ESSD-2025-649

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

Siyu Zhao et al.

In the following document, reviewer comments are presented in blue, and author responses are provided in black.

Reviewer #1

Comment 1:

The topic is interesting; authors found an indirect way to quantify the direct effects of glacier melting by determining the lake areas of Tibetan Platou. Nevertheless, the main idea is not new, there are previous studies in the topic, the MS can have its potentials. The built dataset is based on satellite images, and provides a better insight into the changing lake areas with higher spatial and temporal resolution than the existing data.

Response:

We would like to express our sincere gratitude to the reviewer for the comprehensive and insightful review of our manuscript. These recommendations have significantly enhanced the scope and scientific rigor of our work. We have carefully considered every comment and have made extensive revisions to the manuscript to address these points. Please find our detailed, point by point responses below.

Comment 2:

My first note is that TP as Tibetan Plateau in remote sensing usually referts to True Positive (and TP also appears in the MS later), so I suggest to find another abbreviation, e.g., TiP or something short and informative.

Response:

We thank the reviewer for pointing out this potential ambiguity. We agree that "TP" is commonly used as an acronym for "True Positive" in the context of accuracy assessment (confusion matrices), which could indeed cause confusion in this manuscript. To avoid this ambiguity while adhering to standard conventions in

regional geoscience literature, we have replaced the abbreviation "TP" with "QTP" (Qinghai-Tibet Plateau) throughout the revised manuscript. This abbreviation is widely recognized in the field and distinct from "True Positive."

Comment 3: Introduction

Regarding the introduction, I do not think that it is a good idea to start with the Tibetan Plateau, because it is not the only area in the world where there are high mountains, glaciers, which also suffer by the climate change, global warming. Thus, a wider approach, showing the readers, the global characteristic of the lake-formation and glacier melting would be beneficial. I suggest to add a new first paragraph with this global approach and to find the option to cite all researchers having preliminary knowledge and studies.

Response:

We greatly appreciate this insightful suggestion. We agree that framing the study within a broader global context significantly enhances the manuscript's relevance and readability. Action: We have rewritten the Introduction to begin with a global perspective on high-mountain environments. The new opening paragraph discusses the widespread impact of global warming on glacier-lake systems in major high-altitude regions (e.g., the Andes, the Alps, and the Arctic) before narrowing down to the QTP. The added text is as follows:

Alpine and polar lakes are critical components of the global ecosystem that are currently undergoing rapid transformations driven by accelerating global warming. Across major mountain environments such as the Andes, the Alps, the Himalaya, and the Arctic, rising temperatures have accelerated glacier retreat and permafrost degradation. These cryospheric changes have fundamentally altered hydrological cycles and driven widespread shifts in the formation, expansion, and seasonal variability of glacial lakes (Huggel et al., 2020; Shugar et al., 2020; Zemp et al., 2019). Acting as sentinels of climate change, these lakes integrate signals from glacier melt and shifting water balances while simultaneously influencing downstream freshwater availability, ecosystem stability, and hazards like glacial lake outburst floods

(Carrivick and Tweed, 2016; Emmer et al., 2021). Consequently, systematically characterising the long-term spatio-temporal dynamics of lake extent in these sensitive regions has become a central scientific objective for understanding the evolution of global lake systems and assessing broader ecological stability.

Comment 4:

The "Roof of the World" term is enough to mention once, and both the introduction and Study area description start with it.

Response:

We agree with the reviewer that the repetitive use of descriptive terms such as "Roof of the World" and "Water Tower of Asia" was unnecessary and distracting. We have removed these terms from the Study Area section (Section 2) to maintain a more objective and scientific tone focused strictly on geographical characteristics and lake distribution. The term is now mentioned only once in the Introduction to provide context for the region's significance.

Comment 5: Study area and Methods

Square kilometer can be abbreviated as km2

Response:

We are really sorry for our careless mistakes. We have standardized the unit usage and replaced all instances of "square kilometers" with the abbreviation "km^2" throughout the manuscript.

Comment 6:

In Fig 1, I did not see any lakes of 10-50 km2, a better method/color, background color/transparency should be chosen, in this way these lakes are just mentioned but cannot be seen.

