https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

Open Access Earth System Science Data Discussions

SETP_GLI: An annual 10–30 m glacial lake inventory for the southeastern Tibetan Plateau from 1990 to 2025

Hao Li[1], Jie Dou[1,2], Timothy Kusky[1], Shun Dong[3], Zihao Shi[2], Jie Li[4], Xinjian Xiang[1], and Fange Ding[1]

[1] Badong National Observation and Research Station of Geohazards, China University of Geosciences, Wuhan, 430074, China

[2] School of Future Technology, China University of Geosciences, Wuhan, 430074, China

[3] China Yangtze Power Co., Ltd., Wuhan, 430072, China

[4] National Engineering Research Center of Geographic Information System, China University of Geosciences, Wuhan, 430078, China

Correspondence to: Jie Dou (douj888@gmail.com)

Abstract. Glacial lakes in the southeastern Tibetan Plateau (SETP) have expanded, increasing the potential for cascading hazards associated with glacial lake outburst floods (GLOFs). However, long-term, annual monitoring data that include micro glacial lakes remain relatively limited for this region. To address this gap, this study integrated Landsat series and Sentinel-2 imagery and used the GLA-RCNN deep learning framework with an embedded Convolutional Block Attention Module to construct and release an annual glacial lake inventory (SETP_GLI). The dataset comprises 36 annual vector layers from 1990 to 2025, recording the annual evolution of regional glacial lake numbers and areas. The use of 10 m resolution imagery and model optimization improved the detection of micro glacial lakes (<0.01 km²). The inventory provides annual vector boundaries and standardized physical attributes—including longitude, latitude, area, perimeter, and mean elevation, together with area uncertainty metrics derived from mixed-pixel theory. Quality assessments indicated that the extraction framework is robust against interference from mountain shadows and turbid water. For model performance, the overall F1 scores for typical years remained above 0.82 (with a maximum of 0.895); cross-validation with existing public databases (Hi-MAG and Glacial lake inventory of high-mountain Asia) showed that the matched polygon-level Intersection over Union (IoU) ranged from 0.54 to 0.80, with spatial agreement increasing with improvements in historical image quality. Spatiotemporal analysis revealed a persistent expansion trend, with the annual area growth rate rising from 3.65 ± 1.12 km² a⁻¹ (1990–2012) to 5.95 ± 2.44 km² a⁻¹ (2016–2025). The dataset is archived at the National Tibetan Plateau Data Center (TPDC) (https://doi.org/10.11888/Cryos.tpdc.303491), with processing code released openly. SETP_GLI serves as a baseline dataset for cryospheric response analysis, hydrological modeling, and GLOF risk assessment.

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

Open Access Earth System Science Data Discussions

1. Introduction

Glacial lakes are sensitive indicators of cryospheric change and integral components of hydrological systems in high-mountain regions. In High Mountain Asia (HMA), sustained glacier mass loss and terminus retreat have facilitated the formation, expansion, drainage, and coalescence of glacial lakes (Zhang et al., 2022; Yao et al., 2022). These dynamics influence seasonal water storage and downstream hydrological processes, and can increase the potential risk of glacial lake outburst floods (GLOFs) (Harrison et al., 2018; Veh et al., 2023; Zheng et al., 2021). The southeastern Tibetan Plateau (SETP) is particularly sensitive to glacier–lake interactions due to its steep topographic relief, strong monsoonal influence, rapid glacier thinning, and extensively distributed maritime glacier systems (Zeitler et al., 2014; Yao et al., 2019; Brun et al., 2017). Accordingly, constructing a long-term glacial lake inventory with high temporal consistency is essential for assessing regional cryospheric changes and supporting regional-scale hydrological and hazard-related research.

Although several large-scale glacial lake datasets for the Tibetan Plateau or HMA have been released, providing important baselines for regional assessments (Chen et al., 2021; Wang et al., 2020), three primary limitations remain for refined applications in the SETP. First, many datasets have limited update frequency; many existing inventories cover only specific years or lack records of recent dynamics (Shugar et al., 2020). Second, most large-scale products rely predominantly on 30 m resolution optical imagery, which limits their capability to detect small glacial lakes (<0.01 km²) and newly formed early-stage lakes (Zhang et al., 2022). These small lakes can be important morphological indicators for the early stages of outburst hazards (Watson et al., 2017). Third, existing products are often inadequate in providing polygon-level uncertainty quantification, validation samples, image-source metadata, and user guidelines for inter-annual change analysis, making it difficult to support highly reproducible and continuous annual evolution studies (Emmer et al., 2022; Nie et al., 2017).

The aforementioned limitations are further amplified by the extremely complex geographic environment of the SETP. Challenges such as deep topographic shadows, highly turbid water, supraglacial debris, fragmented moraine landscapes, seasonal snow cover, and persistent cloud contamination create challenges to automated glacial lake mapping (Zhang et al., 2021). In such complex backgrounds, small lakes are easily omitted due to the mixed-pixel effect, while shadows, river segments, and seasonal meltwater ponds are frequently misidentified as glacial lakes by optical sensors (Li et al., 2022; Zhao et al., 2018). To overcome these challenges, there is a need to establish an integrated processing workflow that incorporates multi-source remote sensing imagery, topographic constraints, automated extraction, rigorous manual quality control, and explicit uncertainty quantification (Nie et al., 2017; Emmer et al., 2022).

To address these requirements, this study developed and released an annual glacial lake inventory for the southeastern Tibetan Plateau covering the period from 1990 to 2025 (SETP_GLI). This paper provides a detailed documentation of the input data, image-selection protocols, extraction workflows, manual quality control, attribute calculations, uncertainty estimations, technical validation, data records, and usage instructions. The primary contributions of this dataset include: (1) a 36-year annual vector inventory of glacial lakes at 10–30 m mapping resolution; (2) standardized polygon-level physical attributes and uncertainty estimates based on boundary pixels; (3) stratified validation across multiple sensors, size classes,

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

and complex alpine backgrounds; and (4) explicit usage guidelines to support the scientifically consistent interpretation of long-term glacial lake evolution across different sensors.

2. Study Area

The SETP is situated in the geologically active Eastern Himalayan Syntaxis, geographically spanning 92°E–99°E and 27°N–31°N. It encompasses the eastern segments of the Nyainqêntanglha Mountains, Gangdise Mountains, and Himalayas, as well as the western part of the Hengduan Mountains, with a total area of approximately 150,000 km². The region has complex topography, characterized by deeply incised high mountains and gorges, with an average elevation exceeding 4,000 m (Nie et al., 2017). It is among the regions with high rates of crustal uplift and surface denudation (Zeitler et al., 2014). Climatically, as a major moisture corridor that transports warm and humid air from the Indian Ocean to the interior of the plateau, the SETP exhibits significant spatial heterogeneity in precipitation (Yang et al., 2014), with uneven spatial and seasonal distributions. In the lower-elevation southern regions, annual precipitation can exceed 3,000 mm, occasionally reaching over 5,000 mm, the vast majority of which is concentrated in the summer (Zhang et al., 2024). Meteorological observations indicate that the region is undergoing a pronounced warming and wetting trend; over the past few decades, the mean annual temperature has risen at a rate of approximately 0.30–0.40 °C/10a, exceeding the global average over the same period (Yao et al., 2019). This unique alpine-humid hydrothermal environment has fostered the development of extensive maritime glaciers. According to the Randolph Glacier Inventory (RGI 7.0), the study area contains approximately 8,600 modern glaciers with a total area of roughly 13,200 km², representing one of the highest concentrations of maritime glaciers on the Tibetan Plateau. However, these glaciers are highly sensitive to climate change and have experienced substantial mass loss in recent decades. Recent geodetic estimates reveal that from 2000 to 2019, the glacier mass loss rate in this region reached -0.62 ± 0.10 m w.e. a^{-1} (Hugonnet et al., 2021), indicating substantial glacier ablation in High Mountain Asia. This accelerated ablation has contributed to cascading hydro-geomorphic responses, most notably the proliferation and areal expansion of glacial lakes, making the region an important area for studying rapid glacial lake evolution.

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

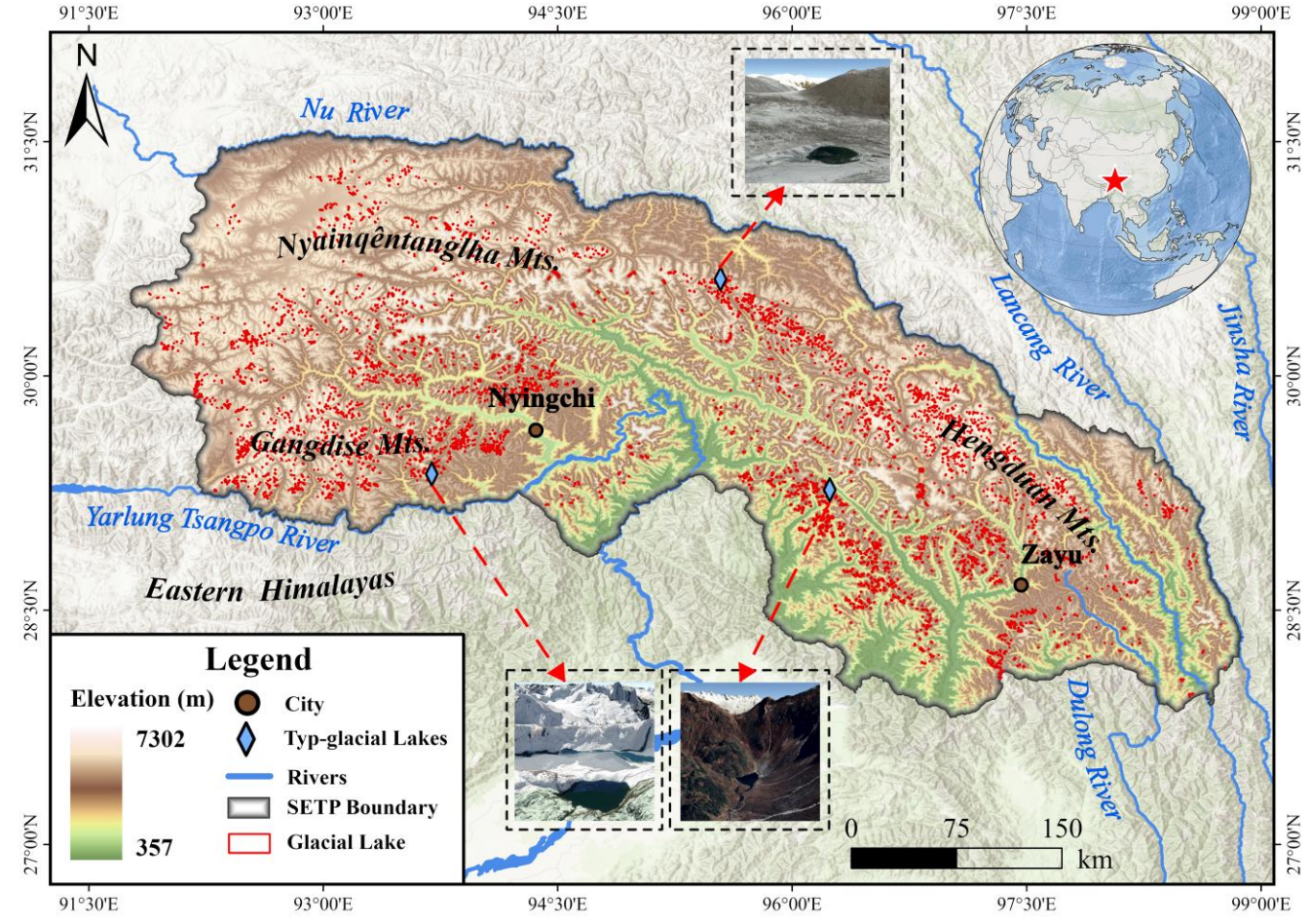

Figure 1. Overview map of the study area. The boundary of the SETP is derived from Zhao et al. (2022). The reference glacial lake locations and boundaries shown for study-area context are based on the Glacial lake inventory of high-mountain Asia (Wang et al., 2020; http://www.ncdc.ac.cn); these reference data are not the final SETP_GLI product generated in this study.

3. Data Sources

3.1 Satellite Imagery and Auxiliary Topographic Data

To construct a long-term glacial lake inventory covering 1990 to 2025, a phased satellite data integration strategy was implemented. For 1990–2012, Landsat 5 TM and Landsat 7 ETM+ Surface Reflectance (SR) products from the USGS were used; for 2013–2015, Landsat 8 OLI SR products were used. Landsat 7 ETM+ was used only as an auxiliary source to fill spatiotemporal gaps, thereby minimizing potential errors introduced by scan-line corrector (SLC-off) issues. For the 2016–2025 period, Sentinel-2 MSI (Level-2A) products from the ESA were incorporated, using their 10 m spatial resolution to enhance monitoring capabilities for micro-scale targets. Regarding topographic and reference data, SRTM DEM was used for slope derivation for the Landsat-period topographic auxiliary layer, and ALOS PALSAR RTC DEM was used for slope

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

95 derivation for the Landsat 8 and Sentinel-2 topographic auxiliary layer; a 10 m DEM was used to derive the lake-surface
96 elevation attribute. Furthermore, the Hi-MAG (Chen et al., 2021) and Glacial lake inventory of high-mountain Asia (Wang et
97 al., 2020) databases were selected as reference sources to assist in label generation for model training and quantitative accuracy
98 assessment of the extraction results.

99 **Table 1. Data sources and characteristics.**

Category	Sensor / Dataset	Time Span	Spatial Res.	Product Level / Source	Key Usage
Optical Imagery	Landsat 5 TM	1990–2011	30 m	USGS Collection 2 Tier 1 (Surface Reflectance)	Historical inventory
	Landsat 7 ETM+	1999–2012	30 m	USGS Collection 2 Tier 1 (Surface Reflectance)	Gap-filling support
	Landsat 8 OLI	2013–2015	15 m	USGS Collection 2 Tier 1 (Pan-sharpened via NNDiffuse)	Transitional inventory
	Sentinel-2 MSI	2016–2025	10 m	ESA Level-2A (BOA Reflectance)	High-resolution monitoring
Terrain Data	SRTM DEM	1990–2012	30 m	NASA LP DAAC	Slope derivation for the Landsat-period topographic auxiliary layer
	ALOS PALSAR RTC	2013–2025	12.5 m	ASF DAAC (Resampled to 10/15 m)	Slope derivation for the Landsat 8 and Sentinel-2 topographic auxiliary layer
Validation	Hi-MAG Database	2008–2017	30 m	NCDC	Accuracy assessment
	Glacial lake inventory of high-mountain Asia	1990 and 2018	30 m	NCDC	Accuracy assessment

100

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026


Open Access Earth System Science Data Discussions

3.2 Image Selection Protocol

To ensure the temporal consistency and reproducibility of the SETP_GLI dataset, the following image selection and preprocessing protocols were implemented:

- Primary window: The preferred observation window was set from September to November each year (i.e., the late ablation season before widespread lake freezing). During this period, seasonal snow and cloud cover are typically minimized, and glacial lake boundaries remain relatively stable during late-season ablation recharge, facilitating differentiation between water bodies and transient snow.
- **Backup Strategy**: Because cloud cover is high in alpine regions, if suitable images with cloud cover < 20% were unavailable in the primary window, the search was expanded to include August. If no suitable observations were found from August to November, images from December or January/February of the following year (corresponding to observation gaps with minimal snow and incomplete lake freezing) were selected. This stepwise expansion strategy improves the temporal continuity of the 1990–2025 sequence.
- Primary-year consistency and documented supplementation: The dataset follows a target-year-priority mapping principle. Images acquired within the target year were used whenever possible to maintain annual phase consistency. In areas affected by persistent cloud, snow, or Landsat 7 SLC-off gaps, adjacent-year images were used only as supplementary gap-filling sources when necessary, and the corresponding scene IDs were recorded in the source_img attribute. If individual images were obscured by clouds, the Google Earth Engine (GEE) platform was used to perform median compositing or least-cloud-cover compositing of available observations within the target-year or documented backup observation window to generate an annual basemap with fewer cloud-contaminated pixels.

