability to accurately reflect seasonal changes in lake water levels within the elevation data. This limitation affects the precision of water body area extraction using the data. Given that most lakes are interconnected, this study employs the maximum connected component analysis algorithm from the Open-CV vision field to define the boundaries of lakes and extract their areas.

By setting the thresholds for slope θ and aspect φ to 0 in the digital elevation model (DEM), lakes are distinguished from mountainous areas:

$$\theta(x,y)=0,\quad \phi(x,y)=0$$

Where $\theta(x,y)$ and $\phi(x,y)$ represent the slope and aspect values at a given point (x,y), respectively. By setting $\theta(x,y)=0$ and $\phi(x,y)=0$ as threshold conditions for the lake area, the lake region is defined as the area where both the slope and aspect are equal to 0.

$$L=\max_i\Big(\sum_{(x,y\in C_i}I(x,y)\Big)$$

Where L represents the total number of pixels in the largest lake area, C_i represents the i-th connected component in the image, the function $\sum$ denotes the summation of pixel points, and $\max_i$ indicates the selection of the largest connected component as the lake area.

（7）Buildings Removal

The construction of the building index currently mainly relies on the fact that the surface temperature of buildings is usually higher than that of surrounding land cover, and the mid-infrared band can effectively reflect surface temperature differences.

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

However, in previous land cover classification studies, the extraction results using this algorithm were not ideal. Considering that most buildings in the study area are not distributed along lakes, the maximum connected component algorithm can effectively exclude parts where buildings are misidentified as water bodies.

Based on the NDWI (Normalized Difference Water Index), a threshold T is used to binarize the image, separating water bodies from non-water bodies.

$$I(x,y)=\begin{cases}1, \ if \ NDVI(x,y)>T \\ 0, \ if \ NDVI(x,y)\le T\end{cases}$$

Connected Component Calculation: In the binarized image, the Connected Components Labeling (CCL) algorithm is used to identify all connected regions. A connected component is determined by scanning the neighboring pixels in the image (up, down, left, right, or diagonally). The formula is expressed as:

$$C_i=\sum_{(x,y)\in R} I(x,y)$$

Where R represents the connected regions in the image, and C_i denotes the connected components.

To eliminate interference from buildings, a threshold condition τ is set, retaining only connected components with an area greater than τ. Since buildings typically have smaller areas, while lakes exhibit larger connected components, the lake regions can be filtered using the following condition:

$$C_i>\tau$$

The lake boundary is extracted using a boundary detection algorithm (e.g., the Canny edge detection algorithm) applied to the selected largest connected region.

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

$$B = Canny(C_i)$$

As shown in the figure 6, the white areas in the original image include both lakes and buildings. When using threshold segmentation to extract water bodies, buildings may also be mistakenly identified as water. By applying the maximum connected component method, buildings can be effectively separated.

Figure 6: The white areas in the original image include both lakes and buildings (a), water bodies extracted using threshold segmentation (b), and buildings separated using the maximum connected component method (c). Source: Landsat imagery courtesy of the U.S. Geological Survey (USGS), processed and interpreted by the authors.

(8) Cloudy Image Filling Processing

The processing steps (5)–(7) are applied to cloud-free images. For cloudy images, cloud-free images are used for filling before executing steps (5)–(7).

The filling approach is as follows: Based on the cloud coverage information (CLOUD_COVER), images with cloud cover less than or equal to 30% (CLOUD_COVER ≤ 30%) are classified as cloud-free images, while others are considered cloudy images. The formula is as follows:

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

$$\text{Cloudy Image} = \{\text{Image} \mid CLOUD_{COVER} \le 30\%\}$$
$$\text{Cloud-Free Image} = \{\text{Image} \mid CLOUD_{COVER} > 30\%\}$$

Then, the Mutual Information (MI) algorithm is used to perform the most similar matching between the cloudy image and all cloud-free images. Next, the most similar image is combined with the original cloudy image through a union operation to obtain the filled cloudy image. Finally, steps (5)-(7) from the cloud-free image processing are executed, resulting in the final water body area. The specific steps are as follows:

Candidate Cloud-Free Image Set: In the time periods before and after the cloudy image, select images with low cloud coverage (CLOUD_COVER ≤ 30%) as the candidate image set.

Mutual Information Algorithm: Use the MI algorithm to calculate the similarity between the cloudy image and the candidate cloud-free images. The formula is as follows:

$$I(I_{cloudy}, I_{clear}) = \sum_{i,j} p(I_{cloudy} = i, I_{clear} = j) \log\left(\frac{p(I_{cloudy} = i, I_{clear} = j)}{p(I_{cloudy} = i), p(I_{clear} = j)}\right)$$

Where I_{cloudy} represents the cloudy image, I_{clear} represents the candidate cloud-free image, and p is the joint probability distribution of the pixel grayscale values. III denotes mutual information, which measures the correlation between the cloudy image and the cloud-free image.

Selecting the Most Similar Image: Based on the mutual information value, the cloud-free image most similar to the cloudy image is selected.

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

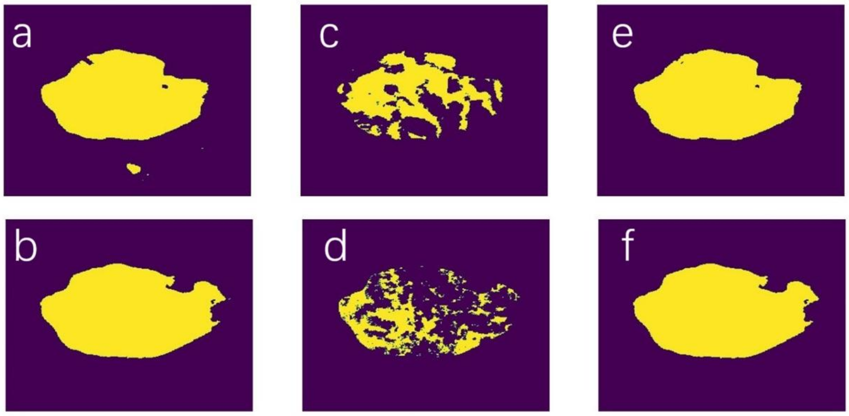


Figure 7: Image a and b show the cloud-free images found to be most similar to the cloudy images images c and d are the cloudy images, and images e and f show the result after cloud-filled processing of the cloudy images. Source: Landsat imagery courtesy of the U.S. Geological Survey (USGS), processed and interpreted by the authors.

