https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

Open Access Earth System Science Data Discussions

Landsat and Sentinel-derived glacial lake dataset in the China-Pakistan Economic Corridor from 1990 to 2020

Muchu Lesi[1], Yong Nie[1, *], Dan H. Shugar[2], Jida Wang[3], Qian Deng[1, 4], Huayong Chen[1]

[1]Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu, China.

[2]Water, Sediment, Hazards, and Earth-surface Dynamics (waterSHED) Lab, Department of Geoscience, University of Calgary, Alberta, T2N 1N4, Canada

[3]Department of Geography and Geospatial Sciences, Kansas State University, Manhattan, Kansas 66506, USA

[4]University of Chinese Academy of Sciences, Beijing 100190, China

[*]Corresponding author, nieyong@imde.ac.cn

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

Abstract. The China-Pakistan Economic Corridor (CPEC) is one of flagship projects of the One Belt One Road Initiative, which faces threats from mountain disasters in the high altitude region, such as glacial lake outburst floods (GLOFs). An up-to-date high-quality glacial lake dataset with critical parameters (e.g. lake types), which is fundamental to flood risk assessments and predicting glacier-lake evolutions, is still largely absent for the entire CPEC. This study describes a glacial lake dataset in 2020 for CPEC at 10-30 m resolution, which was produced from both Landsat and Sentinel optical images as well as glacial lake inventories in 1990 and 2000 from Landsat observation, using an advanced object-oriented mapping method associated with rigorous visual inspection workflows. The results show that Landsat derived 2234 glacial lakes in 2020, covering a total area of 86.31±14.98 km^2 with a minimum mapping unit of 5 pixels (4500 m^2), whereas Sentinel derived 7560 glacial lakes in 2020 with a total area of 103.70±8.45 km^2 with a minimum mapping unit of 5 pixels (500 m^2). The discrepancy implies that there is a significant quantity of small glacier lakes not recognized in existing glacial lake inventories and a more thorough inclusion of them require future efforts using higher resolution data. The total number and area of glacial lakes from consistent 30 m resolution Landsat images remain relatively stable despite a slight increase from 1990 to 2020. A range of critical attributes have been generated in the dataset, including lake types of two classification systems and mapping uncertainty estimated by an improved equation. This comprehensive glacial lake dataset has potentials to be widely applied in studies on glacial lake-related hazards and glacier-lake interactions, and is freely available at https://doi.org/10.12380/Glaci.msdc.000001 (Lesi et al., 2022).

1 Introduction

Glaciers in High-mountain Asia (HMA) play a crucial role in regulating climate, supporting ecosystems, modulating the release of freshwater into rivers, and sustaining municipal water supplies (Wang et al., 2019; Viviroli et al., 2020), agricultural irrigation, and hydropower generation (Pritchard, 2019; Nie et al., 2021). Most HMA glaciers are losing mass in the context of climate change (Brun et al., 2017; Shean et al., 2020; Bhattacharya et al., 2021; Maurer et al., 2019), therefore, unsustainable glacier melt is reducing the hydrological role of glaciers and impacting downstream ecosystem services, agriculture, hydropower and other socioeconomic values (Nie et al., 2021). The present and future glacier changes also alter the frequency and intensity of glacier-related hazards, such as glacier lake outburst floods (GLOFs) (Nie et al., 2018; Zheng et al., 2021; Rounce et al., 2020), and rock and ice avalanches (Shugar et al., 2021). The increasing frequency of GLOFs has been observed in the Karakoram and Himalaya (Nie et al., 2021), and the increasing risk of GLOFs is threatening existing and planned infrastructures in the mountain ranges, such as hydropower plants, railways, and highways.

A large number of major infrastructure construction projects for the One Belt One Road Initiative (BRI) play a fundamental role in strengthening the interconnection of infrastructure between countries and promoting international trade and investment (Battamo et al., 2021; Li et al., 2021). Taking the Karakoram Highway for example, it is a unique land route to link

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

China and Pakistan. The China-Pakistan Economic Corridor (CPEC) is one of the BRI flagship projects, originating from Kashgar of the Xinjiang Uygur Autonomous region, China and extending to Gwadar Port, Pakistan (Ullah et al., 2019; Yao et al., 2020). The northern section of the CPEC passes through Pamir, Karakoram, Hindu Kush and Himalaya mountains where glacier-related hazards such as GLOFs are frequent and severe (Hewitt, 2014; Bhambri et al., 2019), threatening the existing, under-construction and planned infrastructure projects. Understanding the risk posed by GLOFs is a critical step to disaster prevention for infrastructures across the CPEC (Figure 1).