4. Dataset Generation Methods

The methodological framework of this study comprises four primary stages: (1) standardized preprocessing of multi-source data; (2) development of the GLA-RCNN model and extraction of glacial lakes; (3) automated post-processing coupled with refined manual visual correction; and (4) calculation of glacial lake attributes and uncertainty assessment. The overall technical workflow is illustrated in Figure 2.

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

Open Access Earth System Science Data Discussions

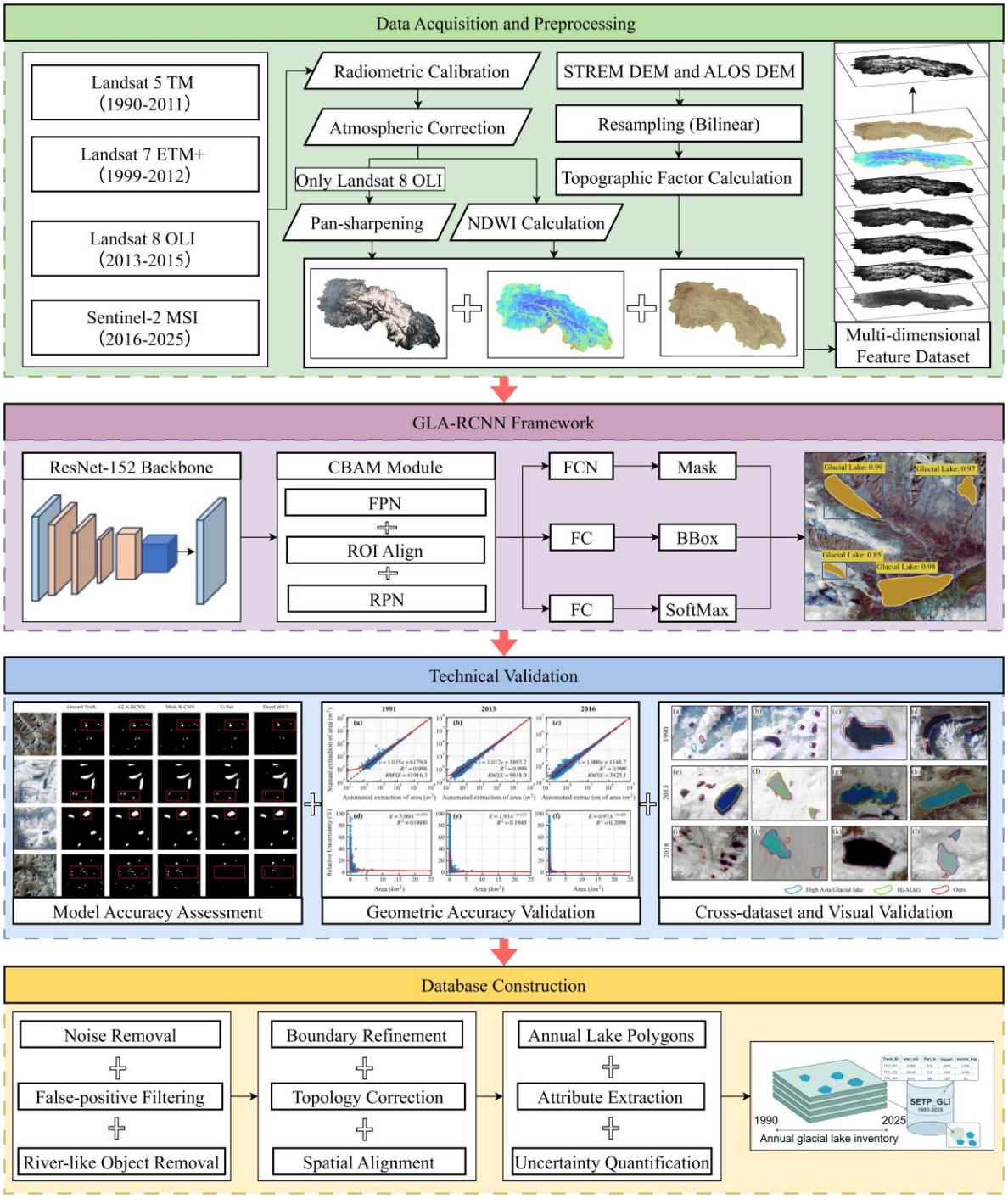

Figure 2. Methodological Workflow

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026


Open Access Earth System Science Data Discussions

4.1 Image Preprocessing

To satisfy the input requirements of the deep learning model and reduce inconsistencies associated with multi-source data selection, a standardized preprocessing workflow was implemented. Initially, sensor-available spectral bands sensitive to water bodies, including Red (R), Green (G), Blue (B), Near-Infrared (NIR), and SWIR where available at the required mapping resolution, were extracted and stacked for Landsat 5 TM, Landsat 7 ETM+, Landsat 8 OLI, and Sentinel-2 MSI images. Because Sentinel-2 does not provide a 10 m SWIR band, the input channel number was inferred automatically in the model configuration rather than forced to a fixed band combination for all sensors. Subsequently, to reduce variations in image tone and brightness caused by differences in acquisition dates, weather conditions, and solar elevation angles, a dodging algorithm was applied to color-balance the imagery for each year. This step improves the temporal consistency of feature extraction. For Landsat 8 OLI imagery, which includes a 15 m panchromatic band, the Nearest Neighbor Diffusion (NNDiffuse) pan-sharpening algorithm was used to fuse the 15 m panchromatic band with the 30 m multispectral bands. This algorithm increases spatial resolution while largely preserving the spectral fidelity of the original multispectral data, thereby reducing the spectral distortion commonly associated with conventional fusion techniques. This process increases the spatial resolution of the transitional-period data toward that of Sentinel-2 (10 m), improving subsequent target recognition accuracy. Furthermore, given the broad numerical range (0–10,000) of the Level-2 surface reflectance products, a Min-Max normalization method was applied to linearly map the pixel values of all bands to the [0, 1] interval, facilitating model convergence.

To achieve pixel-level registration between the topographic constraints and the optical imagery, bilinear interpolation was employed to resample the ALOS DEM to 15 m (to match Landsat 8) and 10 m (to match Sentinel-2), respectively. Compared to the nearest-neighbor method, bilinear interpolation effectively preserves the continuity of the topographic surface and minimizes jagged artifacts. Based on the preprocessed DEMs, a slope layer was generated. Following normalization, this layer was fed into the deep learning model as an auxiliary "topographic channel" to assist in reducing interference from alpine shadows. Concurrently, the Normalized Difference Water Index (NDWI) was calculated using the Green and NIR bands as follows: $NDWI = (Green - NIR) / (Green + NIR)$. Finally, the available optical bands were stacked with the NDWI and slope layers to construct a sensor-adaptive multi-channel dataset for model ingestion; the actual input-channel number was inferred automatically during training and inference.

4.2 Model Architecture

Given the large scale variations of glacial lakes and the complex environmental backgrounds in the SETP, standard convolutional neural networks (CNNs) can have difficulty extracting glacial lake features from high-frequency noise. To address this, we developed the GLA-RCNN deep learning model based on the Mask R-CNN instance segmentation framework by incorporating the Convolutional Block Attention Module (CBAM) (Woo et al., 2018) (Figure 3). This framework employs ResNet-152 as the backbone network for feature extraction, integrating CBAM modules at the ends of the deep residual blocks Stage 4 and Stage 5. The CBAM comprises two sequential sub-modules: channel attention and spatial attention. The channel

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

attention adaptively recalibrates the weights of each channel by applying global average pooling and max pooling to the feature maps, thereby highlighting the specific spectral-textural features of glacial lakes and suppressing non-water background noise. The spatial attention further focuses on high-response regions across the spatial dimensions of the feature maps, assisting the model in precisely locating glacial lake boundaries in low-contrast environments, such as deeply incised gorges. This helps reduce the common classification ambiguity between "mountain shadows" and "dark water bodies". To overcome the inherent flaw of traditional networks losing information on tiny targets during downsampling, we implemented targeted improvements to the Feature Pyramid Network (FPN). CBAM modules were embedded into each fused feature layer (P2–P5) following the top-down pathway and lateral connections of the FPN. This design strengthens the fusion of deep semantic information with shallow detailed information and provides cross-scale feature recalibration capabilities. For tiny glacial lakes smaller than 0.01 km², the improved FPN enhances their saliency on low-resolution feature maps, thereby reducing the omission rate caused by negligible pixel proportions and supporting model robustness across various glacial lake scales. In the final prediction stage of instance segmentation, to reduce spatial misalignment errors introduced by coordinate quantization in the traditional RoI Pooling layer, this study adopted the RoI Align technique. This technique utilizes bilinear interpolation to calculate the floating-point values of feature points within the Region of Interest (RoI), improving pixel-level correspondence between the feature maps and the original image. This is important for the fine delineation of irregular glacial lake boundaries. Model training employs a multi-task joint loss function for end-to-end optimization, with the total loss function defined as:

$$L = L_{cls} + L_{box} + L_{mask} \quad (1)$$

where L_{cls} represents the classification loss, L_{box} is the bounding box regression loss, and L_{mask} denotes the segmentation mask loss based on average binary cross-entropy. Through end-to-end optimization, accurate pixel-level extraction of glacial lake boundaries in complex terrains was achieved.

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

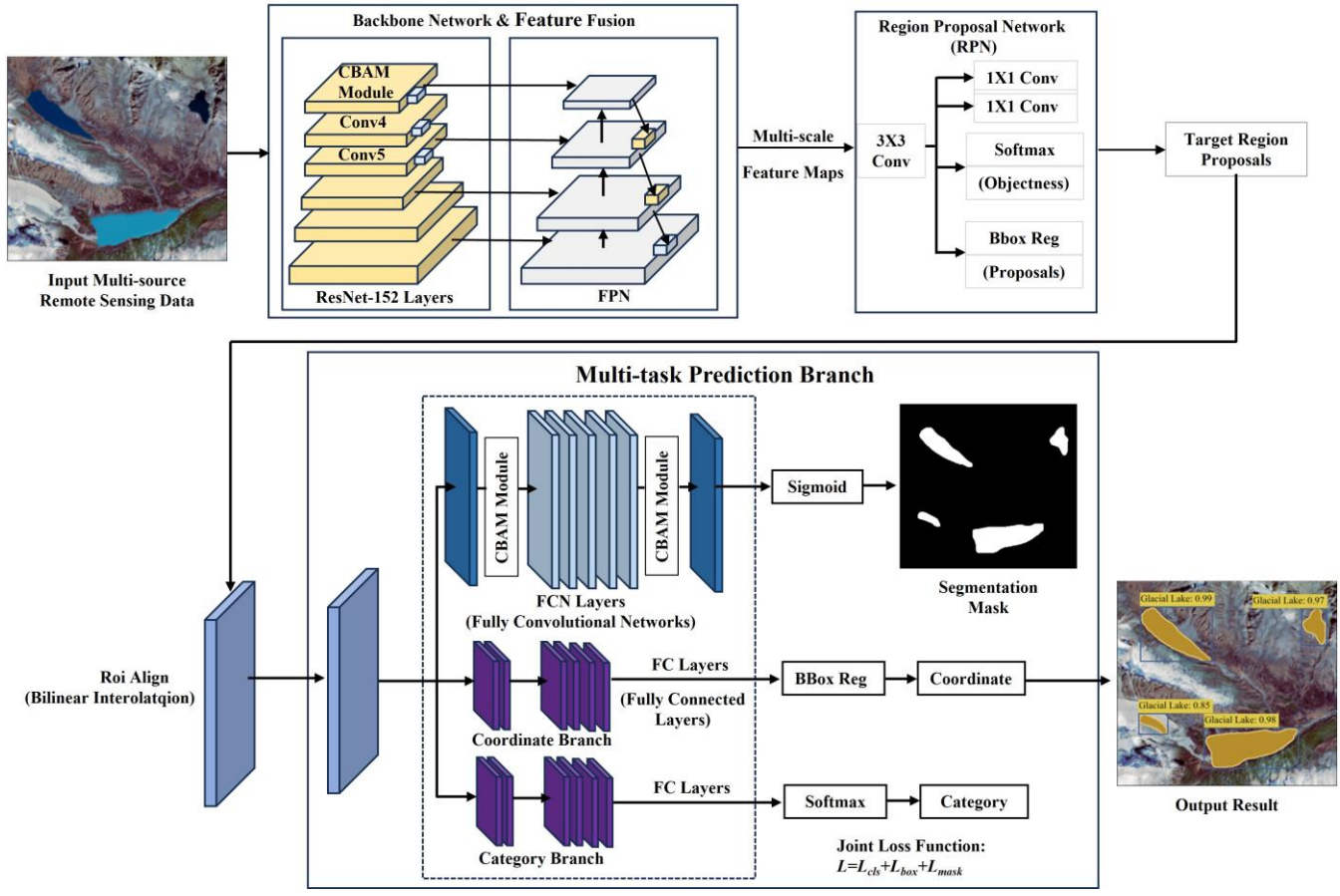

Figure 3. GLA-RCNN model framework

4.3 Experimental Design and Extraction Strategy

To evaluate the extraction performance of the proposed GLA-RCNN framework in complex geomorphic landscapes, and to mitigate potential systematic biases introduced by the transition among long-term multi-source remote-sensing datasets, this study established a "phased multi-scale" training strategy and standardized inference control protocols:

- **Phased data slicing strategy:** The preprocessed multi-channel remote sensing images were sliced using a sliding window of 256 x 256 pixels. Ground truth labels for glacial lakes were manually produced by visual interpretation of optical imagery, with the Glacial lake inventory of high-mountain Asia (Wang et al., 2020) used as a reference database. To reduce scale confusion and subsequent artificial step changes caused by varying physical resolutions (30 m, 15 m, and 10 m), this study generated labels specifically for three key temporal nodes (1991, 2013, and 2016) and trained three targeted sets of GLA-RCNN weights.
- **Sample ratio and spatial partitioning**: To enhance the model's resistance to interference against complex land surfaces, background patches containing deep gorge shadows, flat rivers, and local cloud shadows were

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

introduced into the training set, with the positive-to-negative sample ratio set at 1:2. The training dataset was augmented through techniques such as 90° flipping and rotation. To reduce the risk of "spatial leakage" caused by overlapping features in adjacent patches, a spatial block-based partitioning strategy was employed. The sample set was spatially separated and divided into a training set, a validation set, and an independent test set at a ratio of 8:1:1.