(9) Striped Image Filling

The previously mentioned dataset indicates that Landsat 7 TM images have significant striping interference. Additionally, Landsat 5 TM and Landsat 8 OLI images also experience striping interference in certain months, such as Landsat 5 TM from 2001 to 2003 and Landsat 8 in 2008. To more accurately obtain the temporal changes in lake area, it is necessary to fill the missing portions of striped images. The method is the same as for cloud-filled images. By utilizing the clear contours of historical cloud-free images and applying the MI algorithm, the most similar historical cloud-free images are searched to fill the water pixels in the striped regions. The method for filling striped images is the same as that for cloud-filled images. Afterward, steps (5)-(7) are

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

executed to obtain the final water area extent.

Figure 8: The image found from the cloud-free images that is most similar to the striped interference image (a); The striped interference image(b); The effect after filling the striped interference(c). Source: Landsat imagery courtesy of the U.S. Geological Survey (USGS), processed and interpreted by the authors.

(10) Water Area Extraction

After applying the maximum connectivity component processing to the image, the number of water pixels is counted. Then, based on the spatial resolution of the pixels (30m * 30m), the actual area is calculated.

(11) Interpolation Processing

Collect all known lake area data for specific time points, where t_i dots represent time points with available data. For each missing data point t_{missing}, use the known data points $t_{\text{missing}-1}$ and $t_{\text{missing}+1}$, and apply the selected interpolation method to calculate the lake area $A(t_{\text{missing}})$ at time.

3.2 Lake Area Time Series Construction

The interannual variation of Bahannao is quite drastic, but the overall trend is declining, though not significantly. Before 1999, the changes were relatively stable. In 2000, the lake area shrank severely, decreasing by 82.98% compared to 1999, leaving only 3.12 km². Since then, the lake has exhibited a cyclical fluctuation pattern with a

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

period of approximately 5–6 years. In 2021, the lake area reached its minimum value of just 0.71 km², followed by a rapid increase, reaching its maximum of 23.38 km² in 2023.

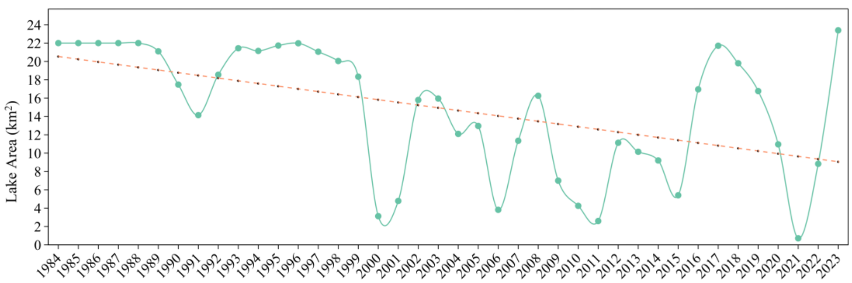


Figure 9 Interannual Variation of Lake Area

Due to its location in the Mu Us Desert and the lack of long-term observational data, this study references the lake area interpreted via remote sensing in the *Comprehensive Lake Water Ecological Management Plan of Uxin Banner*. This report provides remote sensing imagery data for 24 years from 1988 to 2018 (with six years lacking clear images suitable for analysis).

A comparison of the data (Figure 10) shows that the lake area interpreted in this study aligns with the trend reported in the management plan. Over the 23 years of overlapping interpretation, the error remains within 15% for 12 years. However, in years when the lake area was smaller, the error was relatively larger, such as in 2000, 2001, 2009, 2010, 2011, and 2015. According to records, Bahannao Lake shrank significantly during these years but did not completely dry up until 2021, which is consistent with the results of this study.

The interpreted lake area in this study also indicates (Figure 11) that the annual average area of Bahannao Lake in 2021 was only 0.71 km². The lake area was at its smallest in August, September, and October, reaching only 0.2 km², while the largest area was recorded in March at 3.5 km².

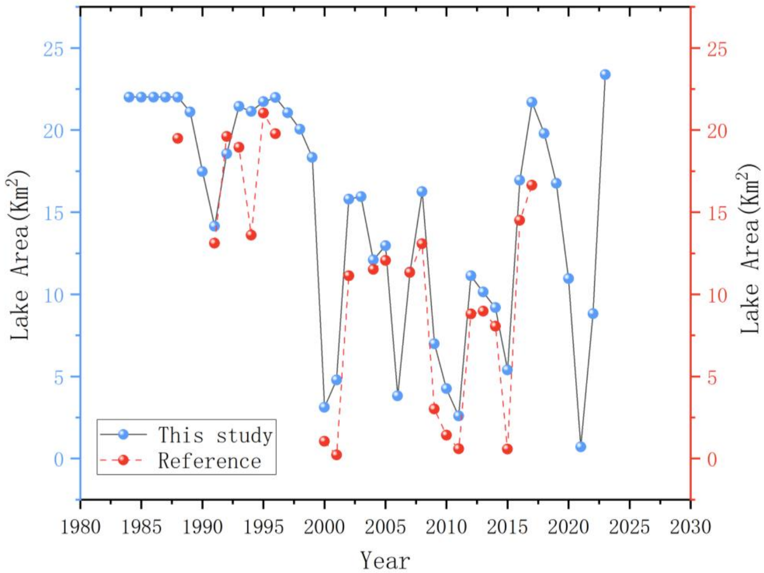
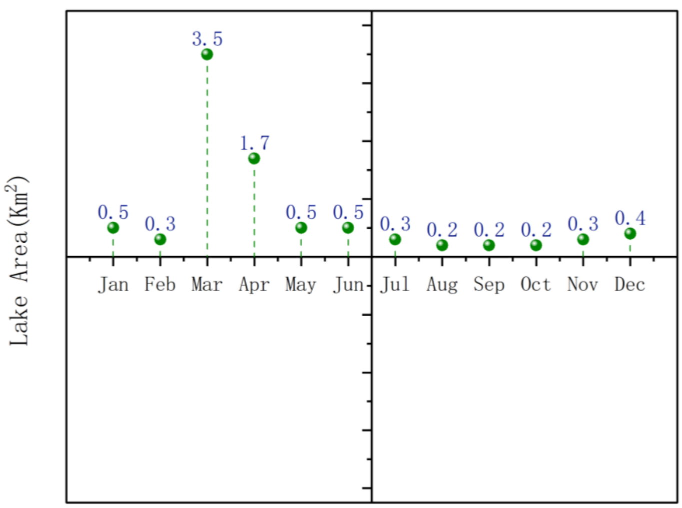