Response:

We thank the reviewer for carefully checking the figures. We acknowledge that the original color scheme and background density made the smaller lakes (10-50 km^2)

difficult to distinguish. We have revised Figure 1 to improve visibility. Specifically, we increased the transparency of the background DEM layer to reduce visual clutter. We also adjusted the color of the 10-50 km² lakes to a more vibrant and contrasting tone to highlight their distribution against the terrain. The revised Figure 1 now clearly displays the spatial distribution of both large and medium sized lakes.
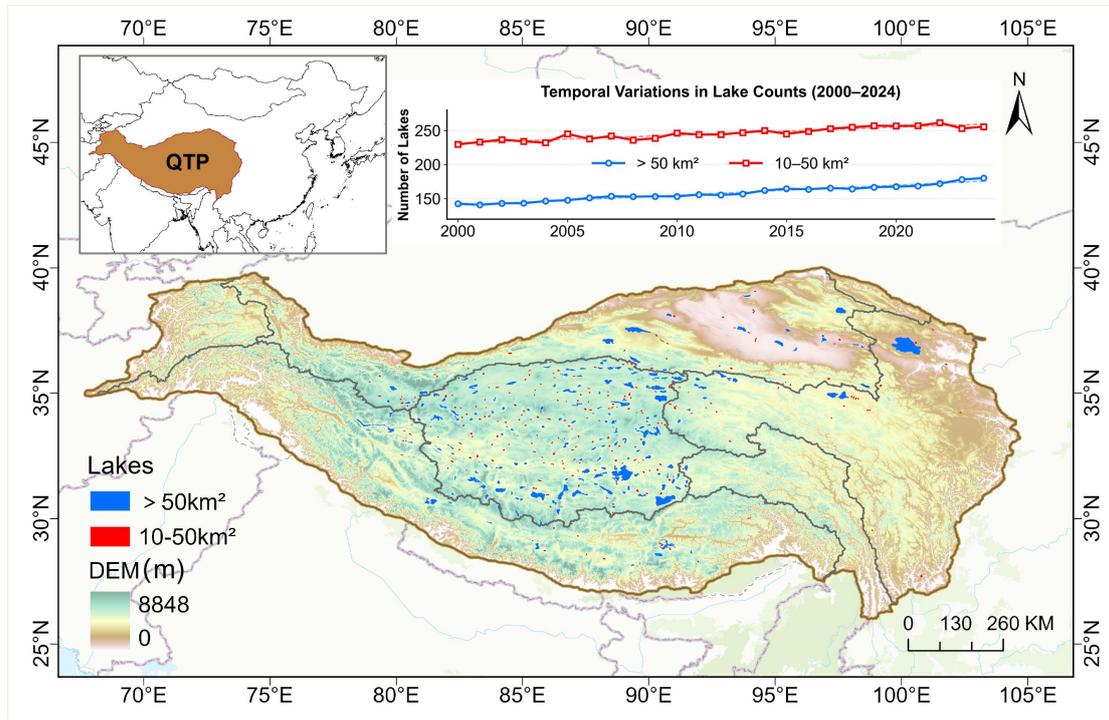

Figure 1: Overview map of the study area. The main map illustrates the spatial distribution of lakes categorized by size, overlaid on the topographic relief. Blue contours indicate large lakes ($>50\ km^2$), while red contours denote medium-sized lakes ($10-50\text{km}^2$). The inset time-series plot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

Comment 7:

I do not see the point of CART when RF is used for classification. Hundreds of decision trees are certainly better than a single one, as far as I see it makes no sense, to apply them together. The description provided for the CART is rather vague, and does explain its significance. Furthermore, the description of SVM and RF (and CART) is very short, do not help to understand the method (while, I agree that

Response:

We sincerely thank the reviewer for this critical and constructive comment.

First, regarding the use of CART, we fully agree that Random Forest (RF), as an ensemble method, is theoretically expected to outperform a single Decision Tree. However, we included CART not to challenge RF's superiority, but to serve as a comparative baseline. By comparing RF against a structurally regularized single tree, we explicitly quantified the performance gains achieved by ensemble learning (an improvement of ~7.5% in accuracy) and demonstrated its necessity for handling the "salt-and-pepper" noise on the Tibetan Plateau.

Second, regarding the model descriptions and hyperparameters, we apologize for the lack of detail in the original manuscript. We have extensively rewritten Section 3.2.2 and added a new Table 2 to explicitly list the configuration for all three classifiers.

We clarified that hyperparameters were determined using grid search with cross validation.

- **RF:** We specified the settings (150 trees, minLeaf=5) that ensured ensemble diversity and generalization.

- **SVM:** We explicitly stated the use of a Radial Basis Function (RBF) kernel. Through empirical tuning, the Cost (C) was set to 10 and Gamma (Y) to 0.5.

- **CART:** We clarified that we applied structural regularization (capping at 128 and setting to 5) to suppress noise, rather using a default unpruned tree.maxNodesminLeafPopulation

We added Table 2 to summarize these hyperparameters, ensuring the experiment is fully reproducible.