- **Training parameters and hyperparameter optimization**: The model was trained end-to-end by jointly minimizing classification loss, bounding-box regression loss, and segmentation mask loss. For hyperparameter configuration, the Non-Maximum Suppression (NMS) threshold for the Region Proposal Network (RPN) was set to 0.7, with an RPN batch size per image of 256 and a positive fraction of 0.5. In the Region of Interest (RoI) prediction stage, the score threshold was set to 0.05, the NMS threshold to 0.5, the RoI batch size per image was increased to 512 (with a positive fraction of 0.25), and the maximum number of detections per image was limited to 100. The model was built on the PyTorch framework, utilizing the AdamW optimizer to update parameters. The learning rate was selected during model training, and the total number of training epochs was set to 30.
- **Automated inference and tile merging:** During the large-scale inference phase across the entire region, an overlapping sliding window strategy was adopted to mitigate edge artifacts. The inference batch size was set to 4, the tile size to 256 pixels, and an overlap padding of 64 pixels was established. For the predicted probabilities in overlapping areas, a mean-probability merge policy was applied. In the mask binarization stage, the classification-confidence threshold was set to 0.9 to suppress false-positive errors in complex backgrounds.
- **Baseline comparison setup**: To quantify the performance of the improved model, using 1991 as a representative year, this study introduced three deep learning models—Mask R-CNN, U-Net, and DeepLab V3—as comparative models. By evaluating these models under a unified environment, we aimed to assess the contribution of the CBAM attention mechanism and the improved Feature Pyramid Network (FPN) in reducing terrain-shadow errors and preserving glacial lake topology.

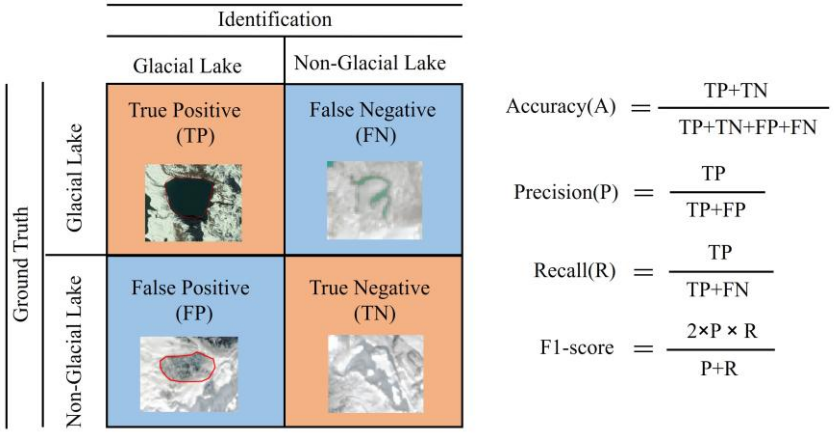

Figure 4. Confusion matrix for dichotomous problems

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026


Open Access Earth System Science Data Discussions

4.4 Post-processing and Manual Correction

Although deep learning models exhibit strong generalization across complex surface backgrounds, the initial extraction results may contain misclassifications or boundary topological errors due to interference from complex alpine topography and local cloud shadows. To meet the accuracy requirements of a benchmark dataset, this study established a post-processing mechanism comprising "automated rule-based filtering and full manual verification," to improve data reliability through multi-level quality control:

Automated Rule Filtering and Elimination: By integrating prior geoscientific knowledge, this study implemented strict morphological and semantic constraints to further enhance the reliability of the extraction results. To suppress isolated noise while maximizing the retention of visually plausible glacial lakes, we first applied a sensor-adaptive mapping threshold based on pixel continuity. Candidate lakes were retained only when they consisted of approximately 6-10 contiguous pixels (Nie et al., 2017; Wang et al., 2020). Therefore, the sensor-adaptive mapping thresholds for 10 m, 15 m, and 30 m imagery were set to 0.001 km² (10 pixels), 0.002025 km² (9 pixels), and 0.0054 km² (6 pixels), respectively. These thresholds were designed to remove obvious false positives while preserving true small glacial lakes as much as possible; they do not imply equivalent detection efficiency among sensors. In particular, lakes close to the theoretical 30 m threshold remain difficult to identify reliably because of mixed pixels, shadows, and boundary uncertainty. Accordingly, an additional uniform analysis threshold of 0.01 km² was applied only in the subsequent long-term trend analysis to standardize the comparison among different sensors. Furthermore, to address the problem that elongated river valleys are frequently misidentified as glacial lakes due to spectral similarities, this study introduced reference river vector data as a spatial mask. Morphological aspect-ratio constraints were used to remove linearly distributed river segments and seasonal meltwater, thereby ensuring the topological consistency of the dataset.

Full manual verification and visual interpretation: Following automated filtering, the verification team overlaid the initial extraction vectors onto the original Landsat and Sentinel-2 optical imagery for hierarchical visual interpretation. The main tasks of manual correction included the following four aspects:

- *Noise removal and false-positive suppression*: With reference to high-resolution historical imagery, residual cloud shadow noise, snow patches, and other false glacial lake polygons were manually removed.
- *Topological relationship restoration*: For merged polygons caused by extreme spatial proximity where the model failed to distinguish adjacent lakes, manual segmentation was performed to ensure the correct topology of independent glacial lake entities.
- *Spatial position geometric calibration*: To address significant coordinate offsets of glacial lakes in certain years caused by geometric registration residuals of the original imagery, this study utilized the stable spatial coordinates of the lake observed across the majority of the time series as a baseline to perform overall translation and alignment correction for the offset polygons.

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

- *Refined boundary reconstruction*: For fuzzy boundaries caused by supraglacial debris coverage or high water turbidity, refined manual trimming and boundary redrawing were conducted based on geomorphological features to improve spatial fidelity of the boundaries.

Quality review and final output: The verification process was independently conducted by three professionals with relevant backgrounds, implementing a cross-checking mechanism. Through the aforementioned quality control, interferences from mountain shadows in deeply incised gorges and sensor transitions were substantially reduced, ultimately yielding an annual glacial lake product (SETP_GLI) with consistent spatial topology and detailed physical boundaries.

4.5 Attribute Calculation and Uncertainty Assessment

During the model accuracy assessment and long-term dataset cross-validation phase, to reduce the bias of non-independent validation, the Hi-MAG database (Chen et al., 2021) and the Glacial lake inventory of high-mountain Asia (Wang et al., 2020) were used as external benchmark inventories for spatial cross-comparison. By conducting a spatial overlay analysis between our automated extraction results and these reference datasets, accuracy metrics, including Recall, Precision, and Intersection over Union (IoU) were calculated, providing an evaluation of the extraction performance for the long-term dataset. Furthermore, considering that existing public validation sets are primarily derived from 30 m resolution imagery, this study introduced meter-scale high-resolution historical imagery from Google Earth as a secondary validation source to further assess the extraction performance for micro glacial lakes (<0.01 km²) and complex background conditions. Using a stratified random sampling strategy, 200 sample points were randomly selected across typical error-prone regions—such as those with mountain shadow occlusion, moraine coverage, and turbid water—to conduct manual visual cross-validation. This dual validation mechanism, combining "quantitative benchmark comparison" with "qualitative visual inspection", supports the reliability of the long-term dataset across different historical periods and varying target scales.

Based on the model-generated annual vector boundaries, this study further calculated multi-dimensional attributes, including geometric, topographic, and uncertainty metrics, to construct a standardized inventory dataset. All vector data were first projected onto the Asia North Albers Equal Area Conic coordinate system to accurately calculate the area and perimeter of the glacial lakes. Subsequently, the mean elevation value from the 10 m DEM within each glacial lake extent was extracted as the lake surface elevation. To quantify the boundary positioning errors introduced by the mixed-pixel effect in remote sensing imagery, this study followed the error assessment method used by Wang et al. (2020) in the Glacial lake inventory of high-mountain Asia, calculating the area uncertainty of individual glacial lakes based on the boundary pixel theory of Hanshaw and Bookhagen (2014). This method assumes that the extraction errors of boundary pixels follow a normal distribution, taking one standard deviation (1 sigma) as the confidence interval (i.e., assuming approximately 68.72% of the peripheral boundary pixels are subject to error). The formula for calculating area uncertainty is as follows:

$$Uncert = \frac{P}{G} \times \frac{G^2}{2} \times 0.6872 \quad (2)$$

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

where P is the perimeter of the glacial lake (m); G is the spatial resolution of the remote sensing imagery (m); P/G represents the total number of mixed pixels falling along the glacial lake boundary; $G^2/2$ represents the maximum theoretical error area for a single boundary pixel (i.e., half the area of a pixel); and 0.6872 is the correction coefficient at the 1σ confidence level.

5. Technical Validation

5.1 Accuracy Assessment and Uncertainty Analysis

Quantitative evaluation of the models on independent test and validation sets (Table 2) indicates that Mask R-CNN exhibits performance limitations, constrained by complex background interference and the difficulty of identifying small targets in high-alpine mountainous regions. Although it maintains a high recall (0.9559) on the test set, its precision is only 0.8637. This suggests that traditional feature networks are more susceptible to false positives and boundary blurring when dealing with complex backgrounds and small glacial lakes. In contrast, the proposed GLA-RCNN achieves improvements across all metrics; its precision on the test set increases to 0.8803 and recall reaches 0.9563, with a balance between precision and recall. Ultimately, GLA-RCNN yields the highest F1 score of 0.9167 and an IoU of 0.8462 on the independent test set (while also reaching an F1 of 0.9236 and an IoU of 0.8580 on the validation set, demonstrating robust generalization stability and limited overfitting). This performance not only significantly surpasses that of mainstream U-Net and DeepLab V3 but also quantitatively validates the contribution of the CBAM attention mechanism in filtering out shadow noise interference and precisely restoring the topological structure of contiguous water bodies. Qualitative visual assessments further confirm that the improved model exhibits excellent robustness in small target detection and spectral anomaly processing within independent regions (Figure 5).

Considering that long-term glacial lake monitoring spans multiple sensor generations, this study evaluated the performance of GLA-RCNN across multiple spatial resolutions (10 m, 15 m, and 30 m) (Table 3). The model performed best on the test set of 10 m high-resolution images, with F1 and IoU values of 0.9485 and 0.9021, respectively. As the spatial resolution degraded to 30 m, physical limitations such as boundary mixed-pixel effects and reduced geometric information led to an expected decline in precision. However, recall remained above 0.95 across all scales. This pattern of high and stable recall indicates that even in historical coarse-resolution imagery, GLA-RCNN maintains a robust capacity for target capture, reducing the omission of small glacial lakes. Overall, the model maintained an F1 score exceeding 0.91 on both the test and validation sets at a 30 m resolution, providing algorithmic support for generating a long-term glacial lake evolution dataset from 1990 to 2025.

Table 2. Comparison of glacial lake recognition accuracy across different deep learning models.

Model	Test F1	Test IoU	Test Precision	Test Recall	Val F1	Val IoU

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

Open Access Earth System Science Data Discussions

GLA-RCNN	0.9167	0.8462	0.8803	0.9563	0.9236	0.8580
Mask R-CNN	0.9074	0.8305	0.8637	0.9559	0.9207	0.8530
U-Net	0.9023	0.8219	0.8994	0.9052	0.9130	0.8399
DeepLab V3	0.8476	0.7355	0.8292	0.8668	0.8620	0.7574

Table 3. Performance comparison of GLA-RCNN at different spatial resolutions

Resolution	Test F1	Test IoU	Test Precision	Test Recall	Val F1	Val IoU
10 m	0.9485	0.9021	0.9359	0.9615	0.9503	0.9054
15 m	0.9280	0.8656	0.9064	0.9506	0.9304	0.8698
30 m	0.9167	0.8462	0.8803	0.9563	0.9236	0.8580

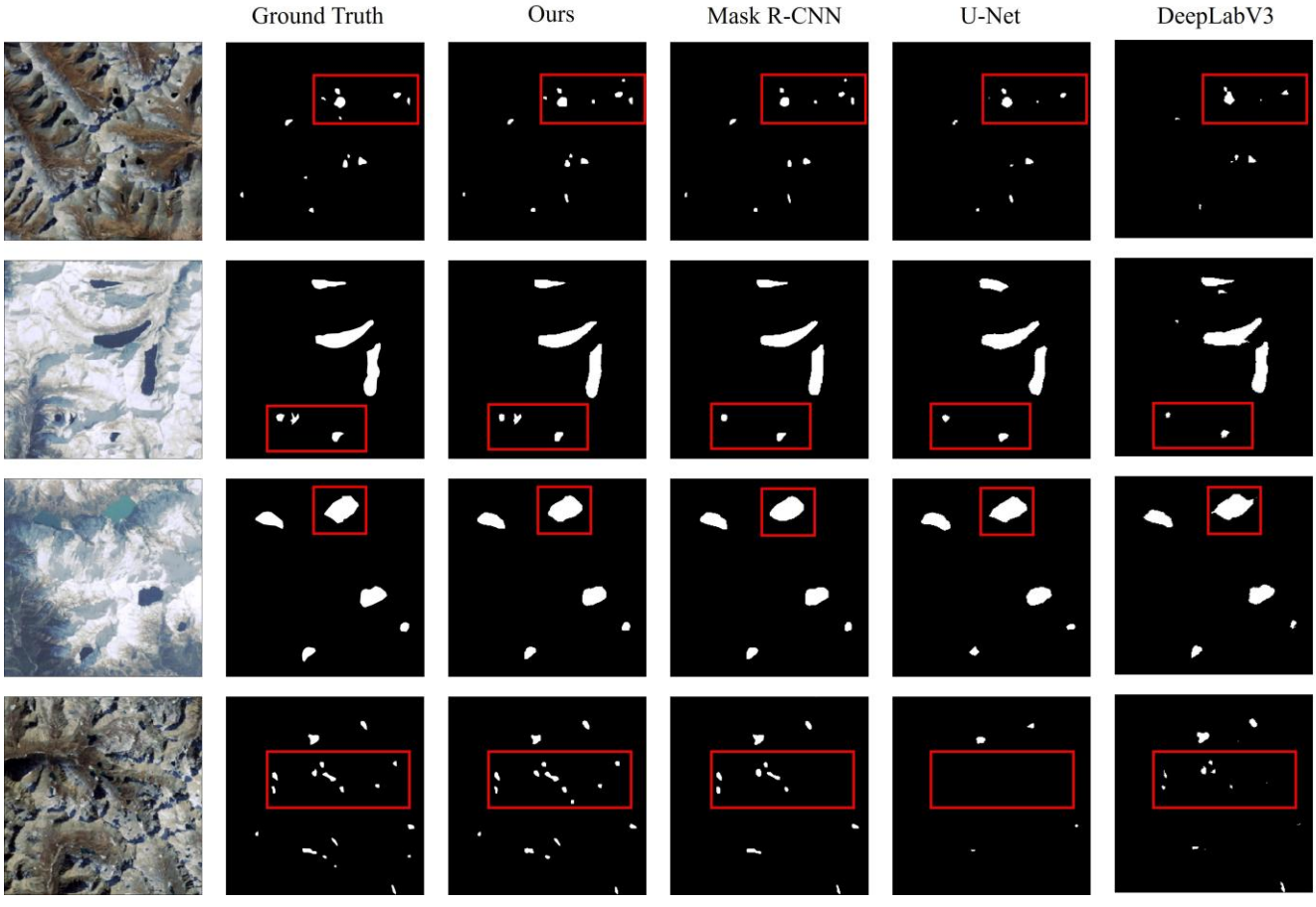

Figure 5. Comparison of results for glacial lake recognition across different models

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026


Open Access Earth System Science Data Discussions

315 **5.2 Geometric Accuracy Validation and Uncertainty Quantification**

316 Building on the model validation, to further evaluate the geospatial geometric accuracy and long-term cross-sensor
317 consistency of the final dataset, three key temporal nodes were selected as baseline validation years: 1991 (representing
318 Landsat 5, 30 m), 2013 (representing Landsat 8 pan-sharpened imagery, 15 m), and 2016 (representing Sentinel-2, 10 m).
319 Using contemporaneous remote sensing imagery and Google Earth images, with the Hi-MAG database (Chen et al., 2021) and
320 the Glacial lake inventory of high-mountain Asia (Wang et al., 2020) as references, manual fine-scale delineation was
321 performed for the vast majority of visible glacial lakes in the study area. This process resulted in the construction of a high-
322 confidence ground truth dataset spanning multiple spatial resolutions.