Figure 10 Comparison of the lake area interpreted in this study with the reference data



https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

Figure 11 Monthly and seasonal variation of lake area in 2021

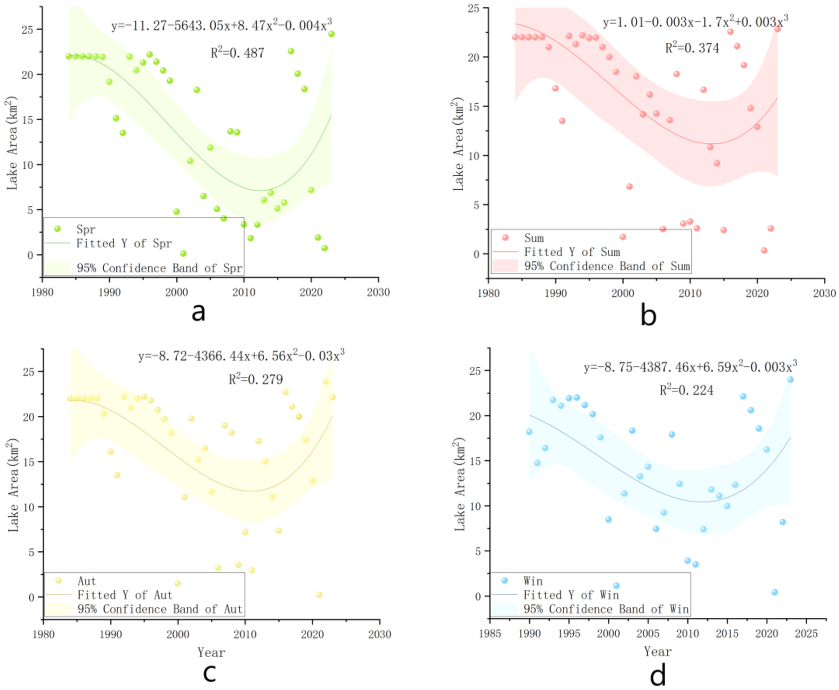


Figure 12 Seasonal Variation of Lake Area

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

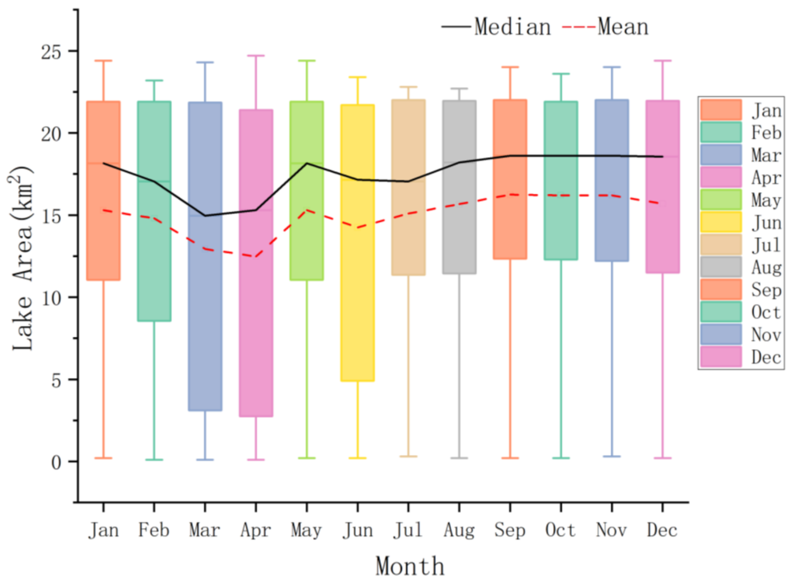


Figure 13 Monthly Variation of Lake Area

From the perspective of seasonal and monthly variation characteristics, Bahannao exhibits significant seasonal differences. The lake area in summer, autumn, and winter is noticeably larger than in spring, with autumn having the largest lake area, averaging 16.21 square kilometers and reaching a peak of 16.24 square kilometers in September. In contrast, spring has the smallest lake area, averaging only 13.57 square kilometers, with the lowest value of 12.48 square kilometers occurring in April.

3.3 Impact of Climate Change

(1) Temperature Variation

1) Temperature

The rise in air temperature directly affects the evaporation rate of the lake. The warming rate is 0.0429°C per year, leading to an increase in the lake surface temperature and, consequently, higher evaporation. High temperatures intensify water

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

evaporation, reducing the lake's water volume and causing a gradual decrease in lake area over the years.

The increase in air temperature enhances heat input into the water body, accelerating evaporation. As more heat is absorbed, surface water transforms more easily into water vapor, leading to a decline in lake water levels. Although the influence of temperature on lake area varies across different time periods, its continuous upward trend has a long-term impact on the reduction of lake area.

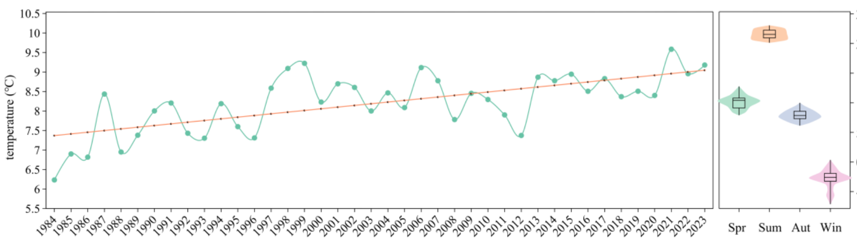


Figure 14 Regional Air Temperature Variation

2) 2m Dew Point Temperature

The 2m dew point temperature increases at a rate of 0.0095°C/a, indicating changes in atmospheric humidity. A rising dew point temperature suggests an increase in water vapor content in the air, typically associated with higher humidity. However, humidity changes do not always directly impact lake area; instead, they influence lake water volume indirectly by affecting evaporation and precipitation. While an increase in dew point temperature usually indicates higher humidity, if precipitation is insufficient or evaporation rates are too high, this increase in humidity may not effectively replenish lake water. Instead, it could contribute to lake shrinkage. The varying influence of the 2m dew point temperature over different periods suggests a

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

complex relationship with lake area changes, requiring a comprehensive analysis alongside other climatic factors.

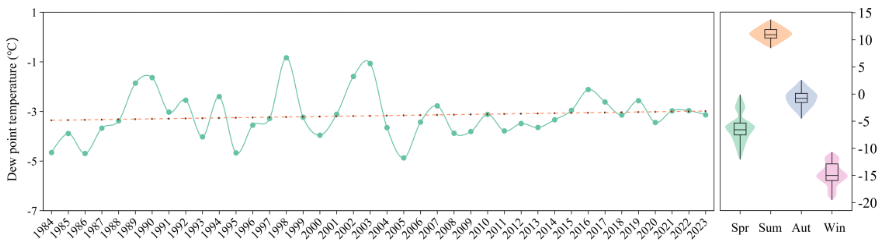


Figure 15 Regional 2m Dew Point Temperature Changes

(2) Changes in Precipitation

The total precipitation is decreasing at a rate of 1.7355 mm per year. Precipitation is one of the primary sources of lake water. A reduction in precipitation leads to insufficient water replenishment for the lake, resulting in a decline in water levels and a reduction in lake area.