Glacial lake inventories with a range of attributes benefit risk assessment and disaster reduction related to GLOFs, and contribute to predicting glacier-lake evolution under climate change (Nie et al., 2017; Brun et al., 2019; Liu et al., 2020; Maurer et al., 2019). Remote sensing is the most viable way to map glacial lakes and detect their spatio-temporal changes in the high-elevation zones where in situ accessibility is extremely low (Huggel et al., 2002; Quincey et al., 2007). Studies in glacial lake inventories using satellite observations have been heavily conducted at regional scales recently, such as in the Tibetan Plateau (Zhang et al., 2015), the Himalaya (Gardelle et al., 2011; Nie et al., 2017), the HMA (Chen et al., 2021; Wang et al., 2020), the Tien Shan (Wang et al., 2013) and the northern Pakistan (Ashraf et al., 2017). However, the latest glacial lake mapping in 2020 is still absent along the CPEC. Among existing studies, Landsat archival images are the most widely used due to their multi-decadal record of earth surface observations, reasonably high spatial resolution (30 m), and publicly available distribution (Roy et al., 2014). Freely available Sentinel-2 satellite images show a better potential than Landsat in glacial lake mapping and inventories due to their higher spatial resolution (10 m) and a global coverage, but have only been available since late 2015 (Williamson et al., 2018; Paul et al., 2020). Glacial lake inventories using Sentinel images are relatively scarce at regional scales, and studies of the latest glacial lake mapping as well as comparisons of glacial lake datasets derived from Sentinel and Landsat observations are still lacking.

Discrepancies between various glacial lake inventories (Zhang et al., 2015; Shugar et al., 2020; Chen et al., 2021; Wang et al., 2020) result from differences in mapping methods, minimum mapping units, definition of glacial lakes, time periods, data sources and other factors. For example, manual vectorization method was widely adopted at the earlier stage for its high accuracy. However, it is time-consuming associated with high labor intensity and is only practical at regional scales (Zhang et al., 2015; Wang et al., 2020). Automated and semi-automated lake mapping methods, such as band ratio and object-oriented classification (Gardelle et al., 2011; Zhang et al., 2018; Nie et al., 2017), have been developed to improve the efficiency of glacial lake inventories, although artificial modification is unavoidable to assure the quality of lake data impacted by cloud cover in optical images, mountain shadows, seasonal snow cover and frozen lake surfaces (Sheng et al., 2016; Wang et al., 2017; Wang et al., 2018). Type classification of glacial lakes provides a crucial attribute for glacier-lake interactions and risk assessment (Emmer and Cuřín, 2021). Glacier lakes in currently available datasets have been traditionally categorized by their spatial relationship with upstream glaciers (Gardelle et al., 2011; Chen et al., 2021; Wang et al., 2020), and classification attributes considering the formation mechanism and the properties of dams are rare or incomplete in the CPEC (Li et al., 2021; Yao et al., 2018). Therefore, an up-to-date

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

Open Access Earth System Science Data Discussions

glacial lake dataset with critical, quality-assured parameters (e.g. lake types) is necessary.

This study aims to (1) employ both Landsat 8 and Sentinel-2 images to create an up-to-date glacial lake dataset in the CPEC to accurately document its detailed lake distribution in 2020; (2) reveal glacial lake changes and the spatial heterogeneity across mountains and basins in the CPEC using consistent 30-m Landsat images at three time periods (1990, 2000 and 2020); and (3) share the glacial lake inventories with a range of critical attributes to benefit hazardous risk assessment of GLOFs and glacio-hydrological modeling in the HMA.