323 Regarding the geometric accuracy of glacial lake outlines, this study conducted an annual linear regression analysis
324 between the automatically extracted areas and the manually delineated ground truth (Figure 6a–c). The results show high
325 agreement between the two across all three periods, with coefficients of determination (R^2) reaching 0.998 or higher, and
326 regression line slopes closely approaching 1:1 (1.035, 1.012, and 1.000, respectively). This indicates the model's strong
327 generalization capability when handling multi-source data extraction spanning 36 years, exhibiting no clear systematic bias of
328 area overestimation or underestimation. A further comparison of the root mean square error (RMSE) across different years
329 shows that as spatial resolution improves in remote sensing imagery, the extraction error exhibits a stepwise decline: from
330 **41,916.3 m²** in 1991, down to **9,818.9 m²** in 2013, and reaching a low value of **3,425.1 m²** in 2016 with the use of Sentinel-2
331 data. This progressive decrease supports the value of using high-resolution imagery to enhance the core accuracy of fine-scale
332 glacial lake mapping in complex alpine regions.

333 To further quantify the reliability of glacial lake area extraction, this study analyzed the nonlinear variation pattern of
334 relative uncertainty with respect to glacial lake area scales (Figure 6d–f). Statistical analysis shows that uncertainty depends
335 strongly on lake size: constrained by the mixed-pixel effect of medium-resolution optical remote sensing, when a glacial lake
336 area is small, boundary pixels account for a larger proportion of the total area, resulting in a higher relative uncertainty;
337 conversely, as the glacial lake area increases, the relative uncertainty exhibits a power-law decreasing characteristic.
338 Comparing the scatterplot distributions across the three periods of varying resolutions, it is apparent that although the goodness
339 of fit fluctuates slightly due to differences in the distribution and detection rate of micro glacial lake samples, the overall L-
340 shaped convergence trend remains highly consistent. When the glacial lake area exceeds **1 km²**, the uncertainty decreases
341 rapidly and stabilizes at a very low level. This pattern not only validates the physical rationality of estimating area uncertainty
342 based on boundary pixels—indicating that this dataset is highly robust in delineating medium-to-large glacial lakes—but also
343 reflects the important role of improved spatial resolution (from 1991 to 2016) in reducing peak uncertainty values for micro
344 glacial lakes.

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

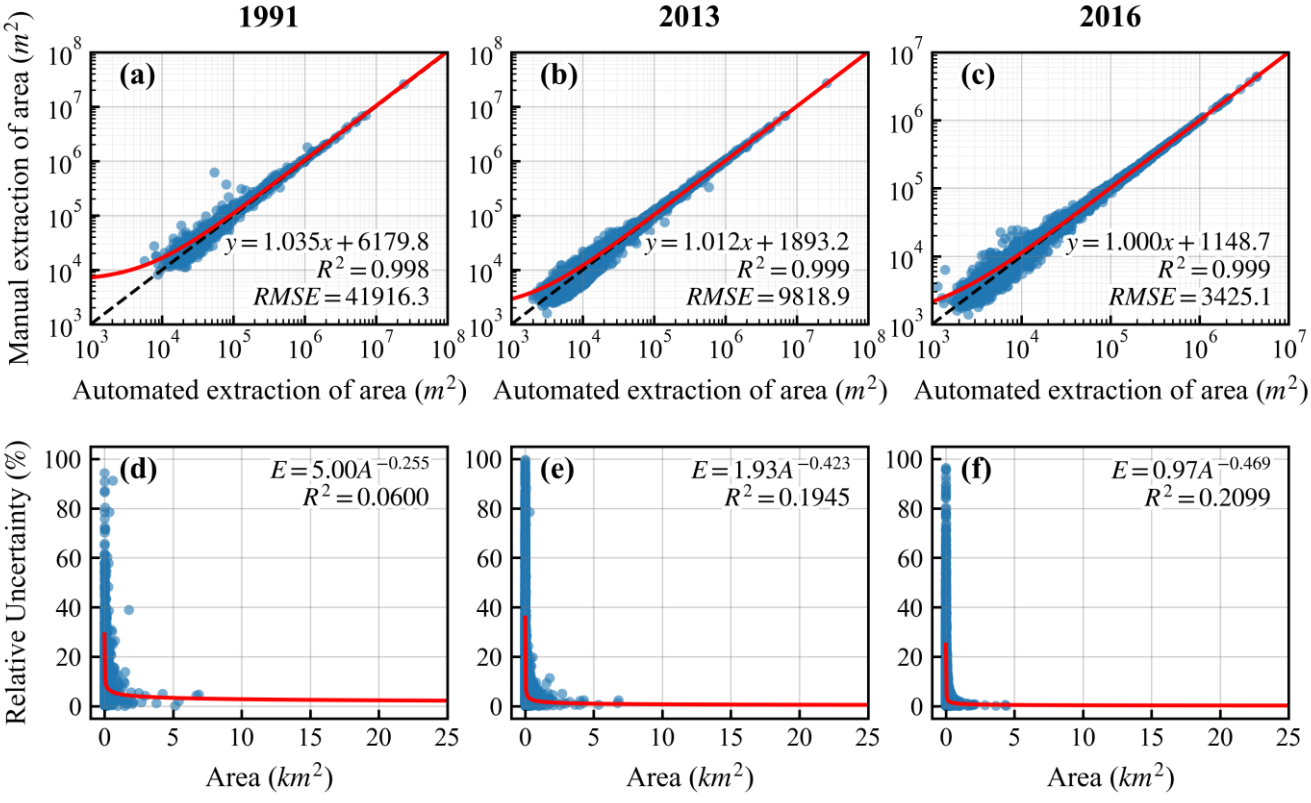

Figure 6. Geometric accuracy and size-dependent uncertainty analysis of glacial lakes. (a–c) Correlation assessment between automatically extracted areas and ground truth. The RMSE shows a stepwise decline as spatial resolution improves from 30 m to 10 m. (d–f) Nonlinear fitting of relative uncertainty against area. The L-shaped convergence of the scatters shows the amplification effect of extraction errors in the micro-glacial lake range.

Building on the aforementioned overall accuracy assessment, this study further conducts a detailed evaluation of the extraction results across different area classes for typical years (1991, 2013, and 2016) (Figure 7). The overall evaluation metrics indicate that despite variations in image characteristics across different periods and sensors, the automated extraction algorithm employed in this study maintains high stability. The global F1 scores for all years remain stable above 0.82 (reaching 0.887 and 0.874 in 2013 and 2016, respectively), demonstrating the high reliability of this dataset at regional spatiotemporal scales.

Detailed analysis of the error distribution characteristics across various area classes reveals that extraction errors depend strongly on lake size, primarily concentrated in the micro glacial lake range with areas less than 0.01 km² (the shaded area in Figure 7). A comparison between 1991 (Figure 7a) and 2016 (Figure 7c) shows that, constrained by the physical limits of the 30 m spatial resolution of early Landsat imagery, the omission error for micro glacial lakes in 1991 approaches 100%, with very few true lakes successfully identified. However, with the introduction of 10 m high-resolution Sentinel-2 imagery in 2016, the omission rate in this range drops to approximately 30%, and the number of successfully extracted true micro glacial lakes

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

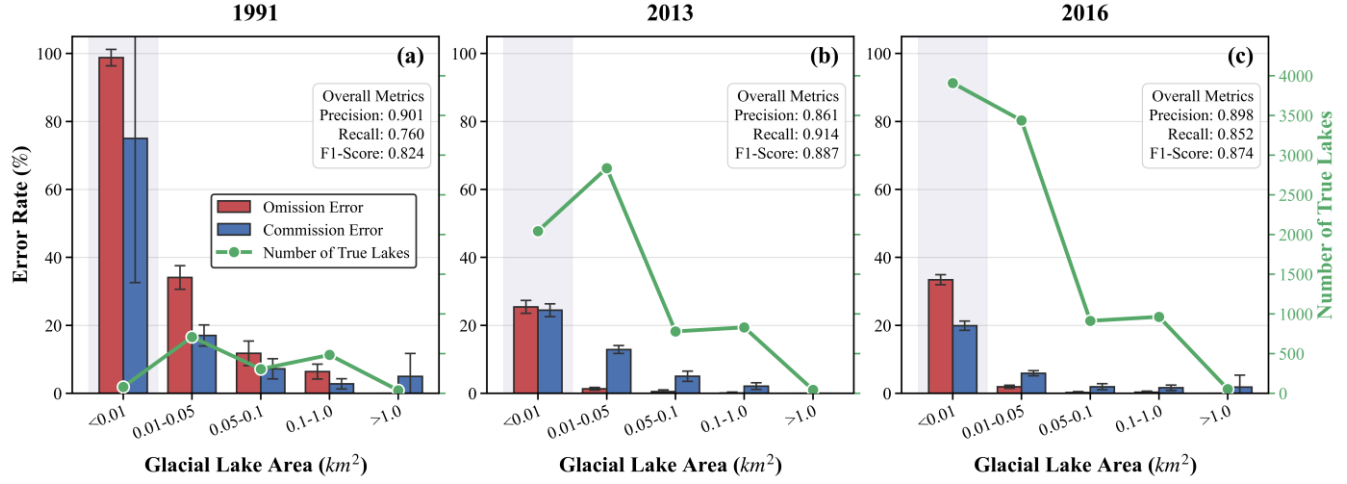

Figure 7. Distribution of omission and commission errors across different glacial lake area categories and overall accuracy evaluation in typical years (1991, 2013, and 2016). The gray shaded area highlights the high error distribution of micro glacial lakes (area < 0.01 km²), showing the limitations of early image resolution for micro-target recognition.

Figure 8 illustrates the extraction results of typical glacial lakes in the SETP generated by the proposed GLA-RCNN model, and compares them with meter-scale high-resolution historical imagery from Google Earth. Given that existing public validation datasets are predominantly constructed based on 30 m resolution imagery, they are insufficient to adequately verify the extraction performance for tiny glacial lakes (area < 0.01 km²) and under complex backgrounds. Therefore, this study introduces meter-scale Google Earth imagery as a secondary validation source to evaluate the model's extraction capability under challenging conditions. The model shows strong extraction performance for the following three typical mapping challenges:

(1) Robust recognition against optical heterogeneity of water bodies (Figures 8a, 8d): Influenced by physicochemical factors such as water depth and suspended mineral concentrations, alpine glacial lakes in the SETP frequently exhibit a wide color spectrum ranging from clear blue-green to turbid muddy brown. The GLA-RCNN model reduces the limitations of traditional fixed-threshold segmentation based on water indices, which are highly susceptible to spectral variations. It adaptively learns the deep semantic features of water bodies across varying turbidities, achieving detailed boundary delineation across a broad color spectrum.

(2) Boundary extraction under severe topographic shadow interference (Figures 8b, 8e): In frigid mountainous terrains with strong topographic relief, the deep shadows cast by steep slopes and the complex interplay of light and dark frequently induce "different objects with similar spectra" (spectral confusion) or "same objects with different spectra" phenomena between non-water areas and glacial lakes. In such backgrounds, the model demonstrates feature-screening capability and resistance to interference. By using spatial contextual relationships rather than isolated grayscale information, it distinguishes

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

water bodies from heavy shadows and background clutter, effectively suppressing the "false-positive" detection of pseudo-glacial lakes.

(3) Precise capture of tiny targets in highly fragmented geomorphology (Figures 8c, 8f): For tiny glacial lakes embedded within highly fragmented moraine ridges or bare rock landscapes, traditional medium- to low-resolution remote sensing is limited by targets being represented by mixed pixels. Comparisons with meter-scale high-resolution Google Earth imagery indicate that the model, augmented by multi-source high-resolution features, shows high sensitivity in capturing tiny glacial lakes (<0.01 km²). It not only successfully localizes these targets but also maintains geometrically detailed outputs even within extremely cluttered backgrounds.

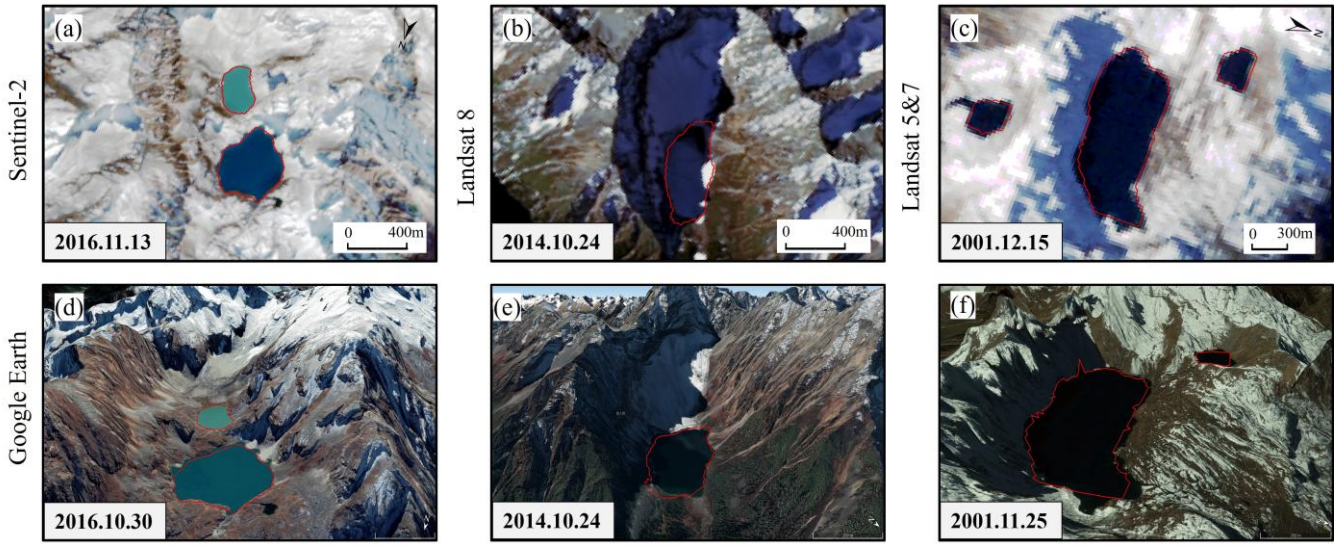

Figure 8. Comparison of GLA-RCNN extraction results under challenging conditions with meter-scale © Google Earth high-resolution imagery.

5.3 Cross-consistency and spatial boundary validation

To quantitatively evaluate the accuracy and consistency of this dataset at the spatial scale, this study conducted a cross-validation of the extraction results against the Hi-MAG database (Chen et al., 2021) and the Glacial lake inventory of high-mountain Asia (Wang et al., 2020). Three typical time points, 1990, 2013, and 2018, were selected for comparative analysis using quantitative metrics such as spatial Intersection over Union (IoU) and Recall (Table 4). It should be noted that due to severe cloud cover in historical optical imagery over high mountainous regions, there are differences in the image temporal selection strategies among the databases: the Glacial lake inventory of high-mountain Asia adopts a strategy of fusing multi-year images around the target year to obtain less-cloudy or cloud-free images near the end of the ablation season (summer and autumn); in contrast, both the Hi-MAG database and this study employ a single-year extraction strategy. To maintain temporal phase consistency in the long-term interannual evolution analysis, this study strictly limits the observation window to

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026


Open Access Earth System Science Data Discussions

September–December of each year. The aforementioned methodological differences and trade-offs serve as the basis for interpreting the various cross-validation metrics.