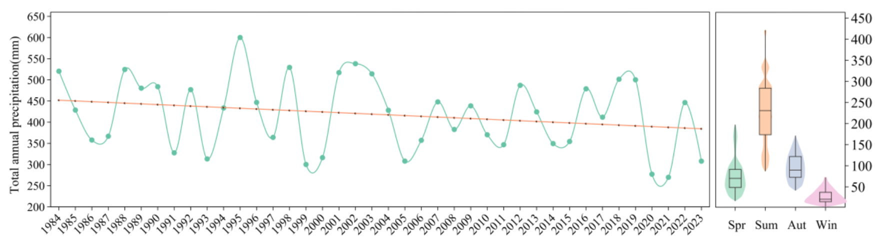


Figure 16 Regional Precipitation Changes

(3) Radiation and Energy Exchange

1） Net Longwave Radiation at the Surface

Net longwave radiation at the surface decreases by 0.0843 W/m² per year. The reduction in longwave radiation means that the lake receives less radiative heat, which theoretically could reduce evaporation. However, this effect is overshadowed by other

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

factors such as reduced precipitation and rising temperatures. While the decrease in longwave radiation could reduce heat loss from the lake, in conditions of drought and high evaporation, the impact of this reduction is likely limited.

2) Net Shortwave Radiation at the Surface

Net shortwave radiation at the surface increases by 0.0653 W/m² per year. The increase in shortwave radiation enhances the evaporation process, thereby reducing the lake's surface area. The rise in shortwave radiation leads to an increase in surface temperature, which accelerates evaporation. The intensified evaporation exacerbates the loss of water from the lake. The effect of increased shortwave radiation on the lake's area is significant during all periods, especially under drought and high-temperature conditions, where its impact is particularly pronounced.

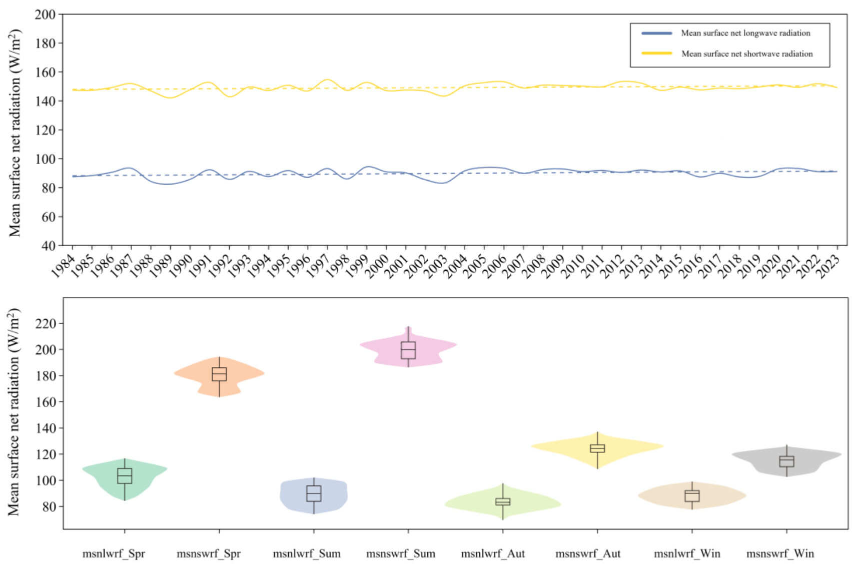


Figure 17 Regional Mean Surface Net Radiation

3)Mean Surface Latent Heat Flux

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

he latent heat flux decreases at a rate of 0.1343 W/m² per year. The decrease in latent heat flux indicates a reduction in the moisture carried by the air, possibly as a result of decreased humidity, which further intensifies evaporation from the water.

4)Mean Surface Sensible Heat Flux

The sensible heat flux increases by 0.0693 W/m² per year, meaning that the heat exchange between the surface and the atmosphere is enhanced. This leads to more evaporation, particularly during the summer when temperatures are higher.

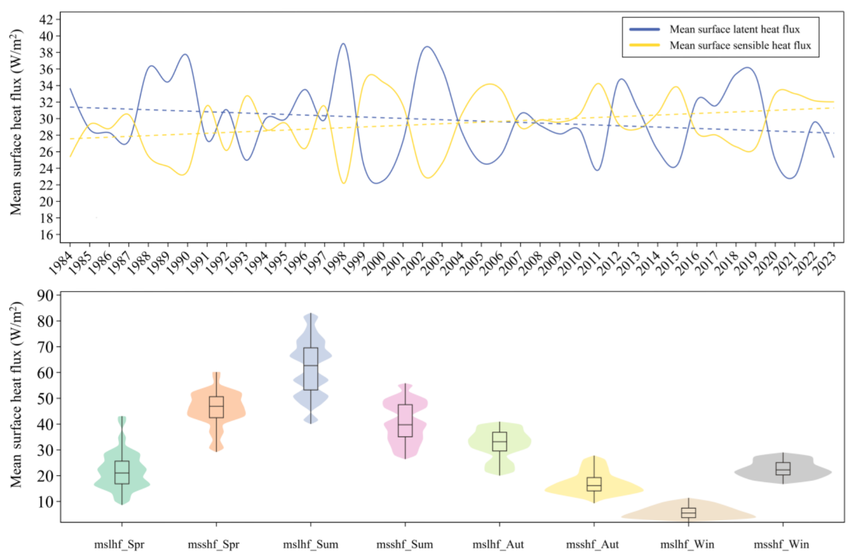


Figure 18 Regional Mean Surface Heat Flux

(4) Humidity and Evapotranspiration

1）Relative Humidity

The relative humidity decreases at a rate of 0.0987 per year. A decrease in humidity typically accelerates evaporation from the lake, leading to a reduction in lake area. The decrease in humidity means that the air becomes drier, and the evaporation rate increases. This accelerates the evaporation of lake water, resulting in a decline in both

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

lake water levels and area, intensifying the process of lake desiccation.