2 Study area

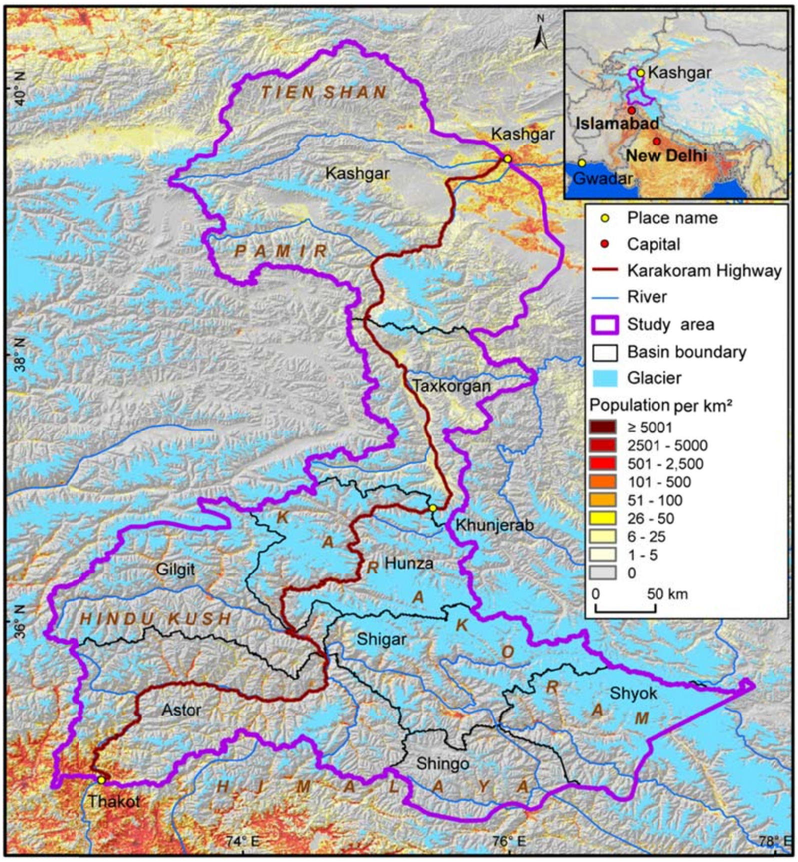


Figure 1. Location of the study area and distribution of glaciers, mountains, basins and population.

The study area (**Figure 1**) covers all the drainage basins along Karakoram Highway starting from Kashgar and ending at Thakot, with a total area of ~125,000 km^2. The upper Indus basins beyond the Pakistani-administrated border are excluded in this study due to little impact of GLOFs there on CPEC infrastructures. The entire study area is divided into eight sub-basins, covering most of the Karakoram with the highest altitude up to 8611 m, western Himalaya and Tien Shan, eastern Hindu Kush and Pamir mountains. The 9710 glaciers in the study area cover a total area of 17,447 km^2 and nearly 60% of glaciers are distributed in the

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

Open Access Earth System Science Data Discussions

Karakoram (5818 glaciers with a total area of 14,067.52 km^2) (RGI Consortium, 2017). Most glaciers in the western Himalaya and eastern Hindu Kush are losing mass in the context of climate change (Kääb et al., 2012; Yao et al., 2012; Shean et al., 2020; Brun et al., 2017; Hugonnet et al., 2021), whereas the glaciers in the eastern Karakoram and Pamir have shown unusually little changes, including unchanged, retreated, advanced and surged glaciers (Nie et al., 2021; Brun et al., 2017; Shean et al., 2020; Kääb et al., 2012; Hewitt, 2005; Bolch et al., 2017). The spatially heterogeneous distribution and changes of glaciers are primarily explained as a result of differences in the dominant precipitation-bearing atmospheric circulation patterns that include the winter westerlies the Indian summer monsoon, their changing trends and their interactions with local extreme topography (Azam et al., 2021; Nie et al., 2021; Yao et al., 2012).

3 Data sources

Both Landsat and Sentinel images have been employed to map glacial lakes between 1990 and 2020 in the CPEC (Figure 2). A total number of 98 Landsat Thematic Mapper (TM), Thematic Mapper Plus (ETM+) and Landsat 8 Operational Land Imager (OLI) images with a consistent spatial resolution of 30 m were downloaded from the United States Geological Survey Global Visualization Viewer (GloVis, https://glovis.usgs.gov/app/) to be used to create glacial lake inventories in 1990, 2000 and 2020. High-quality Landsat images around 2010 are insufficient to cover the entire study area, so we had to give up glacial lake mapping in 2010 as a result of Landsat 7's scan-line corrector errors and significant cloud covers. In addition, 40 Sentinel-2 images were downloaded from Copernicus Open Access Hub (https://scihub.copernicus.eu/) to produce the 10-m resolution glacial lake inventory in 2020.

Cloud and snow covers heavily affect the usability of optical satellite images (Wulder et al., 2019) and their availability in the entire study area, so we took advantage of the images acquired before and after each of the baseline years 1990, 2000 and 2020 to construct the glacial lake inventories. To minimize the impact of intra-annual changes of glacial lakes, most of used images (85% for Sentinel and 82% for Landsat) were acquired from August to October in the given baseline year with cloud coverage of <20% for each image. For some specific scenes where cloud cover exceeded the threshold of 20%, we selected more than one image to remedy the effect of cloud contamination (Nie et al., 2010; Nie et al., 2017; Jiang et al., 2018).