The cross-validation results reflect the combined impacts of temporal control strategies, image resolution, and sensor geometric accuracy on glacial lake inventories. In 1990, the total glacial lake area of our dataset was broadly consistent with the reference datasets, but the total number of patches was significantly lower, and the spatial agreement metrics were lower. This pattern reflects the trade-off made in this study between temporal consistency and the detection rate of micro targets. On the one hand, the multi-year fusion strategy of the High Asia database increases the probability of cloud-free observations for micro glacial lakes in summer and autumn; however, given the limited quality of early Landsat 5 imagery, this strategy tends to include seasonal snowmelt or non-water surface features from different years into glacial lake targets (Figures 9b and 9d). On the other hand, to maintain the single-year extraction standard of September–December, this study selected more images acquired closer to winter for regions with severe cloud and snow cover in 1990. This resulted in some true micro glacial lakes not being recognized by the model because they had lost typical water spectral features due to surface freezing or winter snow cover (Figure 9a). In addition, a small portion of early Landsat images suffered from inherent georeferencing errors and coarse spatial resolution, causing the glacial lake boundaries extracted from these images to be prone to pixel-level misalignments or morphological generalization when compared spatially with the reference datasets (Figure 9b), which lowered boundary overlap metrics such as the Intersection over Union (IoU) in 1990.

With the overall improvement in remote sensing image quality and sensor spatial positioning accuracy, the global recall of this dataset reached 0.865 in 2013, indicating that it covered the main targets in existing databases. Local detail comparisons (Figures 9e–h) further support this interpretation. Compared to the Hi-MAG database, the extraction results of this study show better fidelity along complex water body edges (Figures 9g and 9h) and reduced excessive over-generalization into adjacent non-water regions (Figure 9e). Meanwhile, Figure 9f shows the minor spatial misalignments that still exist between different datasets when dealing with the geometric offsets of remote sensing images.

In the 2018 comparison, for the matched glacial lake pairs with spatial overlap between the two sets, the mean Intersection over Union (mIoU) of this dataset reached 0.682, reflecting a high degree of individual boundary agreement. The apparent precision was 0.541, primarily influenced by the enhancement of the observation scale. Benefiting from the 10 m spatial resolution of Sentinel-2 imagery, this study identified 7,880 glacial lakes within the single year of 2018, with a total area of 501.79 km², exceeding the reference datasets based on 30 m resolution or multi-year mixed observations (4,479 lakes, 349.98 km²). Spatial overlap analysis and visual comparison (Figures 9i–l) indicate that the more than 3,000 additional patches identified in our dataset are mostly micro glacial lakes with areas less than 0.01 km² (Figure 9i). The high resolution not only enables this study to accurately separate closely spaced independent glacial lake clusters spatially (Figure 9l) but also captures subtle isolated patches missed by the reference datasets (Figures 9i and 9j). Furthermore, in the delineation of individual outlines, the extracted boundaries exhibit a close pixel-level agreement with the image pixels (Figure 9k). These newly discovered micro targets brought about by high resolution are typically not labelled in the reference datasets, thereby lowering the calculated precision. The changes in the aforementioned quantitative metrics and the enhancement of visual details

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

441 objectively demonstrate the potential value of high-resolution imagery and deep learning models in the refined monitoring of
442 micro glacial lakes and the extraction of complex boundaries.

443 **Table 4. Statistics of cross-consistency and spatial accuracy validation between our dataset and existing public glacial lake databases**
444 **in 1990, 2013, and 2018.**

Year	Reference Dataset	Ours Area (km²)	Ref. Area (km²)	Ours Count	Ref. Count	Overall Precision	Overall Recall	Instance mIoU	Matched IoU
1990	Glacial lake inventory of high-mountain Asia	301.975	313.136	2,125	4,104	0.4203	0.4058	0.4436	0.5435
2013	Hi-MAG	458.645	297.292	6,920	2,899	0.5607	0.8649	0.6496	0.7849
2018	Glacial lake inventory of high-mountain Asia	501.792	349.982	7,880	4,479	0.5419	0.7769	0.6825	0.8048

445

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

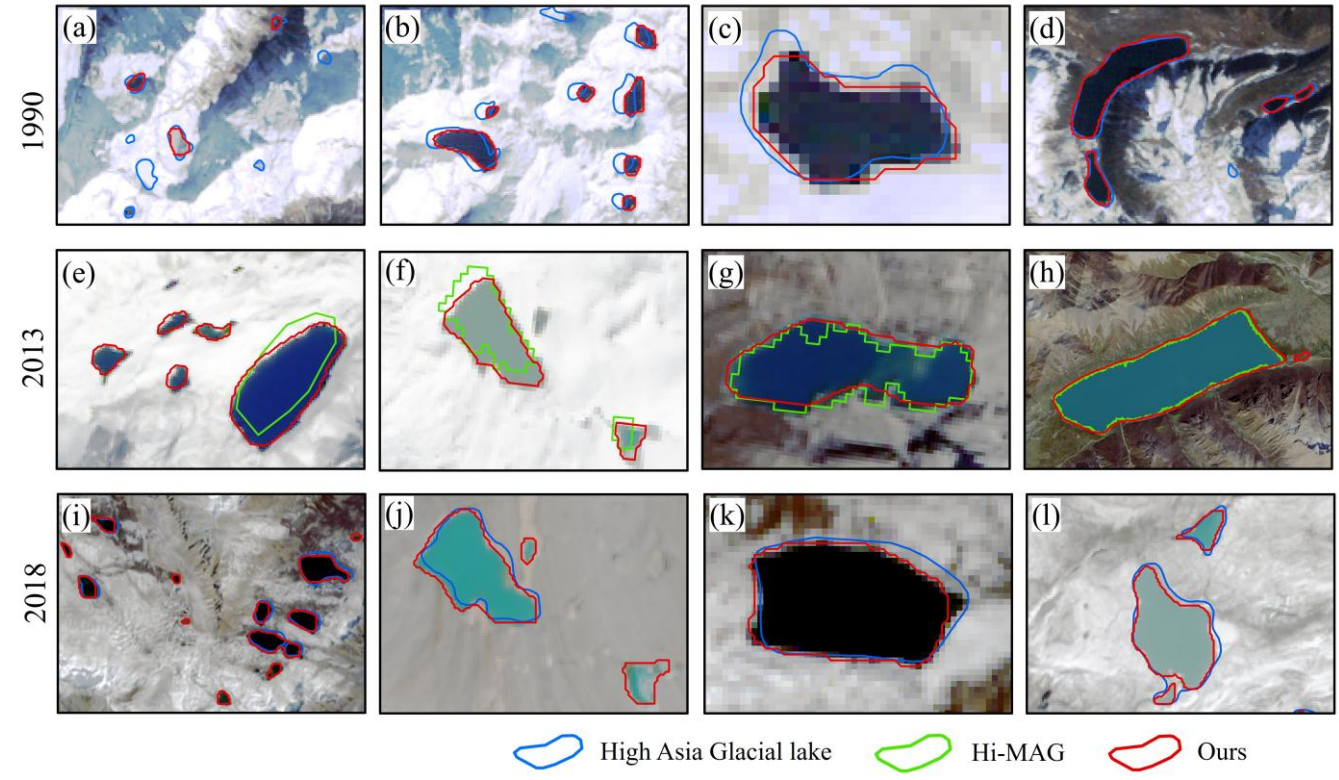

Figure 9. Visual comparison and validation of the extracted glacial lake results against existing reference databases (Glacial lake inventory of high-mountain Asia and Hi-MAG) across multiple typical regions in 1990, 2013, and 2018.

6. Data Records

SETP_GLI_1990–2025 is distributed as annual vector inventories in both GeoPackage (.gpkg) and ESRI Shapefile (.shp) formats. The GeoPackage version is recommended for most users because it preserves field names, coordinate reference information, and multiple layers more reliably than the Shapefile format (Table 5). The archive contains five folders: (1) /data_gpkg, including annual layers named SETP_GL_YYYY; (2) /data_shp, including annual Shapefile versions named SETP_GL_YYYY.shp; (3) /metadata, including annual statistics, field definitions, file checksums, and license information; (4) /study_area, including the study-area boundary; and (5) /validation, including manual validation polygons. All lake polygon geometries are provided in the Asia North Albers Equal Area Conic coordinate system. The Lon and Lat attributes provide WGS84 centroid coordinates in decimal degrees. Areas and perimeters were calculated in the same Albers equal-area projection and stored in the attribute table.

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

Open Access Earth System Science Data Discussions

461 **Table 5. Definitions and descriptions of the attribute fields in the glacial lake dataset (SETP_GLI_1990-2025).**

Field name	Definition	Unit / values	Notes
OBJECTID / FID	Software-managed feature number within each layer	integer	Not used for temporal tracking.
Track_ID	Stable object identifier for lakes matched across years	text	Facilitates the temporal tracking of individual lakes across the time series.
Lon	Longitude of lake centroid	degrees east	WGS 84.
Lat	Latitude of lake centroid	degrees north	WGS 84.
area_m2	Lake area	m²	Calculated in the Albers equal-area projection.
Peri_m	Lake perimeter	m	Calculated in the Albers equal-area projection.
elev_m	Mean lake-surface elevation	m	Derived from 10 m DEM zonal statistics.
Uncert_m2	Boundary-based area uncertainty	m²	Estimated from image resolution and polygon perimeter.
source_img	Source scene ID(s) used for lake delineation	text	Records the source scene ID(s) used for lake delineation.
res_m	Nominal mapping resolution	10/15/30 m	Used for uncertainty calculation.

462 **7. Assessment of Spatiotemporal Characteristics**

463 Long-term monitoring results indicate that glacial lakes in the SETP region exhibited a continuous expansion trend in
464 both number and area from 1990 to 2025. To eliminate statistical biases introduced by interannual climate fluctuations and
465 cross-sensor resolution disparities, this study conducted a comparative analysis of the long-term evolution of regional glacial
466 lakes—both in segmented periods and across the entire sequence—by combining the full inventory retained after sensor-
467 adaptive mapping thresholds with a standardized subset constrained by a uniform analysis threshold (≥ 0.01 km²) (Figure 10).
468 The full inventory preserves the maximum amount of reliable mapping information, whereas the ≥ 0.01 km² subset provides a
469 more consistent basis for cross-sensor trend comparison.

470 In the initial unconstrained time series, the evolution of glacial lakes demonstrated distinct staged characteristics. Between
471 1990 and 2012, regional glacial lakes showed a steady increase, with their number growing at an average rate of 55.0 lakes a^{-1},
472 and their area expanding at an average annual rate of 3.65 ± 1.12 km² a^{-1} ($p < 0.01$). Conversely, from 2016 to 2025, the mean
473 annual area expansion rate climbed to 5.95 ± 2.44 km² a^{-1} ($p < 0.05$), and the apparent numerical growth rate reached 203.7

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

lakes a^{-1}. Notably, a systematic "step-effect" was recorded on the data curve at the transition node between sensors (2012–2016). In-depth analysis reveals that this statistically discontinuous growth is largely attributable to the increase in observation resolution. The leap in spatial resolution from 30 m to 10 m significantly suppressed extraction uncertainties at the mathematical level. In 2012, the total area of the unconstrained glacial lakes was 339.02 ± 35.68 km² (with the area error ratio based on boundary pixels being approximately 10.5%). However, following the introduction of Sentinel-2 imagery in 2016, the total area of the unconstrained glacial lakes rose to 536.86 ± 23.79 km², and its uncertainty ratio converged to approximately 4.4%. This step-wise decrease in relative error supports the contribution of high-resolution data to improved geometric boundary fidelity.

Nevertheless, the improvement in observation resolution inevitably led to a sudden surge in the detection rate of minute water bodies. To decouple this statistical interference caused by the generational gap in spatial resolution, this study comparatively analyzed the evolutionary trend of the entire sequence (1990–2025) before and after applying the uniform 0.01 km² analysis threshold. The comparative results reveal two distinct geoscientific patterns. First, regarding the numerical trend, the linear fitting of the overall growth rate for the unconstrained original sequence tended to be overestimated due to the superimposition of the "step-effect" in the middle and later stages. After applying this threshold—taking 2016 as an example—a large number of newly identified minute patches, specifically detected by high-resolution imagery, were uniformly filtered out. This caused the total number of glacial lakes to drop sharply from 8,754 to 5,304 (a numerical reduction rate of 39.4%). Consequently, the observational scales of historical and recent sequences were statistically aligned, and the smoothed numerical curve exhibited a more continuous and physically logical progressive growth process. Second, concerning the area trend, the evolutionary trajectories of the two datasets demonstrated a high degree of consistency. Data show that in 2016, after removing nearly 40% of the minute glacial lakes by number, the total area only experienced a slight decrease from 536.86 ± 23.79 km² to 516.90 ± 20.16 km² (an area reduction rate of merely about 3.7%). In Figure 10, the total area trend lines before and after the threshold constraint are nearly parallel. The high degree of overlap in this trend not only indicates that minute glacial lakes make a limited contribution to the long-term evolution of the regional total water volume, but also substantiates a core conclusion through cross-validation: the accelerated expansion of the glacial lake area in the SETP region over the past decade (2016–2025), compared to the historical period (1990–2012), does not originate from a sensor-driven apparent increase caused by high-resolution observations. Instead, it is primarily controlled by the continuous expansion of medium-to-large glacial lakes and the coalescence of adjacent patches. This result indicates a sustained regional cryospheric and hydrological response to climate warming.