Table2: Hyperparameter settings for the machine learning classifiers

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

Comment 8:

I do not see the point to present the equations of Precision, Recall and F1 + RMSE, ubRMSE, Bias, R2, MAPE – as far as see, all papers have these equations, but I never see the reason. If these are shown, why the equations of RF, SVM, or CART are not shown?

Response:

We fully accept the reviewer's suggestion. We acknowledge that metrics such as Precision, Recall, F1-score, RMSE, and R^2 are standard performance indicators in the remote sensing community, and listing their mathematical formulas is indeed redundant.

We have removed Equations (6)–(13) from the revised manuscript (Section 3.2.3 and Section 3.4). Instead, we now describe these evaluation metrics directly in the text and provide appropriate citations to standard references (e.g., Congalton, 1991; Li et al., 2015) for readers interested in the definitions. This modification has significantly streamlined the methodology section.

Comment 9:

Response:

We appreciate this helpful suggestion regarding the scientific writing style. We agree that direct citations (e.g., "Figure 3 shows...") can interrupt the flow of the narrative and that the text should focus on interpreting the scientific findings rather than merely pointing to the illustrations.

We have carefully reviewed the entire manuscript and strictly revised all instances of direct figure citations to ensure the narrative focuses on interpretation. Specific revisions include:

- **Section 3.3 (Original P8 L206):**

  Original: "Figure 3 illustrates the processing workflow and example results for selected lakes."

  Revised: "The complete post-processing workflow and representative outcomes are visualized in the schematic diagram (Fig. 3)."

- **Section 3 (Workflow description):**

  Original: "Figure 2 summarised the monthly lake area data generation workflow..."

  Revised: "The monthly lake area data generation workflow comprises four main stages, seamlessly integrating data preprocessing with machine learning classification (Fig. 2)."

- **Section 4.1.2 (Intra-annual analysis):**

  Original: "Figures (b-i) show the intra-annual maximum and minimum boundary differences..."

  Revised: "Analysis of lakes with significant variability reveals distinct intra-annual maximum and minimum boundary differences, highlighting the intensity of seasonal fluctuations (Fig. 6b-i)."

- **Section 4.2 (Validation results):**

  Original: "Figure 8 presents the water body boundary extraction results for five representative lakes..."

Revised: "Representative extraction results for five typical lakes (e.g., Qinghai Lake, Nam Co) demonstrate robust consistency across spring, summer, autumn, and winter (Fig. 8)."

We have ensured that figure citations are now consistently placed in parentheses at the end of sentences throughout the revised manuscript.

Comment 10:

Fig 3. is a part of Fig 2. I do not think it would be a good idea to replicate the same figure in parts. Refer to Fig 2.

Response:

We thank the reviewer for pointing out this redundancy. We agree that Figure 3 illustrates a processing step that is already encompassed within the overall framework presented in Figure 2. We have removed Figure 3 from the manuscript to avoid repetition. The text in Section 3.3 has been updated to reference Figure 2. All subsequent figures (formerly Figs. 4–10) have been renumbered strictly throughout the manuscript and the text citations.

Comment 11: Results

The results are correctly presented; the figures help the understanding and serve as good background.

Comment 12: Discussion

The section of "Comparison with other products" is a good point, but a wider discussion would be needed. As I noted for the introduction, a global outlook would make the paper more popular, even if the comparison is not that direct.

https://global-surface-water.appspot.com/map

https://www.hydrosheds.org/

https://www.arcgis.com/apps/mapviewer/index.html?webmap=5d65be95ccc341d5878 96a81794021bf

https://essd.copernicus.org/articles/17/2277/2025/

Response:

We are sincerely grateful to the reviewer for this insightful suggestion and for providing the valuable references to global datasets (JRC, HydroLAKES, etc.). We agree that placing our dataset in a broader global context and conducting a direct comparison with established products significantly enhances the paper's impact.

To address this, we have expanded the "Comparison with other products" section to include a comprehensive comparison between our TPLake-MED dataset and global products like the JRC Global Surface Water and HydroLAKES.

We have added a new composite figure to visually compare our TPLake-MED dataset with the JRC Global Surface Water product and the HydroLAKES baseline (Fig. 11).

Similarities: Both datasets demonstrate a high degree of consistency during clear-sky conditions in summer. This alignment validates the baseline accuracy of our water extraction algorithm, confirming that our MODIS-based results are reliable and comparable to standard global products when atmospheric conditions are ideal.