Other datasets used include the Randolph Glacier Inventory version 6.0 (Pfeffer et al., 2014; RGI Consortium, 2017) and the Glacier Area Mapping for Discharge from the Asian Mountains (GAMDAM) glacier inventory (Sakai, 2019). These two glacier datasets were used to determine glacial lake attributes. The Shuttle Radar Topography Mission Digital Elevation Model (SRTM DEM) at a 1-arc second (30 m) resolution (Jarvis et al., 2008) was employed to extract the altitudinal characteristics of the glacial lakes. The absolute vertical accuracy of the SRTM DEM is 16 m (90%) (Farr et al., 2007; Rabus et al., 2003). We also applied other published glacial lake datasets for comparative analysis. They include the glacial lake inventories of HMA in 1990 and 2018 downloaded from http://doi.org/10.12072/casnw.064.2019.db (Wang et al., 2020), the Third Pole region in 1990, 2000 and 2010 publicly shared at http://en.tpedatabase.cn/ (Zhang et al., 2015), the Tibet Plateau from 2008 to 2017 accessed at https://doi.org/10.5281/zenodo.3700282 (Chen et al.,

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

2021), and the entire world in 1990, 2000 and 2015 provided at https://nsidc.org/data /HMA_ GLI/versions/1 (Shugar et al., 2020). In addition, field survey data collected between 2017 and 2018 were also used to assist in lake mapping and glacial lake type classification.

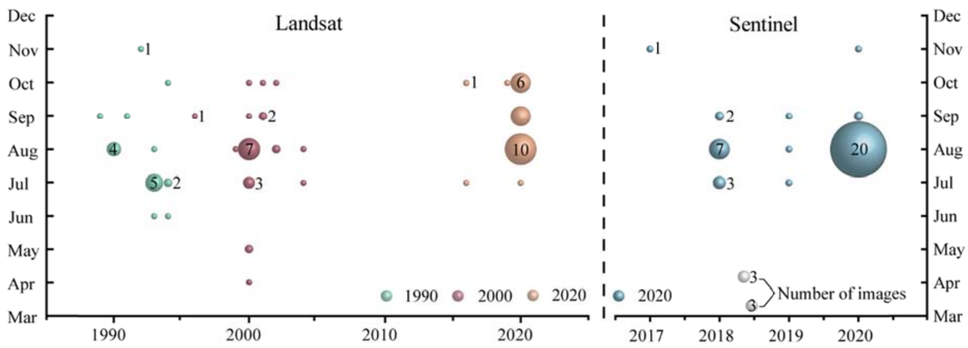


Figure 2. Acquisition years and months of Landsat and Sentinel images selected for glacial lake inventories. The bubble size indicates the available image number.

4 Glacial lake inventory methods

4.1 Definition of glacial lakes

We consider a glacial lake as one that formed as a result of modern or ancient glaciation. Contemporary glacial lakes are easily recognized using a combination of glacier inventories and remote sensing images. Ancient glacial lakes can be identified from periglacial geomorphological characteristics, including moraine remnants and U-shaped valleys that are discernible from satellite observations (Post and Mayo, 1971; Nie et al., 2018; Martín et al., 2021; Westoby et al., 2014). Landslide-dammed lakes (Chen et al., 2017) in the periglacial environment were excluded in our inventories because of their irrelevance to glaciation. We abandoned the definition that considers all lakes surrounding a specific buffering distance of other glaciers also as glacier lakes, although this definition has been widely used in previous studies assuming glacial meltwater as the main water supply (Zhang et al., 2015; Wang et al., 2020). This is because the contribution of glacial meltwater to the lake supply is arduous to be quantified without an accurate modeling of the cryosphere-hydrological processes (Lutz et al., 2014). All glacial lakes in the study area were mapped according to our definition without any distance limit between lakes and glaciers. We were able to implement this definition by carefully leveraging the spectral properties of glacial lakes and the periglacial geomorphological features that are often evident in remote sensing images (see more in sections 4.3 and 4.4).

4.2 Interactive lake mapping

A human-interactive and semi-automated lake mapping method (Wang et al., 2014; Nie et al., 2017; Nie et al., 2020) was adopted to accurately extract glacial lake extents using Landsat and Sentinel-2 images, based on the Normalized Difference Water Index (NDWI) (Mcfeeters,

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

1996). The NDWI uses the green and near infrared bands and is calculated by the following equation:

$$NDWI = \frac{Band_{Green} - Band_{NIR}}{Band_{Green} + Band_{NIR}} \quad (1)$$

where the green band and near infrared band were provided by both Landsat and Sentinel multispectral images.

Specifically, the method automatically generated the histogram of NDWI in each user-defined region of interest. The NDWI threshold that separates lake surface from land was interactively determined by screening the NDWI histogram against the lake region in the imagery (Nie et al., 2020; Wang et al., 2012). This way, the determined NDWI threshold can be well-tuned to adapt various spectral conditions of the studied glacier lakes. The raster lake extents segmented by the thresholds were then converted to vector polygons. We first completed the glacial lake inventory in 2020 using this interactive mapping method, and the 2020 inventory was then used as a reference to facilitate the lake mapping for other periods.