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

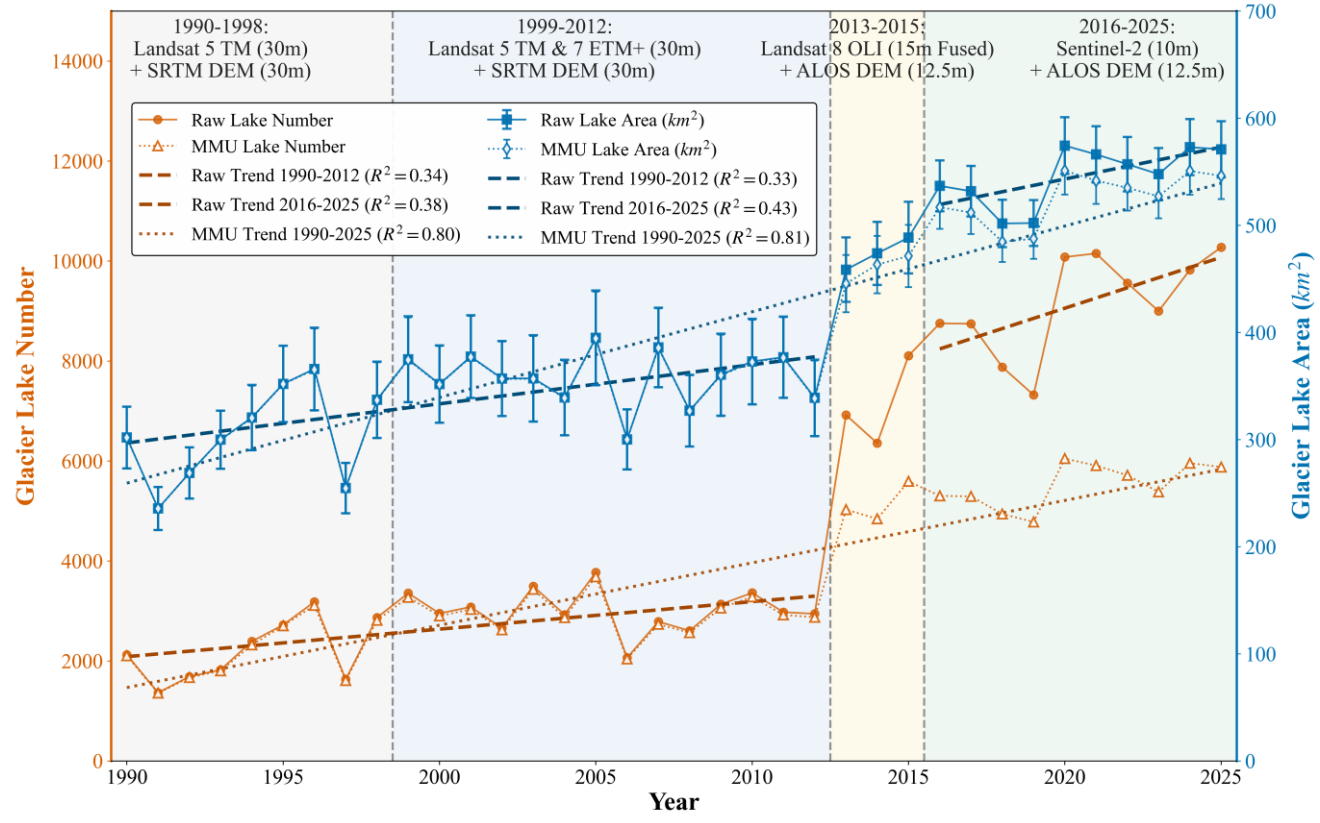

Figure 10. Statistics of changes in the number and area of glacial lakes in the SETP from 1990 to 2025.

After establishing the temporal evolution pattern of the overall accelerated expansion of regional glacial lakes, to further explore the spatial heterogeneity of this expansion trend, this study constructed a 25 km × 25 km spatial grid across the entire southeastern Tibetan Plateau based on the quality-controlled long-term dataset constrained by the uniform 0.01 km² analysis threshold. Combined with the Theil-Sen median trend estimation, this study quantified the area expansion rate of glacial lakes at the grid scale from 1990 to 2025 (Figure 11). The results show that the expansion of glacial lakes in the study area exhibits spatial clustering. The region-wide average expansion rate was 0.0306 km² a^{-1}, whereas in the expansion hotspots, the expansion rate generally exceeded 0.0550 km² a^{-1}, with the core extreme-value areas reaching 0.2042 km² a^{-1}. These high expansion-rate zones are primarily distributed in a belt along the eastern section of the Nyainqêntanglha Mountains and the high-altitude glacier margins of the Bomi-Zayu region, which further indicates spatially that the region is undergoing a marked landscape transition from snow/ice and land to liquid water bodies.

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

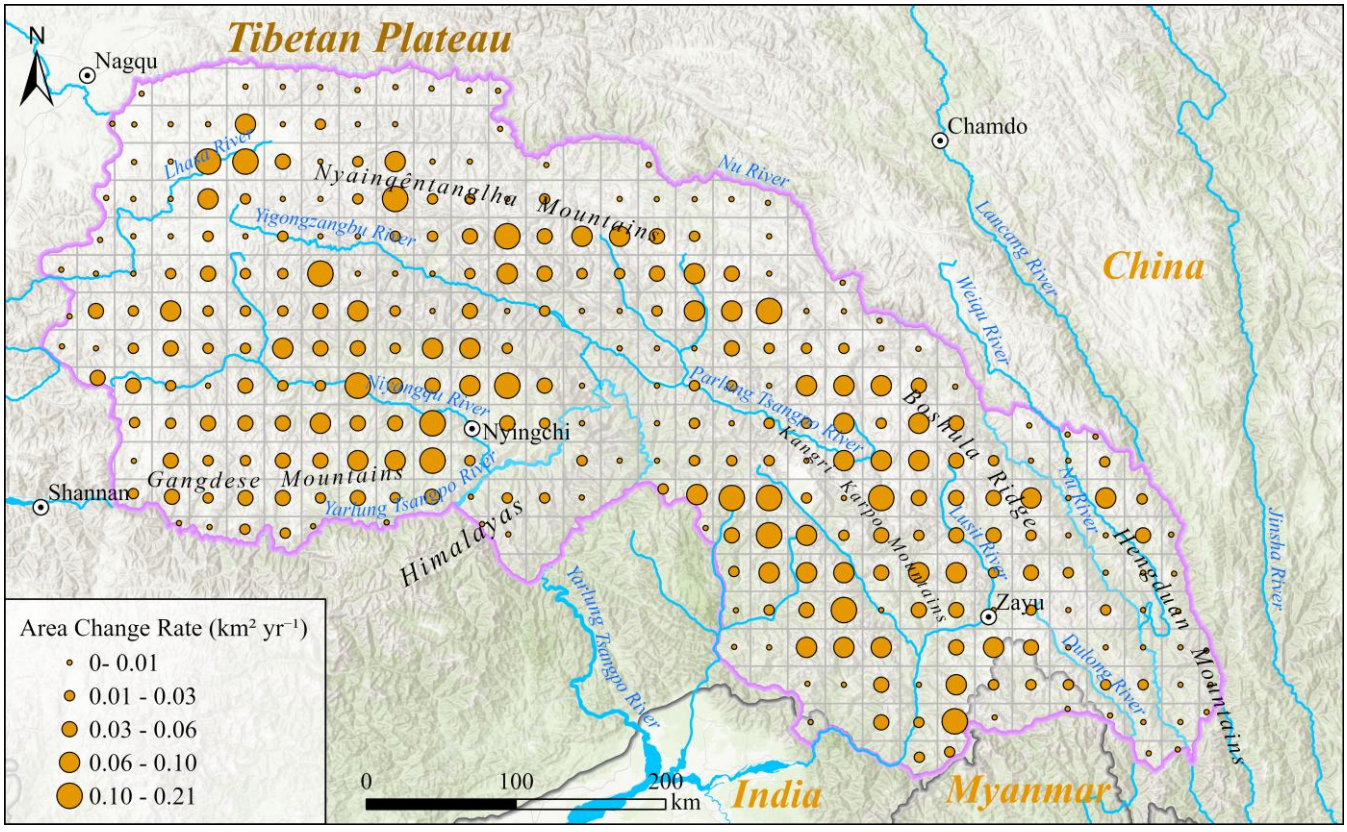

Figure 11. Spatial distribution of the glacial lake area expansion rate in the southeastern Tibetan Plateau from 1990 to 2025. The absolute expansion rates (km² a⁻¹) were calculated based on a 25 km × 25 km grid under the constraint of the uniform 0.01 km² analysis threshold, using the Theil-Sen median estimator combined with the Mann-Kendall significance test ($p < 0.05$).

8. Discussion

8.1 Advances of SETP_GLI and Comparison with Existing Datasets

The SETP_GLI dataset provides a 36-year annual glacial lake inventory for the southeastern Tibetan Plateau, with improvements over existing large-scale products in terms of spatiotemporal resolution, algorithmic robustness, and data standardization. As summarized in Table 6, this dataset addresses the limitations in refined regional hazard monitoring across three core dimensions:

First, regarding timeliness and observation frequency, it fills a critical monitoring gap in recent years. While existing large-scale datasets for High Mountain Asia (e.g., Wang et al., 2020; Chen et al., 2021) provide essential baselines, many of them end around 2018 or provide only intermittent observations, making it difficult to capture recent glacial lake dynamics under ongoing climate change (Shugar et al., 2020). By extending the observation sequence to 2025 with annual continuity,

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

this dataset provides a continuous record for analyzing the recent non-linear expansion and abrupt hydrological events of glacial lakes.

Second, concerning spatial resolution and micro-target detection, it addresses limitations in small-lake detection. Existing large-scale products predominantly rely on 30 m Landsat imagery, which is limited by severe mixed-pixel effects for lakes smaller than 0.01 km² (Zhang et al., 2022). Benefiting from the introduction of 10 m Sentinel-2 imagery, the sensor-adaptive mapping threshold of this dataset is as low as approximately 0.001 km² for recent imagery, allowing more visually confirmed micro-lakes to be retained. For temporal trend assessment, these micro-lakes are further evaluated using the standardized $\geqslant$ 0.01 km² subset to reduce cross-sensor detectability bias. Qualitative visual comparisons (Figure 12 and Figure 13) demonstrate that the GLA-RCNN framework, optimized for alpine topography with its embedded CBAM attention mechanism, reduces interference from mountain shadows, high-turbidity water, and supraglacial debris (Zhang et al., 2021; Zhao et al., 2018). It captures a large number of micro-lakes omitted in previous inventories, which are often important early-stage morphological indicators of hazard development (Watson et al., 2017).

Finally, regarding data standardization, it introduces polygon-level uncertainty quantification. Unlike previous inventories that typically provide only single geometric parameters, the SETP_GLI dataset integrates a boundary-pixel-based area uncertainty indicator. This responds to the need for standardized error estimates in benchmark datasets and helps ensure that the dataset supports reproducible dynamic evolution analyses and provides data support for disaster risk assessment (Emmer et al., 2022; Nie et al., 2017).

Table 6. Comparison between the SETP_GLI_1990-2025 dataset and other mainstream glacial lake inventories.

Comparison Dimension	This Study	Wang et al. (2020)	Chen et al. (2021)	Key Improvements
Temporal Span	1990–2025 (36 annual layers)	1990 and 2018	2008–2017	Fills the data gap for the recent warming period (2019–2025).
Temporal Resolution	Annual	Two epochs	Annual	Enables continuous monitoring of rapid lake expansion events.
Spatial Resolution	10 m (Sentinel-2) / 30 m	30 m (Landsat)	30 m (Landsat)	10 m resolution improves the detection of tiny glacial lakes (<0.01 km²).
Methodology	GLA-RCNN	Deep Learning (U-Net based)	Manual / Semi-automated	CBAM module effectively suppresses noise from mountain shadows and turbid water.

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

Attributes	Full (area, perimeter, elevation, uncertainty)	Area, Perimeter	Area, Perimeter	Includes elevation and uncertainty quantification for reliable risk assessment.
Thresholds	Sensor-adaptive mapping threshold: 0.001-0.0054 km ²; trend-analysis threshold: ⩾ 0.01 km²	>0.0054 km² (6 pixels)	>0.0081 km² (9 pixels)	Balances micro-lake retention with standardized cross-sensor trend analysis.

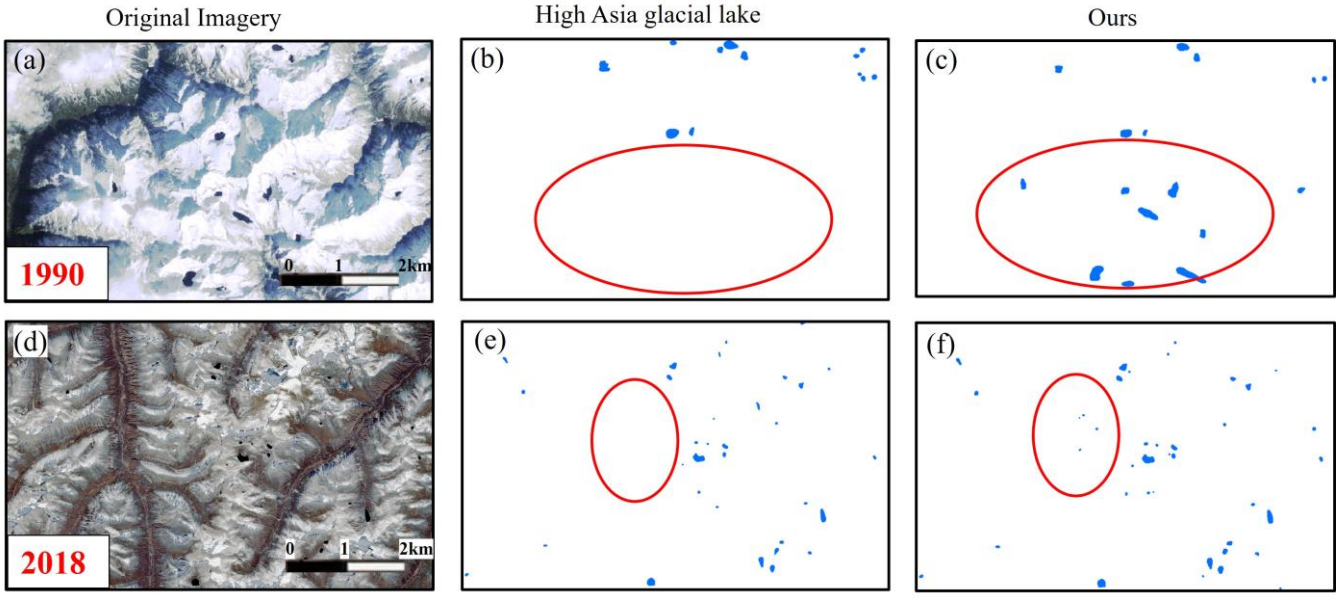

Figure 12. Spatial comparison between the proposed dataset and the High Asia Glacial Lake Database (Wang et al., 2020).

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

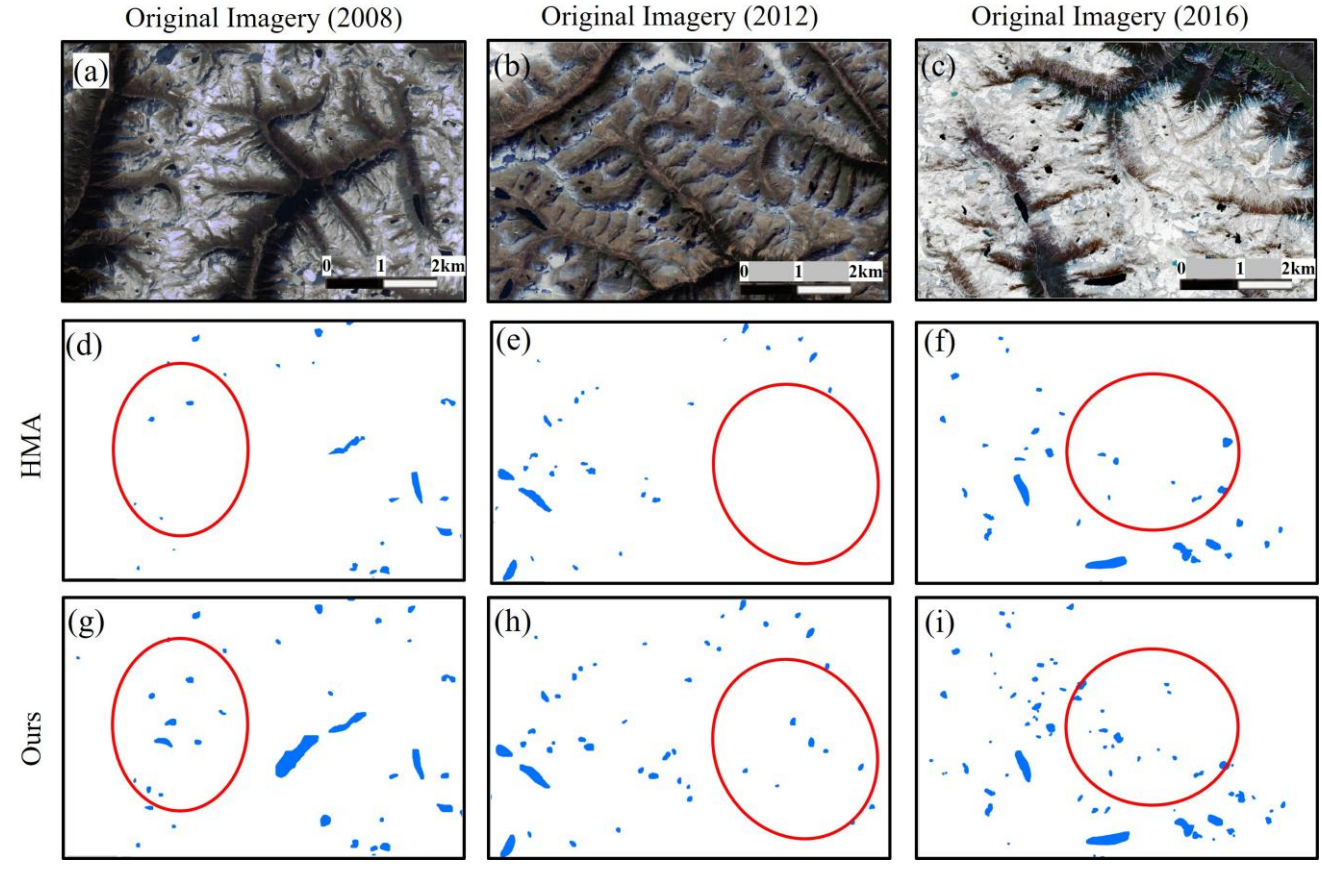

Figure 13. Spatial comparison between the proposed dataset and the High Mountain Asia Glacial Lake Inventory (Hi-MAG) (Chen et al., 2021).