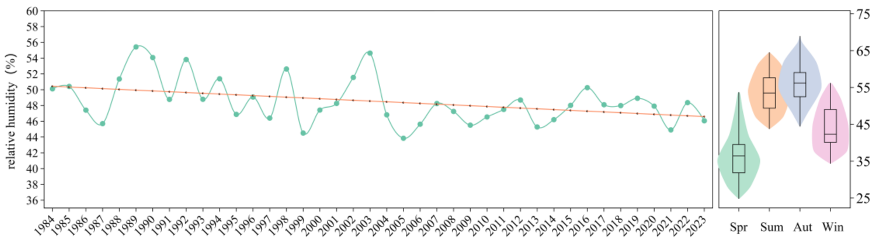


Figure 19 Regional Relative Humidity

2）Potential Evapotranspiration

Potential evapotranspiration increases at a rate of 1.9369 mm per year. The increase in evapotranspiration directly leads to the loss of water from the lake, making it an important factor contributing to the reduction in lake area. The rise in potential evapotranspiration indicates that both evaporation and plant transpiration in the lake area are increasing, further reducing the water volume of the lake. The increase in potential evapotranspiration has a significant impact on the lake area in all time periods, especially under drought and high-temperature conditions, where its effect is even more pronounced.

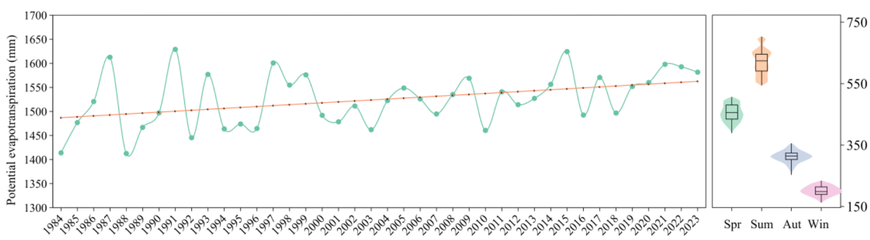


Figure 20 Regional Potential Evapotranspiration

(5) Drought

The drought index decreases at a rate of 0.0019 per year, indicating that the

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

drought conditions in the region are intensifying, further contributing to the shrinkage of the lake.

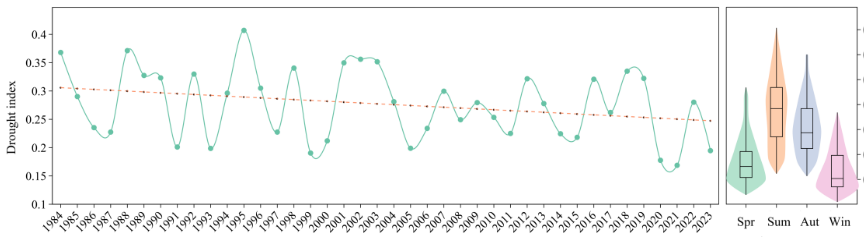


Figure 21 Regional Drought Index

A sliding T-test on the lake area reveals two turning points in the lake's area change, specifically in 2000 and 2015. Therefore, we divide the study period into three time segments: the first period from January 1984 to December 1999, the second period from January 2000 to December 2014, and the third period from January 2015 to July 2024, to investigate the causes of the changes in lake area.

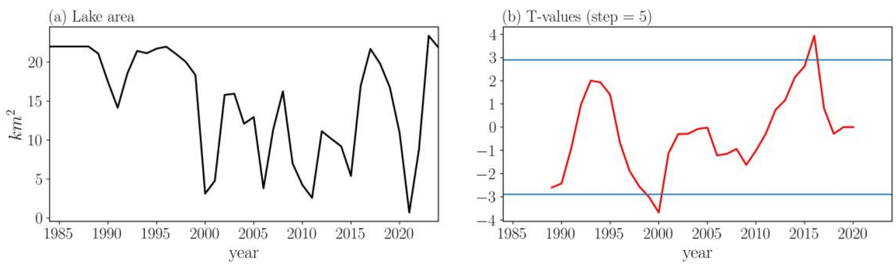


Figure 22 Sliding T-test

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

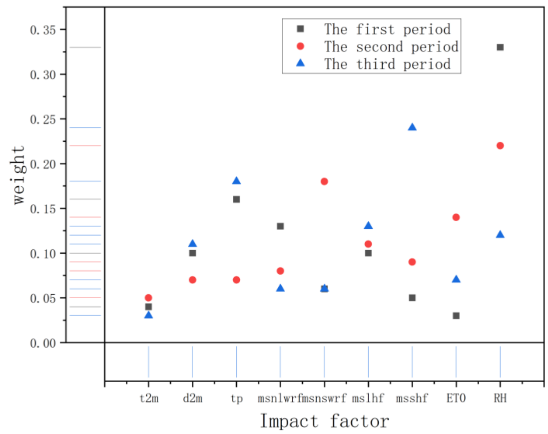


Figure 23 Factor Weight Diagram (Three Time Periods)

This paper conducts a factor analysis for three time periods, as shown in the figure 23. To better understand the causes of the change in the Bahai Nao lake area, we can explore several aspects in detail, including the direct and indirect effects of climate change, the roles of precipitation and evaporation, the effects of radiation and energy exchanges, and the combined effects of humidity and drought.

From 1984 to 1999, the change in Bahai Nao lake area was mainly driven by a decrease in precipitation and a decline in humidity. During this period, the main factors affecting lake area were humidity and precipitation. The weight of humidity reached 0.33, and that of precipitation was 0.16. This suggests that during this period, the decline in humidity significantly increased evaporation, leading to a reduction in lake area. Precipitation also decreased at a rate of 1.7355 mm per year, further exacerbating the loss of lake water. Although temperature rose (at a rate of 0.0429°C per year), its impact on lake area was relatively small (weight of 0.04).

From January 2000 to December 2014, the main influencing factors were humidity

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

(weight of 0.22) and surface net longwave radiation (weight of 0.18). The decline in humidity intensified evaporation, and the increase in surface net shortwave radiation (at a rate of 0.0653 W/m² per year) also significantly influenced evaporation (weight of 0.08). The continuous decrease in precipitation (weight of 0.07) and changes in latent heat flux (weight of 0.09) gradually reduced their impact on lake area.

From January 2015 to July 2024, the impact of sensible heat flux significantly increased (weight of 0.24), reflecting an increased effect of surface heat exchange on water evaporation. Meanwhile, the increase in potential evapotranspiration (weight of 0.13, with a rate of 1.9369 mm per year) indicated a sustained rise in water loss in the region. Although temperature continued to rise (at a rate of 0.0429°C per year), its direct impact on lake area was relatively limited (weight of 0.03). Additionally, the ongoing decrease in precipitation (at a rate of 1.7355 mm per year) continued to contribute to the shrinking of the lake area, and the decline in humidity (at a rate of 0.0987) further exacerbated evaporation (weight of 0.07).