The minimum mapping unit (MMU) was set to 5 pixels for both Landsat (0.0045 km^2) and Sentinel-2 images (0.0005 km^2) in this study. MMU determines the total number and area of glacial lakes in the dataset, and varies in the previous studies, such as 3 pixels (Zhang et al., 2015), 9 pixels (Chen et al., 2021), or 55 pixels (Shugar et al., 2020) for Landsat images for various objectives and spatial scales. While a smaller threshold leads to a large quantity of lakes mapped, it also generates larger mapping noises or uncertainties. Considering this signal-noise balance and our focus on identifying prominent glacier lake dynamics in the study area, we opted to use 5 pixels as the minimum mapping unit for both Landsat and Sentinel-2 images.

Several procedures were taken to assure the quality assurance and quality control for lake mapping, including 1) visual inspection and modification for each lake based on Landsat, Sentinel-2 and Google Earth high-resolution images overlaying preliminarily lake boundary extraction at the given time period; 2) time series check for Landsat-derived glacial lake datasets from 1990 and 2020, and cross-check between Landsat and Sentinel-2-derived lake dataset in 2020 to reduce errors of omission and commission; 3) topological validation of glacial lake mapping, such as repeated removal, elimination of small sliver polygons; and 4) logical check for lake types between two classification systems of glacial lakes. False lake extents resulting from cloud or snow cover, lake ice, and topographic shadows (Nie et al., 2020; Nie et al., 2017) and were modified using alternative images acquired in adjacent years. Those procedures were time-consuming, but helped to minimize the effect of cloud and snow covers, lake mapping errors, and to maximize the quality of the produced lake product and the derived glacial lake changes.

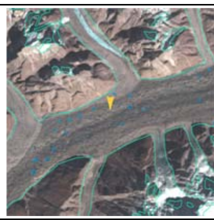
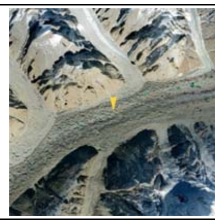
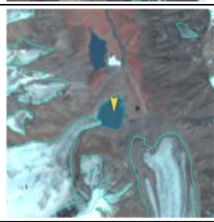
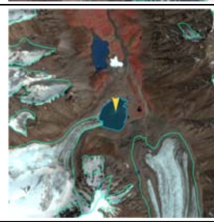
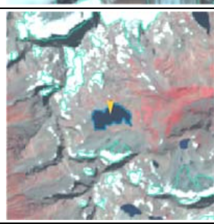
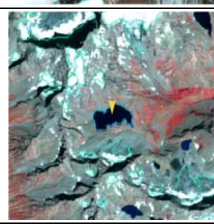
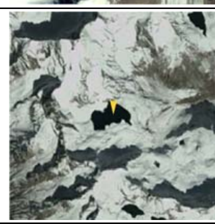
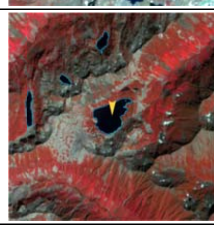
4.3 Classification of glacial lakes

Two glacial lake classification systems (GLCS) have been established based on relationship of interaction between glacial lakes and glaciers as well as lake formation mechanism and dam material properties. In the first GLCS (GLCS1), glacial lakes were classified into four types based on their spatial relationship to upstream glaciers: supraglacial, proglacial, unconnected-glacier-fed lakes, and non-glacier-fed lakes according to Gardelle et al. (2011). Alternatively, combining the formation mechanism of glacial lakes and the properties of

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

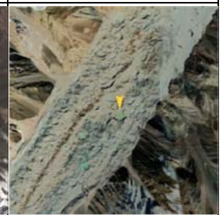
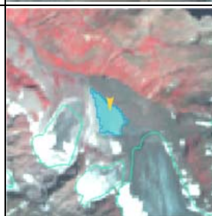
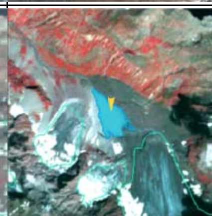
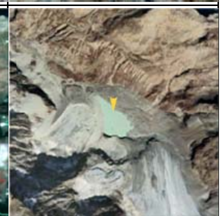
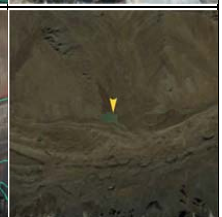
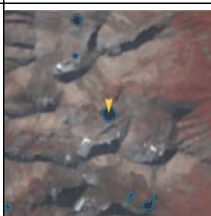
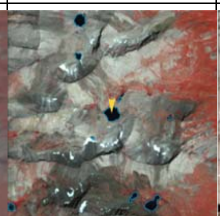
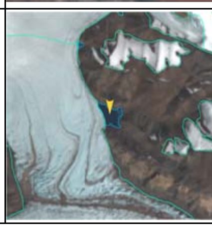
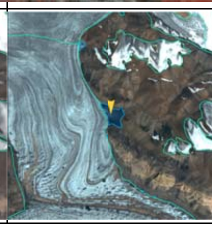
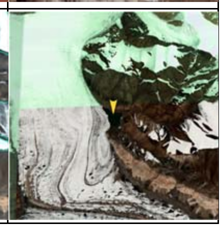
natural dam features, glacial lakes were classified into five categories (herein named GLCS2) modified from Yao's classification system (2018): supraglacial, end-moraine dammed, lateral-moraine dammed, glacial erosion lakes and ice-blocked lakes. Characterization and examples for each type are provided in Table 1 and Table 2. Individual glacial lakes were categorized to the specific types for each GLCS according to available glacier inventory data, geomorphological and spectral characteristics interpreted from Landsat, Sentinel and Google Earth images. The synergy of these two GLCSs is beneficial to predicting glacier-lake evolutions and providing fundamental data for glacial lake disaster risk assessment.