8.2 Influence of Sensor Transition

In long-term glacial lake inventories, the transition between sensors with different spatial resolutions presents a key methodological challenge. The use of 10 m Sentinel-2 imagery has significantly improved data completeness; however, this jump in observational scale inevitably introduces a methodological "step effect" or discontinuity in lake-number statistics (Zhang et al., 2022). Therefore, direct comparisons of total lake numbers between the Landsat and Sentinel-2 periods may overestimate the actual rate of lake proliferation. For long-term trend analysis, we recommend that users apply a standardized analysis threshold ($\geqslant$ 0.01 km² in this study) or conduct separate analyses within Landsat and Sentinel-2 periods to mitigate statistical biases driven by generational sensor shifts (Nie et al., 2017).

8.3 Potential Applications

This dataset is designed to provide foundational support for multidisciplinary research in the High Mountain Asia cryosphere. Specific application scenarios include: (1) Assessing cryospheric change, such as quantifying the response

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

563 thresholds of alpine glacial lakes to climate change (Yao et al., 2022); (2) Initializing hydrological models and analyzing
564 climate forcing, exploring the combined effects of "climate forcing-topographic filtering" on lake evolution (Brun et al., 2017);
565 (3) Preliminary sensitivity screening for Glacial Lake Outburst Floods (GLOFs) (Veh et al., 2023; Zheng et al., 2021); and (4)
566 Identifying candidate lakes for high-precision UAV mapping or field investigations. Since each polygon is accompanied by
567 source-sensor information and area uncertainty estimates, users can flexibly filter subsets according to their research objectives
568 and confidence requirements.

569 **8.4 Limitations and Uncertainties**

570 Despite rigorous quality control, several limitations remain. First, detectability of micro-lakes is not constant over time
571 due to the transition between sensors and the limited spatiotemporal coverage of early Landsat 5/7 archives. Second, although
572 images were primarily selected from the late ablation season, differences in actual acquisition dates and seasonal water-level
573 fluctuations may introduce additional uncertainty. Third, persistent snow cover, lake freezing, deep topographic shadows, and
574 supraglacial debris may still cause minor omissions or boundary displacements (Li et al., 2022). Fourth, the provided area
575 uncertainty mainly quantifies boundary-related mixed-pixel errors and does not encompass all possible sources, such as
576 seasonal hydrological variability or georeferencing residuals (Nie et al., 2017). Finally, in the absence of high-precision
577 bathymetry and field-survey information, this dataset should not be used directly for engineering-scale GLOF modeling or
578 emergency planning.

579 **8.5 Future Updates**

580 To improve data accuracy, future iterations of the dataset will focus on expanding data sources and strengthening
581 empirical validation. First, we plan to incorporate Sentinel-1 Synthetic Aperture Radar (SAR) imagery, using its reduced
582 sensitivity to cloud cover to overcome the optical data gaps common during the monsoon season in the SETP, thereby
583 improving year-round monitoring capacity. Second, comprehensive ground-truth validation and Unmanned Aerial Vehicle
584 (UAV) field surveys will be conducted in critical regions to acquire high-precision data on glacial lake boundaries and
585 bathymetry (water depths).

586 **9. Conclusions**

587 To address the lack of long-term, high-resolution glacial lake inventories in the SETP, this study integrated Landsat and
588 Sentinel-2 imagery and developed the GLA-RCNN deep learning model embedded with a convolutional block attention
589 mechanism. We release a 10–30 m annual glacial lake inventory dataset for this region spanning 1990–2025. The main
590 conclusions are as follows:

591 (1) Detailed extraction of tiny glacial lakes and complex boundaries was achieved. Through rigorous cross-validation
592 against independent 30 m inventory data, the target recall reached 77.69% in the 2018 comparison with the Glacial lake

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

inventory of high-mountain Asia. For spatially matched co-occurring targets in that comparison, the matched IoU reached 0.8048. Combined with qualitative validation using meter-scale high-resolution imagery, the results suggest that the model reduces mountain-shadow misclassification in deeply incised gorges, showing robust segmentation when dealing with tiny glacial lakes (< 0.01 km²) and severe topographic shadow interference.

(2) Accelerated glacial lake evolution and sensor-dependent detectability were identified. Trend analysis based on the annual full-inventory sequence indicates that glacial lakes in the SETP exhibited a significant expansion trend over the past 36 years. Between 2016 and 2025, the apparent mean annual growth rates in number and area reached 203.7 lakes a^{-1} and 5.95 ± 2.44 km^2 a^{-1}, respectively, compared with 55.0 lakes a^{-1} and 3.65 ± 1.12 km^2 a^{-1} during 1990–2012. The increase in lake number is partly affected by sensor-dependent detectability, especially the improved capture of micro glacial lakes after the introduction of high-resolution Sentinel-2 imagery; in contrast, the area trend remains more robust after applying the standardized $\geqslant$ 0.01 km^2 analysis threshold, indicating that recent area expansion is mainly related to the growth and coalescence of medium-to-large glacial lakes.

(3) The data gap for fine-scale hydrological hazard assessment was reduced. This dataset provides not only high-precision vector boundaries but also multi-dimensional attributes with annual timeliness and uncertainty quantification. These foundational data help fill the critical gap in the long-term dynamic monitoring of small and medium-sized glacial lakes, offering essential scientific support for GLOF susceptibility assessments on the Tibetan Plateau, regional water resource management, and investigation of multi-sphere coupling mechanisms among climate, glaciers, and glacial lakes.

Usage Notes

Users should consider sensor-dependent detectability when analyzing long-term changes in lake number. The introduction of Sentinel-2 imagery after 2016 substantially improved the detection of small lakes, especially those smaller than 0.01 km^2. Therefore, trends in total lake area are generally more robust than trends in lake number. Long-term analyses of lake number should either be restricted to a standardized comparison threshold, such as the $\geqslant$ 0.01 km^2 threshold used in this study, or be conducted separately within Landsat and Sentinel-2 periods.

The dataset is suitable for regional glacial lake change assessment, hydrological model initialization, comparison with glacier change products, preliminary GLOF susceptibility screening, and selection of candidate lakes for detailed field or UAV surveys. It should not be used as a direct substitute for bathymetric surveys, dam-structure assessment, or engineering-scale GLOF modeling without additional local geomorphic, hydrological, and field information. For lakes affected by persistent shadow, debris cover, seasonal snow, ice cover, or strong seasonal water-level fluctuations, users are encouraged to inspect the original imagery and source-image metadata before drawing lake-specific conclusions.

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

Code availability

The source code used for image preprocessing, GLA-RCNN training and inference, rule-based post-processing, attribute calculation, uncertainty estimation, and validation is archived at Zenodo (AI-Geohazard, 2026) under DOI: https://doi.org/10.5281/zenodo.20555222. The repository contains the core Python source code, configuration files, environment specifications, and usage instructions required to reproduce the model training, inference, and post-processing workflow. Large remote-sensing images, manual labels, trained model weights, example image tiles, Jupyter notebooks, and full intermediate outputs are not included in the code repository because of data volume and licensing constraints; the released SETP_GLI inventory and validation vectors are provided separately with the dataset archive. Scripts for full model retraining are provided together with the training-sample format and parameter settings.

Author contributions

Hao Li and Jie Dou conceived the study and designed the experiments. Hao Li performed the remote-sensing data processing, glacial lake extraction, attribute calculation, uncertainty estimation, and formal analysis. Zihao Shi contributed to data organization and manuscript drafting. Hao Li, Zihao Shi, Xinjian Xiang, and Fange Ding contributed to visualization and figure preparation. Shun Dong and Jie Li contributed to data validation, result interpretation, and dataset quality checking. Timothy Kusky contributed to conceptual guidance and manuscript review and editing. Jie Dou supervised the study and revised the manuscript. All authors have read and approved the submitted manuscript.

Data availability

The SETP_GLI dataset (Li et al., 2026) is publicly available from the National Tibetan Plateau Data Center (TPDC) at https://doi.org/10.11888/Cryos.tpdc.303491 (direct link: https://data.tpdc.ac.cn/en/data/a3fd1136-d6df-44db-bc30-7d36899d61ff). The archive includes annual glacial lake inventory files from 1990 to 2025 in GeoPackage and Shapefile formats, metadata files, image-source records, attribute descriptions, validation polygons, and uncertainty estimates. The dataset is released under the Creative Commons Attribution 4.0 International License.

Competing interests

The authors declare that they have no competing interests.

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

Acknowledgements

We thank the USGS, ESA, NASA LP DAAC, ASF DAAC, Google Earth, and the National Tibetan Plateau Data Center for providing access to satellite imagery, DEM products, high-resolution reference imagery, and data archiving services. We also thank the developers and maintainers of the open-source geospatial and deep-learning tools used in the data-processing workflow.

Financial support

This research was supported by the National Natural Science Foundation of China (No. 42477170 and No. 42090054), the Natural Science Foundation of Hubei Province of China (Innovation Group Program: No. 2022CFA002 and No. 2024AFD358; No. 2025AFD435), and the Open Fund of National Engineering Research Center of Geographic Information System, China University of Geosciences (Grant No. 2023KFJJ04).

References

AI-Geohazard: douj888/GLA-RCNN-SETP-GLI: GLA-RCNN code for SETP_GLI v1.0.1, Zenodo [code], https://doi.org/10.5281/zenodo.20555222, 2026.

Bolch, T., Shea, J. M., Liu, S., Azam, F. M., Gao, Y., Gruber, S., Immerzeel, W. W., Kulkarni, A., Li, H., Tahir, A. A., Zhang, G., and Zhang, Y.: Status and change of the cryosphere in the extended Hindu Kush Himalaya region, in: The Hindu Kush Himalaya Assessment: Mountains, Climate Change, Sustainability and People, edited by: Wester, P., Mishra, A., Mukherji, A., and Shrestha, A. B., Springer, Cham, Switzerland, 209–255, https://doi.org/10.1007/978-3-319-92288-1_7, 2019.

Brun, F., Berthier, E., Wagnon, P., Kääb, A., and Treichler, D.: A spatially resolved estimate of High Mountain Asia glacier mass balances from 2000 to 2016, Nat. Geosci., 10, 668–673, https://doi.org/10.1038/ngeo2999, 2017.

Chen, F., Zhang, M., Guo, H., Allen, S., Kargel, J. S., Haritashya, U. K., and Watson, C. S.: Annual 30 m dataset for glacial lakes in High Mountain Asia from 2008 to 2017, Earth Syst. Sci. Data, 13, 741–766, https://doi.org/10.5194/essd-13-741-2021, 2021.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L.: DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, IEEE Trans. Pattern Anal. Mach. Intell., 40, 834–848, https://doi.org/10.1109/TPAMI.2017.2699184, 2018.

Crippen, R., Buckley, S., Agram, P., Belz, E., Gurrola, E., Hensley, S., Kobrick, M., Lavalle, M., Martin, J., Neumann, M., Nguyen, Q., Rosen, P., Shimada, J., Simard, M., and Tung, W.: NASADEM global elevation model: methods and progress, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLI-B4, 125–128, https://doi.org/10.5194/isprsarchives-XLI-B4-125-2016, 2016.

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

Dehecq, A., Gourmelen, N., Gardner, A. S., Brun, F., and Goldberg, D.: Twenty-first century glacier slowdown driven by mass loss in High Mountain Asia, Nat. Geosci., 12, 22–27, https://doi.org/10.1038/s41561-018-0271-9, 2019.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., and Bargellini, P.: Sentinel-2: ESA's optical high-resolution mission for GMES operational services, Remote Sens. Environ., 120, 25–36, https://doi.org/10.1016/j.rse.2011.11.026, 2012.

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., Diaz Moreno, A., Drenkhan, F., Fischer, M., Immerzeel, W. W., Izagirre, E., Joshi, R. C., Kougkoulos, I., Kuyakanon Knapp, R., Li, D., Majeed, U., Matti, S., Moulton, H., Nick, F., Piroton, V., Rashid, I., Reza, M., Ribeiro de Figueiredo, A., 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.

Fan, Y. and van den Dool, H.: A global monthly land surface air temperature analysis for 1948–present, J. Geophys. Res.-Atmos., 113, D01103, https://doi.org/10.1029/2007JD008470, 2008.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J.: The climate hazards infrared precipitation with stations – a new environmental record for monitoring extremes, Sci. Data, 2, 150066, https://doi.org/10.1038/sdata.2015.66, 2015.

Hanshaw, M. N. and Bookhagen, B.: Glacial areas, lake areas, and snow lines from 1975 to 2012: status of the Cordillera Vilcanota, including the Quelccaya Ice Cap, northern central Andes, Peru, The Cryosphere, 8, 359–376, https://doi.org/10.5194/tc-8-359-2014, 2014.

Harris, I., Osborn, T. J., Jones, P., and Lister, D.: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset, Sci. Data, 7, 109, https://doi.org/10.1038/s41597-020-0453-3, 2020.

Harrison, S., Kargel, J. S., Huggel, C., Reynolds, J. M., Shugar, D. H., Betts, R. A., Emmer, A., Glasser, N., Haritashya, U. K., Klimeš, J., Reinhardt, L., Schaub, Y., Wiltshire, A., Regmi, D., and Vilímek, V.: Climate change and the global pattern of moraine-dammed glacial lake outburst floods, The Cryosphere, 12, 1195–1209, https://doi.org/10.5194/tc-12-1195-2018, 2018.

He, K., Gkioxari, G., Dollár, P., and Girshick, R.: Mask R-CNN, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017, 2961–2969, https://doi.org/10.1109/ICCV.2017.322, 2017.