Differences: As shown in the time-series comparison, while global products like JRC provide excellent long-term trends, they suffer from significant data gaps (dropping to zero) in the Tibetan Plateau due to frequent cloud cover and seasonal snow/ice. In contrast, our dataset maintains high temporal continuity and successfully captures seasonal dynamics.
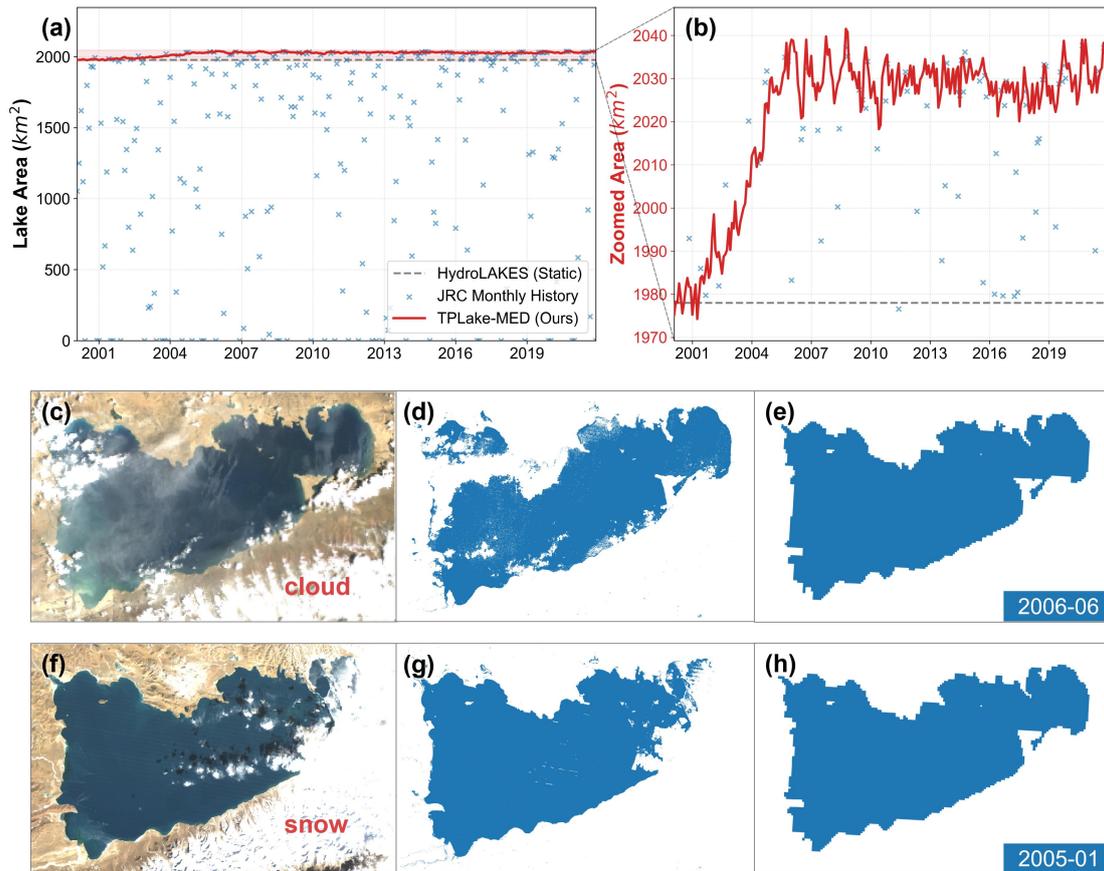
Figure 10: Comparison of temporal consistency and spatial completeness between the TPLake-MED dataset and the JRC Global Surface Water product under complex atmospheric conditions. (a) The long-term time-series comparison (2000–2024) of lake area for Nam Co. The overview highlights the frequent data gaps and near-zero values in the JRC dataset compared to the continuous monitoring capability of TPLake-MED. (b) The zoomed-in panel details the seasonal dynamics captured by TPLake-MED, demonstrating high consistency with valid JRC observations during clear-sky periods and reflecting the significant decadal expansion compared to the static HydroLAKES baseline. (c–e) Visual verification of spatial completeness during a cloud-contaminated month (June 2006). (c) Reference Landsat imagery obscured by clouds; (d) fragmented JRC water extent; (e) complete lake boundary extracted by TPLake-MED. (f–h) Comparison during a snow-covered month (January 2005). (f) Reference imagery; (g) JRC result with snow interference; (h) optimized TPLake-MED boundary. (JRC data © European Commission Joint Research Centre; Landsat imagery courtesy of USGS).

It may be my fault, but I did not find a download link to see the data itself, so in this phase I do not have direct impression on the quality of the dataset.

Response:

We sincerely apologize for this inconvenience. There was a temporary technical issue with our data hosting server at the time of your review, which prevented the download link from functioning properly. We have immediately fixed this issue, and the dataset is now fully accessible. We invite you to access and evaluate the data quality via the following link:

https://doi.org/10.12443/BNU.RSEC.TPLake-MED20251028