Table 1. Classification system of glacial lake types according to the relationship between glacial lakes and glaciers (© Google Earth 2019).

Lake types	Characteristics	Landsat	Sentinel	Google earth
Supraglacial	Lakes formed on the surface of glaciers, generally dammed by ice and thin debris. Case location: 35°43'49.74" N 76°13'53.88" E			
Proglacial	Lakes dammed by moraine, ice or bedrock, supplied by glacial meltwater and connected with glaciers. Case location: 39°09'32.40" N 73°43'12.00" E			
Unconnected-glacier-fed	Lakes currently supplied by upstream glacial meltwater but disconnected with glaciers. Case location: 35°47'60.00" N 72°55'15.60" E			
Non-glacier-fed	Lakes formed by glaciology, dammed by moraine or bed rock, and currently not supplied by glacial meltwater. Case location: 34°50'39.99" N 74°48'29.31" E			

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

Table 2. Classification system of glacial lake types according to the formation mechanism of glacial lakes and dam material properties (© Google Earth 2019).

Lake types	Characteristics	Landsat	Sentinel	Google earth
Supraglacial	Lakes formed on the surface of glaciers, generally dammed by ice and thin debris. Case location: 36°46'7.39" N 74°20'7.59" E			
End-moraine-dammed	Lakes formed behind moraines as a result of glacier retreat and downwasting. Case location: 35°42'50.40" N 73°09'57.60" E			
Lateral moraine-dammed	Lakes formed behind lateral glacial moraine ridges and dammed by debris, different from ice-blocked glacial lake. Case location: 38°28'45.62" N 75°20'52.30" E			
Glacial erosion	Lakes formed in depressions created by glacial over-deepening. Bedrock dam dominates, partially superimposed by top moraine in rugged terrain. Dams are unclear in the satellite images. Case location: 35°55'55.56" N 73°38'20.13" E			
Ice-blocked	Lakes formed behind glaciers, dammed by glacier ices (partially covered by debris on the top). Case location: 35°28'31.32" N 77°30'46.81" E			

4.4 Attributes of glacial lake data

A total of 17 attribute fields were input into our glacial lake datasets (**Table 3**). They include lake location (longitude and latitude), lake elevation (centroid elevation), orbital number of the image source, image acquisition date, lake area, lake perimeter, lake types of the two GLCSs, mapping uncertainty, and the country, sub-basin, and mountain range associated with the lake. Amongst the attributes, lake location was calculated based on the centroid of each glacial lake polygon associated with the DEM, N represents northing and E represents easting.

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

Orbital number of the image source was filled with the corresponding satellite image, with the codes expressed as "PxxxRxxx" or "Txxxxx", where P and R indicate the path and row for Landsat image and T represents the tile of Sentinel image associated with 5 digits code of military grid reference system. Area and perimeter were automatically calculated based on glacial lake extents. Lake types were attributed using the characterization and interpretation marks described in Section 4.3. Mapping uncertainty was estimated using our modified equation which will be introduced in section 4.5 and supplementary tutorial. Located country, sub-basin and mountain range of each glacial lake was identified by overlapping the geographic boundaries of countries, basins and mountain ranges.

Table 3. Classification system of glacial lake types according to the formation mechanism of glacial lakes and dam material properties.

Field Name	Type	Description	Note
FID or OBJECTID	Object ID	Unique code of glacial lake	Number
Shape	Geometry	Feature type of glacial lake	Polygon
Latitude	String	Latitude of the centroid of glacial lake polygon	Degree minute second
Longitude	String	Longitude of the centroid of glacial lake polygon	Degree minute second
Elevation	Double	Altitude of the centroid of glacial lake polygon	Unit: meter above sea level
IMGSOURCE	String	Path and row numbers for Landsat image based on World Reference System 2 or Tile number for Sentinel image based on military grid reference system	PxxxRxxx or Txxxxx
ACQDATE	String	Acquisition date of source image	YYYYMMDD
GLCS1	String	The first classification system of glacial lakes based on relationship of interaction between glacial lakes and glaciers	Supraglacial, Proglacial, Unconnected-glacier-fed, None-glacier-fed
GLCS2	String	The second classification system of glacial lakes based on lake formation mechanism and dam material properties	Supraglacial, End-moraine-dammed, Lateral-moraine-dammed, Glacial erosion and