The driving factors of the Bahai Nao lake area changes show significant differences in different time periods. From 1984 to 1999, humidity and precipitation were the primary factors determining lake area change. Over time, from 2000 to 2014, the impact of declining humidity and increasing shortwave radiation gradually strengthened, while the effects of reduced precipitation and changes in latent heat flux weakened. From 2015 to 2024, the rise in temperature, increase in sensible heat flux, and the increase in potential evapotranspiration became the major drivers, making the trend of lake area shrinkage more significant.

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

Overall, the reduction in the Bahai Nao lake area is primarily driven by the combined effects of climate warming, enhanced evaporation, and reduced precipitation. Particularly under the changes in humidity and evapotranspiration, the evaporation rate of the lake has notably accelerated.

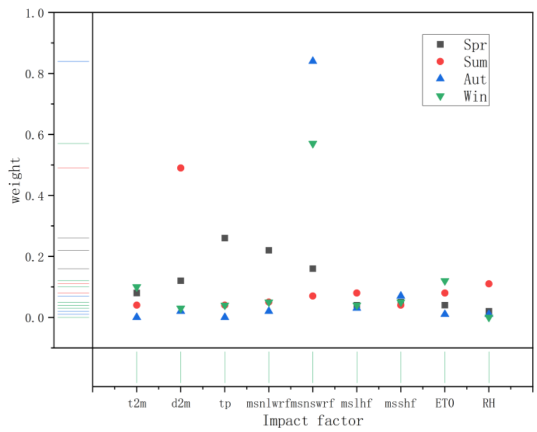


Figure 24 Weight of Influencing Factors by Season

Subsequently, a seasonal analysis of the influencing factors on the lake area of Bahai Nao Lake in spring, summer, autumn, and winter was conducted, as shown in Figure 24.

From the previous analysis, the climate conditions in spring were: temperature 10.16°C, 2-meter dew point temperature 6.28°C, rainfall 76.14 mm, average net long-wave radiation 103.13 W/m², potential evapotranspiration 456.37 mm, and humidity 36.62%.

Spring is the main replenishment period for rainfall, with a weight of 0.26, significantly higher than the other seasons. Therefore, although the 76.14 mm of rainfall is not as abundant as in summer, it still plays an important role in replenishing the lake's

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

water volume. At the same time, radiation energy (103.13 W/m²) and relatively low humidity lead to significant evaporation. The high potential evapotranspiration of 456.37 mm also indicates that the evaporation potential in spring is high, which could partly offset the water replenishment brought by the rainfall. This is why the increase in lake water volume is slow in spring.

In summer, the climate conditions were: temperature 21.58°C, 2-meter dew point temperature 11.06°C, rainfall 229.87 mm, average net long-wave radiation 89.14 W/m², potential evapotranspiration 620.36 mm, and humidity 53.58%.

Summer is the season with the most abundant rainfall (229.87 mm), which is a key replenishment period for the lake's water volume. However, the high weight of the 2-meter dew point temperature (0.49) indicates that humidity controls the evaporation of the water body. Due to the high humidity (53.58%), the evaporation rate of the lake is relatively low. Despite the very high potential evapotranspiration (620.36 mm), the impact of humidity significantly slows down the evaporation of moisture, allowing the lake area to maintain relatively well during the summer.

In autumn, the climate conditions were: temperature 7.86°C, 2-meter dew point temperature 0.78°C, rainfall 96.58 mm, average net long-wave radiation 83.52 W/m², average net short-wave radiation 124.12 W/m², potential evapotranspiration 314.29 mm, and humidity 55.91%.

The variation in lake water volume in autumn is mainly driven by solar short-wave radiation, with a weight of 0.84. This indicates that although the rainfall in autumn is moderate (96.58 mm), the higher short-wave radiation (124.12 W/m²) leads to intense

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

evaporation. The potential evapotranspiration is 314.29 mm, showing that the lake evaporation is large, and although the relative humidity is relatively high (55.91%), it is insufficient to prevent the reduction of the lake water volume. The low temperature in autumn (7.86°C) further indicates that although the temperature impact is minimal, the radiation intensity still determines the seasonal reduction in lake area.

in winter, the climate conditions were: temperature 2.79°C, 2-meter dew point temperature 14.66°C, rainfall 26.54 mm, average net long-wave radiation 88.52 W/m², average net short-wave radiation 114.66 W/m², potential evapotranspiration 200.32 mm, and humidity 44.03%.

Winter sees a significant decrease in temperature (2.79°C), and evaporation is suppressed. However, the surface short-wave radiation remains the main influencing factor in winter, with a weight of 0.57. This suggests that, despite the lower rainfall (26.54 mm) in winter, radiation still plays a role in moisture evaporation. The potential evapotranspiration is 200.32 mm, which is lower compared to other seasons, but still enough to affect the lake's water volume. The temperature has a small contribution to the lake water volume change (0.10), indicating that in winter, the main evaporation driving force is solar radiation.

In summary, the seasonal variation of Bahai Nao Lake's water volume is mainly affected by rainfall, radiation, humidity, and evapotranspiration. In spring, the change in lake area is primarily determined by rainfall replenishment, but higher radiation and potential evapotranspiration weaken the accumulation of water. In summer, humidity and dew point temperature are the dominant factors affecting the lake area. Despite

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

abundant rainfall, high humidity slows down evaporation, maintaining the lake's water volume. In autumn, due to intense solar short-wave radiation, the lake water volume decreases significantly, with evaporation being the dominant factor. In winter, despite the low temperature, radiation remains the main driving factor for evaporation, leading to a continued reduction in the lake's water volume.

This seasonal hydrological change in the lake suggests that different seasonal factors influencing the lake area focus on the interaction between rainfall and radiation, as well as the regulatory effect of humidity on evaporation.

4. Discussion

This study established a continuous monthly record of Bahannao Lake from 1984 to 2024 using an optimized extraction framework that integrates seasonal index selection, maximum connectivity analysis, and mutual information – based gap filling. Unlike previous long-term products such as the JRC Global Surface Water dataset, which are often constrained by cloud contamination and temporal discontinuity, our framework ensures both higher temporal continuity and robustness under complex environmental conditions.

The methodological improvements provide several advantages. First, the seasonal use of NDWI and MNDSI effectively distinguishes water bodies under freezing and non-freezing conditions, outperforming traditional single-index approaches (McFeeters, 1996; Yao et al., 2015). Second, the combination of Otsu thresholding with DEM constraints reduces misclassification from shadows and topography, a common issue in arid-region lakes with irregular terrain. Third, the MI-based filling strategy reconstructs

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

cloud- and stripe-contaminated images, extending the applicability of Landsat data and providing a longer, more reliable time series compared with interpolation-only methods (Zhao and Gao, 2018). Together, these innovations establish a transferable framework for dynamic lake monitoring, particularly suited for arid and data-scarce regions where conventional products often fail. Importantly, these improvements systematically address the common challenges highlighted in earlier studies, including cloud contamination, seasonal variations, topographic interference, and spectral complexity in inland waters (Mouw et al., 2015; Palmer et al., 2015; Shen et al., 2017; Cao et al., 2019).