Hugonnet, R., McNabb, R., Berthier, E., Menounos, B., Nuth, C., Girod, L., Farinotti, D., Huss, M., Dussaillant, I., Brun, F., and Kääb, A.: Accelerated global glacier mass loss in the early twenty-first century, Nature, 592, 726–731, https://doi.org/10.1038/s41586-021-03436-z, 2021.

Immerzeel, W. W., Lutz, A. F., Andrade, M., Bahl, A., Biemans, H., Bolch, T., Hyde, S., Brumby, S., Davies, B. J., Elmore, A. C., Emmer, A., Feng, M., Fernández, A., Haritashya, U., Kargel, J. S., Koppes, M., Kraaijenbrink, P. D. A., Kulkarni, A. V., Mayewski, P. A., Nepal, S., Pacheco, P., Painter, T. H., Pellicciotti, F., Rajaram, H., Rupper, S., Sinisalo, A., Shrestha, A. B., Viviroli, D., Wada, Y., Xiao, C., Yao, T., and Baillie, J. E. M.: Importance and vulnerability of the world's water towers, Nature, 577, 364–369, https://doi.org/10.1038/s41586-019-1822-y, 2020.

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

Li, H., Dou, J., Kusky, T., Dong, S., Shi, Z., Li, J., Xiang, X., and Ding, F.: Annual glacial lake inventory dataset for the southeastern Tibetan Plateau from 1990 to 2025, National Tibetan Plateau Data Center [data set], https://doi.org/10.11888/Cryos.tpdc.303491, 2026.

Li, J., Sheng, Y., Luo, J., and Dong, P.: Automatic extraction of Himalayan glacial lakes with high-resolution satellite images using deep learning, IEEE Geosci. Remote Sens. Lett., 19, 1–5, https://doi.org/10.1109/LGRS.2021.3053738, 2022.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S.: Feature pyramid networks for object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 21–26 July 2017, 2117–2125, https://doi.org/10.1109/CVPR.2017.106, 2017.

Matthew, M. W., Adler-Golden, S. M., Berk, A., Richtsmeier, S. C., Levine, R. Y., Bernstein, L. S., Acharya, P. K., Anderson, G. P., Felde, G., Hoke, M. P., Ratkowski, A., Burke, H.-H., Kaiser, R. D., and Miller, D. P.: Status of atmospheric correction using a MODTRAN4-based algorithm, in: Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI, SPIE, 4049, 199–207, https://doi.org/10.1117/12.410341, 2000.

Maussion, F., Scherer, D., Mölg, T., Collier, E., Curio, J., and Finkelnburg, R.: Precipitation seasonality and variability over the Tibetan Plateau as resolved by the High Asia Reanalysis, J. Climate, 27, 1910–1927, https://doi.org/10.1175/JCLI-D-13-00282.1, 2014.

Nie, Y., Sheng, Y., Liu, Q., Liu, L., Liu, S., Zhang, Y., and Song, C.: A regional-scale assessment of Himalayan glacial lake changes using satellite observations from 1990 to 2015, Remote Sens. Environ., 189, 1–13, https://doi.org/10.1016/j.rse.2016.11.008, 2017.

Nie, Y., Pritchard, H. D., Liu, Q., Hennig, T., Wang, W., Wang, X., Liu, S., Nepal, S., Samyn, D., Hewitt, K., and Chen, X.: Glacier changes on the Tibetan Plateau over the past four decades, in: The Third Pole Environment, edited by: Wang, Y., Yao, T., and Xu, B., Springer, Cham, Switzerland, 123–145, https://doi.org/10.1007/978-3-030-65043-4_8, 2021.

Qayyum, N., Ghuffar, S., Ahmad, H. M., Yousaf, A., and Shahid, I.: Glacial lakes mapping using multi satellite PlanetScope imagery and deep learning, ISPRS Int. J. Geo-Inf., 9, 560, https://doi.org/10.3390/ijgi9090560, 2020.

RGI Consortium: Randolph Glacier Inventory – Version 7.0, National Snow and Ice Data Center [data set], https://doi.org/10.5067/f6jmovy5navz, 2023.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, edited by: Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., Springer, Cham, Switzerland, 234–241, https://doi.org/10.1007/978-3-319-24574-4_28, 2015.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., and Ziese, M.: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle, Theor. Appl. Climatol., 115, 15–40, https://doi.org/10.1007/s00704-013-0860-x, 2014.

Shean, D. E., Bhushan, S., Montesano, P., Rounce, D. R., Arendt, A., and Osmanoğlu, B.: A systematic, regional assessment of High Mountain Asia glacier mass balance, Front. Earth Sci., 7, 363, https://doi.org/10.3389/feart.2019.00363, 2020.

743 Shugar, D. H., Burr, A., Haritashya, U. K., Kargel, J. S., Watson, C. S., Kennedy, M. C., Bevington, A. R., Betts, R. A.,
744 Harrison, S., and Strattman, K.: Rapid worldwide growth of glacial lakes since 1990, Nat. Clim. Chang., 10, 939–945,
745 https://doi.org/10.1038/s41558-020-0855-4, 2020.
746 Sun, W., Chen, B., and Messinger, D. W.: Nearest-neighbor diffusion-based pan-sharpening for spectral distortion
747 minimization, Opt. Eng., 53, 013107, https://doi.org/10.1117/1.OE.53.1.013107, 2014.
748 Tadono, T., Ishida, H., Oda, F., Naito, S., Minakawa, K., and Iwamoto, H.: Precise global DEM generation by ALOS PRISM,
749 ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., II-4, 71–76, https://doi.org/10.5194/isprsannals-II-4-71-2014,
750 2014.
751 Veh, G., Korup, O., von Specht, S., Roessner, S., and Walz, A.: Enhanced glacial lake activity threatens communities in the
752 Third Pole, Nat. Commun., 14, 7973, https://doi.org/10.1038/s41467-023-43356-4, 2023.
753 Walder, J. S. and Costa, J. E.: Outburst floods from glacier-dammed lakes: the effect of mode of lake drainage on flood
754 magnitude, Earth Surf. Proc. Land., 21, 701–723, https://doi.org/10.1002/(SICI)1096-9837(199607)21:8%3C701::AID-
755 ESP615%3E3.0.CO;2-2, 1996.
756 Wang, S., Zhang, M., Li, Z., Wang, F., Li, H., Li, Y., and Huang, X.: Glacier area variation and climate change in the Chinese
757 Tianshan Mountains since 1960, J. Geogr. Sci., 21, 263–273, https://doi.org/10.1007/s11442-011-0843-8, 2011.
758 Wang, W., Yao, T., Yang, X., and Zhang, G.: Glacial lake area changes in High Mountain Asia during 1990–2020, Research,
759 2022, 9821275, https://doi.org/10.34133/2022/9821275, 2022.
760 Wang, X., Guo, X., Yang, C., Liu, Q., Wei, J., Zhang, Y., Liu, S., Zhang, Y., Jiang, Z., and Tang, Z.: Glacial lake inventory
761 of high-mountain Asia in 1990 and 2018 derived from Landsat images, Earth Syst. Sci. Data, 12, 2169–2182,
762 https://doi.org/10.5194/essd-12-2169-2020, 2020.
763 Wang, X., Chen, F., Zhang, M., and Tian, B.: Refined glacial lake extraction in a high-Asia region by deep neural network,
764 The Cryosphere, 18, 153–172, https://doi.org/10.5194/tc-18-153-2024, 2024.
765 Wangchuk, S. and Bolch, T.: Mapping of glacial lakes using Sentinel-1 and Sentinel-2 data and a random forest classifier:
766 strengths and challenges, Sci. Remote Sens., 2, 100008, https://doi.org/10.1016/j.srs.2020.100008, 2020.
767 Watson, C. S., Quincey, D. J., Carrivick, J. L., and Smith, M. W.: Ice cliff dynamics in the Everest region of the Central
768 Himalaya, Geomorphology, 278, 238–251, https://doi.org/10.1016/j.geomorph.2016.11.017, 2017.
769 Wen, J., Wang, G., Zhou, R., Li, R., Zhaxi, S., and Bai, M.: Seasonal variation in vertical structure for stratiform rain at Mêdog
770 site in southeastern Tibetan Plateau, Remote Sens., 16, 1230, https://doi.org/10.3390/rs16071230, 2024.
771 Westoby, M. J., Glasser, N. F., Hambrey, M. J., Brasington, J., Reynolds, J. M., and Hassan, M. A. A. M.: Reconstructing
772 historic glacial lake outburst floods through numerical modelling and geomorphological assessment: extreme events in
773 the Himalaya, Earth Surf. Proc. Land., 39, 1675–1692, https://doi.org/10.1002/esp.3617, 2014.
774 Wilson, R., Glasser, N. F., Reynolds, J. M., Harrison, S., Anacona, P. I., Schaefer, M., and Shannon, S.: Glacial lakes of the
775 Central and Patagonian Andes, Global Planet. Change, 162, 275–291, https://doi.org/10.1016/j.gloplacha.2018.01.004,
776 2018.

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S.: CBAM: convolutional block attention module, in: Computer Vision – ECCV 2018, edited by: Ferrari, V., Hebert, M., Sminchisescu, C., and Weiss, Y., Springer, Cham, Switzerland, 3–19, https://doi.org/10.1007/978-3-030-01234-2_1, 2018.

Worni, R., Stoffel, M., Huggel, C., Volz, C., Casteller, A., and Luckman, B.: Analysis and dynamic modeling of a moraine failure and glacier lake outburst flood at Ventisquero Negro, Patagonian Andes (Argentina), J. Hydrol., 444–445, 134–145, https://doi.org/10.1016/j.jhydrol.2012.02.013, 2012.

Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., Allen, R. G., Anderson, M. C., Belward, A. S., Cohen, W. B., Dwyer, J., Erb, A., Gao, F., Griffiths, P., Helder, D., Hermosilla, T., Hipple, J. D., Hostert, P., Hughes, M. J., Huntington, J., Johnson, D. M., Kennedy, R., Kilic, A., Li, Z., Lymburner, L., McCorkel, J., Pahlevan, N., Scambos, T. A., Schaaf, C., Schott, J. R., Sheng, Y., Storey, J., Vermote, E., Vogelmann, J., White, J. C., Wynne, R. H., and Zhu, Z.: Current status of Landsat program, science, and applications, Remote Sens. Environ., 225, 127–147, https://doi.org/10.1016/j.rse.2019.02.015, 2019.

Xu, D.: Characteristics of debris flow caused by outburst of glacial lake in Boqu River, Xizang, China, 1981, GeoJournal, 17, 569–580, https://doi.org/10.1007/BF00209443, 1988.

Yang, K., Wu, H., Qin, J., Lin, C., Tang, W., and Chen, Y.: Recent climate changes over the Tibetan Plateau and their impacts on energy and water cycle: a review, Global Planet. Change, 112, 79–91, https://doi.org/10.1016/j.gloplacha.2013.12.001, 2014.

Yao, T., Xue, Y., Chen, D., Chen, F., Thompson, L., Cui, P., Koike, T., Lau, W. K.-M., Lettenmaier, D., Mosbrugger, V., Zhang, R., Xu, B., Dozier, J., Gillespie, T., Gu, Y., Kang, S., Piao, S., Sugimoto, S., Ueno, K., Wang, L., Wang, W., Zhang, F., Sheng, Y., Guo, W., Ailikun, Yang, X., Ma, Y., and Shen, S. S. P.: Recent Third Pole's rapid warming accompanies cryospheric melt and water cycle intensification and interactions between monsoon and environment: multidisciplinary approach with observations, modeling, and analysis, Bull. Am. Meteorol. Soc., 100, 423–444, https://doi.org/10.1175/BAMS-D-17-0057.1, 2019.

Yao, T., Bolch, T., Chen, D., Gao, J., Immerzeel, W., Piao, S., Su, F., Thompson, L., Wada, Y., Wang, L., Wang, T., Wu, G., Xu, B., Yang, W., Zhang, G., and Zhao, P.: The imbalance of the Asian water tower, Nat. Rev. Earth Environ., 3, 618–632, https://doi.org/10.1038/s43017-022-00299-4, 2022.

Zeitler, P. K., Meltzer, A. S., Brown, L., Kidd, W. S. F., Lim, C., and Enkelmann, E.: Tectonics and topographic evolution of Namche Barwa and the eastern Himalayan syntaxis, Geol. Soc. Am. Spec. Pap., 507, 23–58, https://doi.org/10.1130/2014.2507(02), 2014.

Zemp, M., Frey, H., Gärtner-Roer, I., Nussbaumer, S. U., Hoelzle, M., Paul, F., Haeberli, W., Denzinger, F., Ahlstrøm, A. P., Anderson, B., Bajracharya, S., Baroni, C., Braun, L. N., Cáceres, B. E., Casassa, G., Cobos, G., Dávila, L. R., Delgado Granados, H., Demuth, M. N., Espizua, L., Fischer, A., Fujita, K., Gadek, B., Ghazanfar, A., Hagen, J. O., Holmlund, P., Karimi, N., Li, Z., Pelto, M., Pitte, P., Popovnin, V. V., Portocarrero, C. A., Prinz, R., Sangewar, C. V., Severskiy, I.,

https://doi.org/10.5194/essd-2026-452
Preprint. Discussion started: 9 July 2026

Sigurðsson, O., Soruco, A., Usubaliev, R., and Vincent, C.: Historically unprecedented global glacier decline in the early 21st century, J. Glaciol., 61, 745–762, https://doi.org/10.3189/2015JoG15J017, 2015.

Zhang, C., Zhang, X., Tang, Q., Chen, D., Huang, J., Wu, S., and Liu, Y.: Quantifying precipitation moisture contributed by different atmospheric circulations across the Tibetan Plateau, J. Hydrol., 628, 130517, https://doi.org/10.1016/j.jhydrol.2023.130517, 2024.

Zhang, G., Yao, T., Xie, H., Zhang, K., and Zhu, F.: An inventory of glacial lakes in the Third Pole region and their changes in response to global warming, Global Planet. Change, 131, 148–157, https://doi.org/10.1016/j.gloplacha.2015.05.013, 2015.

Zhang, M., Chen, F., and Tian, B.: An automated method for glacial lake mapping in High Mountain Asia using Landsat 8 imagery, ISPRS J. Photogramm. Remote Sens., 181, 196–215, https://doi.org/10.1016/j.isprsjprs.2021.09.016, 2021.

Zhang, M., Chen, F., Tian, B., and Liang, D.: Landsat- and Sentinel-derived glacial lake dataset in the China–Pakistan Economic Corridor from 1990 to 2020, Earth Syst. Sci. Data, 14, 5489–5512, https://doi.org/10.5194/essd-14-5489-2022, 2022.

Zhao, F., Long, D., Li, X., Huang, Q., and Han, P.: Boundary and topographic data of the Southeastern Tibetan Plateau (2000), National Tibetan Plateau/Third Pole Environment Data Center [data set], https://doi.org/10.1016/j.rse.2021.112853, 2022.

Zhao, H., Chen, F., and Zhang, M.: A systematic extraction approach for mapping glacial lakes in high mountain regions of Asia, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 11, 2780–2793, https://doi.org/10.1109/JSTARS.2018.2859841, 2018.

Zheng, G., Allen, S. K., Bao, A., Ballesteros-Cánovas, J. A., Huss, M., Zhang, G., Li, J., Yuan, Y., Jiang, L., Yu, T., Chen, W., and Stoffel, M.: The 2020 glacial lake outburst flood at Jinwuco, Tibet, The Cryosphere, 15, 3159–3175, https://doi.org/10.5194/tc-15-3159-2021, 2021.