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

Field Name	Type	Description	Note
			Ice-blocked
Basin	String	Basin name where glacial lake locates in	
Mountains	String	Mountain name where glacial lake locates in	
Country	String	Country name where glacial lake locates in	
Perimeter	Double	Perimeter of glacial lake boundary	Unit: meter
Area	Double	Area of glacial lake coverage	Unit: square meter
Uncertainty	Double	Uncertainty of glacial lake mapping estimated based on modified Hanshaw's equation (2014).	Unit: square meter
Operator	String	Operator of glacial lake dataset	Muchu, Lesi
Examiner	String	Examiner of glacial lake dataset	Yong, Nie

4.5 Improved uncertainty estimating method

We modified Hanshaw's (2014) equation that had been used to calculate lake-area mapping uncertainty. Lake perimeter and displacement error are widely used to estimate the uncertainty of glacier and lake mapping from satellite observation. Hanshaw and Bookhagen (2014) proposed an equation to calculate the error of area measurement by the number of edge pixels of the lake boundary multiplied by half of a single pixel area. The number of edge pixels is simply calculated by the perimeter divided by the grid size. The equation is expressed as below:

$$Error(1\sigma) = \frac{P}{G} \times 0.6872 \times \frac{G^2}{2} \quad (2)$$

$$D = \frac{Error(1\sigma)}{A} \times 100\% \quad (3)$$

Where G is the cell size of the remote sensing imagery (10 m for Sentinel-2 image and 30 m for Landsat image). P is the perimeter of individual glacial lake (m), and the revised coefficient of 0.6872 was chosen assuming that area measurement errors follow a Gaussian distribution. Relative error (D) was calculated by equation 3, in which A is the area of an individual glacial lake.

In the original equation 2, the number of edge pixels varies by the shape of lake and is indicated by $\frac{P}{G}$. However, the pixels in the corner are double counted (Figure 3). The total number of repeatedly calculated edge pixels equals the number of inner nodes. Therefore, we adjusted the calculation of the actual number of edge pixels as the maximum of edge pixels ($\frac{P}{G}$) subtracting the number of inner nodes. Accordingly, the equation of uncertainty estimation

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

for lake mapping is modified as below:

$$Error(1\sigma) = \left(\frac{P}{G} - N_{inner}\right) \times 0.6872 \times \frac{G^2}{2} \quad (4)$$

Where N_{inner} is the number of inner nodes (inflection points) of each lake. The modified equation is also suitable for lakes with islands (as illustrated in Figure 3b).

For polygons without islands (Figure S3a), use the following equation:

$$N_{inner} = \left(\frac{N_{Total}-4-1}{2}\right) \quad (5)$$

N_{Total} is the total number of nodes, including both the outer and inner. N_{Total} were calculated by the “Field Calculator” in ArcGIS, in some cases, it is necessary to remove the redundant nodes before calculating the total number of nodes (See the Supplement for more details). An inner node is a polygon vertex where the interior angle surrounding it is greater than 180 degrees. An outer node is the opposite of the inner node, where the interior angle is less than 180 degrees. We found that the outer nodes are usually four more than the inner nodes in our glacial lake dataset. The total nodes in ArcGIS contain one overlapping node to close the polygon, meaning the endpoint is also the startpoint. This extra count was deleted in the calculation (equation 5).

For polygons with island (Figure S3b) use the following equation:

$$N_{inner} = \left(\frac{N_{Total}-(N_{island}+1)\times 5}{2}\right) \quad (6)$$

N_{island} is the number of islands within each polygon. A calculation method of N_{island} is given in the Supplement.

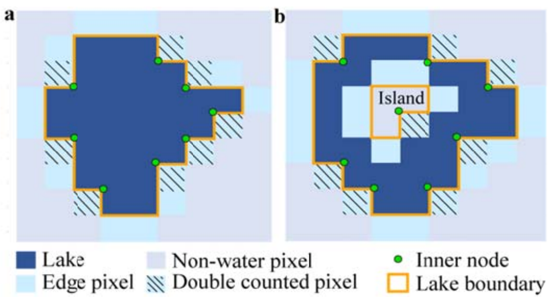


Figure 3. Sketch of estimating the actual edge pixels for uncertainty calculation of individual glacial lake (with and without islands).