Beyond methodological advances, the long-term record reveals important insights into the hydro-climatic controls of arid-region lakes. Precipitation was identified as the dominant driver of lake expansion in spring and summer, while shortwave radiation governed evaporation in autumn and winter. This seasonal contrast aligns with findings from other arid and high-altitude regions, such as Nam Co on the Tibetan Plateau (Li et al., 2017) and lakes on the Mongolian Plateau (Tao et al., 2015), where radiation and humidity strongly modulate evaporation under limited precipitation inputs. However, our results also highlight a pronounced nonlinear shift in dominant drivers over time: humidity and precipitation before 2000, radiation and humidity variability during 2000 – 2014, and energy flux intensification after 2015. This temporal evolution differs from some humid-region lakes, where nutrient enrichment or human disturbance dominate changes (Jeppesen et al., 2014), suggesting that climatic forcing plays a more persistent role in arid environments.

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

These findings carry broader implications for ecohydrological research and water resource management. By quantifying the nonlinear interactions of multiple climatic factors, our study demonstrates that lake dynamics in arid regions cannot be attributed to a single driver but emerge from the shifting balance of precipitation, radiation, and evapotranspiration. This highlights the vulnerability of arid-region lakes to climate change, where even moderate increases in radiation or evapotranspiration can outweigh precipitation recovery. Such insights are crucial for improving hydrological models, projecting future lake dynamics, and informing adaptive management strategies under intensified drought risk.

Several limitations should be acknowledged. First, while remote sensing provides a robust record of surface area, subsurface processes such as groundwater inflow and outflow were not explicitly considered, which may contribute to lake water balance. Second, the spatial resolution of Landsat (30 m) limits the detection of small-scale shoreline changes, and higher-resolution sensors (e.g., Sentinel-2) could improve accuracy in future studies. Third, although XGBoost effectively captured nonlinear relationships, its “black-box” nature limits interpretability compared with process-based hydrological models. Future research could combine machine learning with ecohydrological modeling and socioeconomic datasets to better quantify the combined impacts of climate variability and human activities on arid-region lakes.

5. Conclusion

This study developed an optimized lake area extraction framework and applied it to construct a continuous monthly record of Bahannao Lake from 1984 to 2024. The method integrates seasonal index selection, adaptive thresholding, maximum connectivity analysis, and mutual information–based gap filling, effectively addressing cloud contamination, seasonal freezing, and data gaps in remote sensing images.

The long-term record reveals both significant interannual variability and clear seasonal differences in lake dynamics. Precipitation was the dominant driver of lake expansion in spring and summer, whereas shortwave radiation controlled evaporation in autumn and winter. Factor weights further demonstrate a temporal shift in dominant mechanisms: humidity decline and precipitation reduction before 2000; enhanced radiation and humidity variability during 2000–2014; and intensified sensible heat flux and potential evapotranspiration after 2015.

These findings highlight the nonlinear and evolving interactions of hydro-climatic drivers regulating arid-region lakes. The proposed framework not only improves the reliability of long-term lake monitoring but also provides actionable insights for ecohydrological research, water resource management, and climate change adaptation in arid environments.

Competing interests

The authors declare that they have no conflict of interest.

Code/Data availability

The data and code that support the findings of this study are available from the corresponding author upon reasonable request.

Author contribution

R Z and X W conceived and designed the study, developed the methodology, curated the data, and performed the formal analysis. R Z was responsible for visualization. R Z and X W prepared the original draft of the manuscript, and all authors contributed to reviewing and editing the paper. X W provided overall supervision.

Acknowledgments

We are grateful to the National Key R&D Program of China (No. 2023YFC3206504), National Natural Science Foundation of China (No. 52121006, 41961124006), Postgraduate Thesis Fund of Nanjing Hydraulic Research Institute(Yy524010), Young Top-Notch Talent Support Program of National High-level Talents Special Support Plan, and Research Project of Ministry of Natural Resources (No. 20210103)，Research Project of Academy of Science and Technology of Inner Mongolia (No. 2024RCYJ05003) for providing financial support for this research. We are also thankful international key authors and their agencies. We are also thankful to anonymous reviewers and editors for their helpful comments and suggestions.

References

Adrian R, O'Reilly C M, Zagarese H, et al. Lakes as sentinels of climate change [J]. Limnology and Oceanography, 2009, 54(6part2): 2283-2297

Bergé-Nguyen M, Crétaux J F. Inundations in the Inner Niger Delta: Monitoring and analysis using MODIS and global precipitation datasets [J]. Remote Sensing, 2015, 7(2): 2127-2151.

Busker T, de Roo A, Gelati E, et al. A global lake and reservoir volume analysis using

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

a surface water dataset and satellite altimetry[J]. Hydrology and Earth System Sciences, 2019, 23(2): 669-690.

Cao Z G, Ma R H, Duan H T, et al. A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes [J]. Remote Sensing of Environment, 2020, 248: 111974

Cao Z G, Ma R H, Duan H T, et al.. Effects of broad bandwidth on the remote sensing of inland waters: implications for high spatial resolution satellite data applications [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 153: 110-122

Carroll M L, Townshend J R G, DiMiceli C M, et al. Shrinking lakes of the Arctic: Spatial relationships and trajectory of change[J]. Geophysical Research Letters, 2011, 38(20).

Cooley S W, Smith L C, Stepan L, et al. Tracking dynamic northern surface water changes with high-frequency planet CubeSat imagery[J]. Remote Sensing, 2017, 9(12): 1306.

Donchyts G, Baart F, Winsemius H, et al. Earth's surface water change over the past 30 years[J]. Nature Climate Change, 2016, 6(9): 810-813.

Grant L,Vanderkelen l, Gudmundsson L, et al. Attribution ofglobal lake systems change to anthropogenic forcing[J].Nature Geoscience, 2021, 14(11):849-854.

Huang C, Chen Y, Zhang S, et al. Detecting, extracting, and monitoring surface water from space using optical sensors: A review[J]. Reviews of Geophysics, 2018, 56(2): 333-360.

Jeppesen E, Meerhoff M, Davidson T, et al. Climate change impacts on lakes: an

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

integrated ecological perspective based on a multi-faceted approach, with special focus on shallow lakes[J]. Journal of Limnology, 2014, 73.

Klein I, Gessner U, Dietz A J, et al. Global WaterPack – A 250 m resolution dataset revealing the daily dynamics of global inland water bodies[J]. Remote sensing of environment, 2017, 198: 345-362.

Kravitz J, Matthews M, Lain L, et al. Potential for high fidelity global mapping of common inland water quality products at high spatial and temporal resolutions based on a synthetic data and machine learning approach[J]. Frontiers in Environmental Science, 2021, 9: 58766

LABA Zhuoma, DEJI Yangzong, LA Ba, et al. Remote sensing analysis on the area variations of Tangra Yutso in Tibetan Plateau over the past 40 years [J]. Lake Science,2017,29(02):480-489.

Li, L., Long, D., Wang, Y. et al. Global dominance of seasonality in shaping lake-surface-extent dynamics. Nature (2025). https://doi.org/10.1038/s41586-025-09046-3

Li M, Yan D H, Liu S H, et al, Variation Characteristics of Water Surface Area and Water Storage Capacity of Namucuo Lake in Recent 40 Years [J].Water Resources and Power, 2017, 35 (02): 41-43+52.

Liu Y B, Wu G P, Zhao X S, et al. Remote sensing for watershed hydrology: issues and challenges [J]. Advances in Earth Science, 2020, 35(5): 488-496

Liu Z W, Su Y L and Yang L. Limnology is a multidisciplinary and integrative science for studying inland waters: with special reference to the challenges and

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

opportunities for the development of limnology in China [J]. Journal of Lake Sciences, 2020, 32(5): 1244-1253

Ma R H, Duan H T, Hu C M, et al. half-century of changes in China's lakes: global warming or human influence? [J]. Geophysical Research Letters, 2010, 37(24): L24106

Ma R H, Yang G S, Duan H T, et al. China's lakes at present: number, area and spatial distribution [J]. Science China Earth Sciences, 2011, 54(2): 283-289

Ma Y, Xu N, Zhang W H, et al. Increasing water levels of global lakes between 2003 and 2009 [J]. IEEE Geoscience and Remote Sensing Letters, 2020b, 17(2): 187-191

McFeeters S K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features[J]. International journal of remote sensing, 1996, 17(7): 1425-1432.

Mouw C B, Greb S, Aurin D, et al. Aquatic color radiometry remote sensing of coastal and inland waters: challenges and recommendations for future satellite missions [J]. Remote Sensing of Environment, 2015, 160: 15-30

Palmer S C J, Kutser T and Hunter P D. Remote sensing of inland waters: challenges, progress and future directions [J]. Remote Sensing of Environment, 2015, 157: 1-8

Pekel J F, Cottam A, Gorelick N, et al. High-resolution mapping of global surface water and its long-term changes [J]. Nature, 2016, 540(7633): 418-422

Pekel J F, Vancutsem C, Bastin L, et al. A near real-time water surface detection method

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

based on HSV transformation of MODIS multi-spectral time series data[J]. Remote sensing of environment, 2014, 140: 704-716.

Pickens A H, Hansen M C, Hancher M, et al. Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full Landsat time-series [J]. Remote Sensing of Environment, 2020, 243: 111792

Plug L J, Walls C, Scott B M. Tundra lake changes from 1978 to 2001 on the Tuktoyaktuk Peninsula, western Canadian Arctic[J]. Geophysical Research Letters, 2008, 35(3).

Preston D L, Caine N, McKnight D M, et al. Climate regulates alpine lake ice cover phenology and aquatic ecosystem structure[J]. Geophysical Research Letters, 2016, 43(10): 5353-5360.

Råman Vinnå L, Medhaug I, Schmid M, et al. The vulnerability of lakes to climate change along an altitudinal gradient[J]. Communications Earth & Environment, 2021, 2(1): 35.

Rossow W B, Schiffer R A. Advances in understanding clouds from ISCCP[J]. Bulletin of the American Meteorological Society, 1999, 80(11): 2261-2288.

Schmid M,Hunziker S, Wüest A. Lake surface temperaturesin a changing climate: a global sensitivity analysis[J]. Climatichange,2014,124:301-315.

Secretariat G. Implementation plan for the global observing system for climate in support of the UNFCCC (2010 Update) [C]//Proceedings of the Conference of the Parties (COP), Copenhagen, Denmark. 2009: 7-18.

Shen M, Duan H T, Cao Z G, et al. Determination of the downwelling diffuse

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

attenuation coefficient of lake water with the Sentinel-3A OLCI [J]. Remote Sensing, 2017, 9(12): 1246

Tao S, Fang J, Zhao X, et al. Rapid loss of lakes on the Mongolian Plateau[J]. Proceedings of the National Academy of Sciences, 2015, 112(7): 2281-2286.

Tong Y, Feng L, Wang X, et al. Global lakes are warmingslower than surface air temperature due to accelerate devaporation [J]. Nature Water, 2023,1(11):929-940

Vincent W F. Effects of climate change on lakes[J]. 2009.

Wang J, Song C, Reager J T, et al. Recent global decline in endorheic basin water storages[J]. Nature geoscience, 2018, 11(12): 926-932.

Woolway R I, Kraemer B M, Lenters J D, et al. Global lake responses to climate change[J]. Nature Reviews Earth & Environment, 2020, 1(8): 388-403.

Yang G S, Ma R H, Zhang L, et al. Lake status, major problems and protection strategy in China [J]. Journal of Lake Sciences, 2010, 22(6): 799-810

Yao F, Wang C, Dong D, et al. High-resolution map of urban surface water using ZY-3 multi-spectral imagery[J]. Remote Sensing, 2015, 7(9): 12336-12355.

Yao F, Wang J, Yang K, et al. Lake storage variation on the endorheic Tibetan Plateau and its attribution to climate change since the new millennium[J]. Environmental Research Letters, 2018, 13(6): 064011.

Zhang B, Li J S, Shen Q, et al. Recent research progress on long time series and large scale optical remote sensing of inland water [J]. National Remote Sensing Bulletin, 2021, 25(1): 37-52

Zhang G Q, Yao T D, Chen W F, et al. Regional differences of lake evolution across

https://doi.org/10.5194/egusphere-2025-4356
Preprint. Discussion started: 23 September 2025

China during 1960s—2015 and its natural and anthropogenic causes [J]. Remote Sensing of Environment, 2019, 221: 386-404

Zhang Y L. Progress and prospect in lake optics: a review [J]. Journal of Lake Sciences, 2011, 23(4): 483-497

Zhao G, Gao H. Automatic correction of contaminated images for assessment of reservoir surface area dynamics[J]. Geophysical Research Letters, 2018, 45(12): 6092-6099.

Zhou W, Wang L, Li D, et al. Spatial pattern of lake evaporation increases under global warming linked to regional hydroclimate change [J]. Communications Earth & Environment, 2021, 2(1): 255.