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

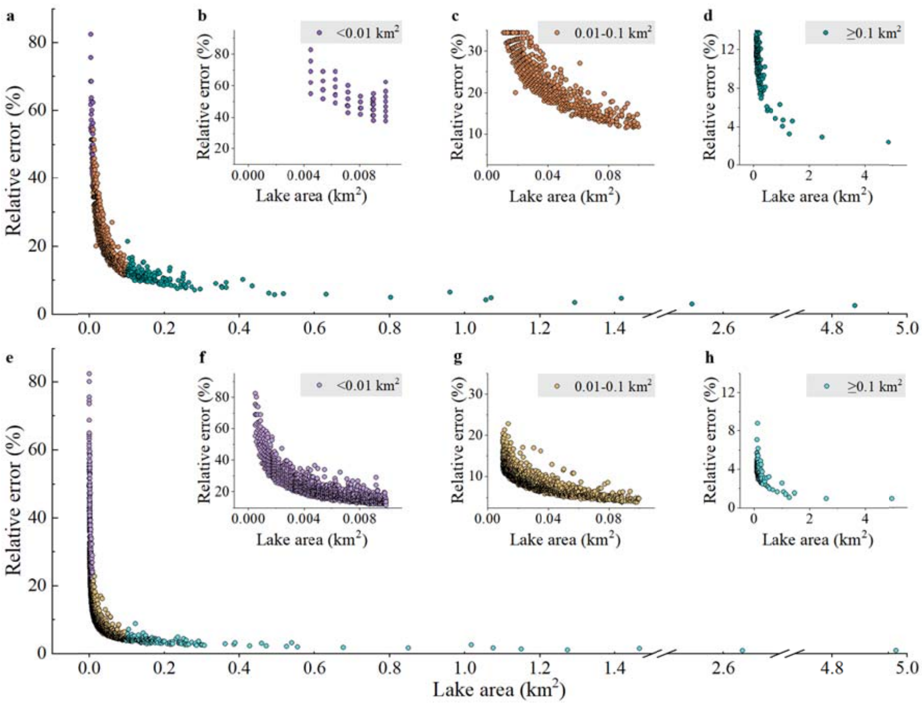


Figure 4. Relationships between individual lake size and its estimated relative error for glacial lakes of all or specific size ranges in study area. Error estimation is based on the modified equation and lake data extracted from Landsat (a-d) and Sentinel images (e-h).

The uncertainty estimated from our improved equation shows that the relative error of individual glacial lake decreases when lake size increases or cell size of remote sensing images reduces (Lyons et al., 2013) (Figure 4). Total area error of glacial lakes in study area is approximate ±14.98 km^2 and ±8.45 km^2 in 2020 for Landsat and Sentinel images, respectively, and the average relative error is ±17.36% and ±8.15%. Generally, small lakes have greater relative errors. For example, the mean relative error is 35.38% for Landsat derived glacial lakes between 0.0045 and 0.1 km^2 and 10.63% for glacial lakes greater than 0.1 km^2. The mean area error of Sentinel-derived glacial lakes is almost one sixth of that extracted from Landsat images for glacial lakes of all or specific size group.

5 Results

5.1 Glacier lake distribution and changes observed from Landsat

We mapped 2,234 glacial lakes for 2020 across the studied CPEC from Landsat-8 images, with a total area of 86.31±14.98 km^2 (Figure 5a and b). The majority of these glacial lakes (1,870 or 83.71%) are smaller than 0.05 km^2 and contribute 36.5% of the total area. 45 (2.01%) of the lakes are larger than 0.2 km^2 and contribute 28.8% of the total area (Figure 6). With the increase of lake size, the abundance (count) of glacial lakes consistently decreases but the total lake area first reduces and then increases. Unconnected-glacier-fed lakes are

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

Open Access Earth System Science Data Discussions

dominant in the first classification system, followed by non-glacier-fed lakes (Figure 7) whereas glacial erosion lakes dominate at both number (1478) and area (57.02 km^2) in the second classification system (Figure 8), followed by end moraine-dammed lakes and supraglacial lakes. Among the classified lakes, 137 are proglacial lakes and cover an area of 5.56 km^2, implying a higher mean size of proglacier lakes than supraglacial lakes.

Glacial lakes are spatially heterogeneous among various mountain ranges and basins in the study area. Himalaya sub-region has the maximum glacier lake count and area across the entire study area, followed by Hindu Kush. Supraglacial lakes are mainly distributed in the Karakoram but they cover less area than those in the Pamir. Tien Shan has fewer glacial lakes. Astor, Gilgit and Shingo basins have the largest percentages of glacier lakes in both number and area (>17%) (Figure 9a), and each of the other basins contributes less than 10% except Kashgar basin in area due to several large ancient glacial lakes. Glacial lakes of less than 0.05 km^2 dominate in number within each basin and the total number decreases as lake size increases. Small lakes consistently account for the maximum percentage in area except Kashgar basin as a result of the disproportionally large lakes.

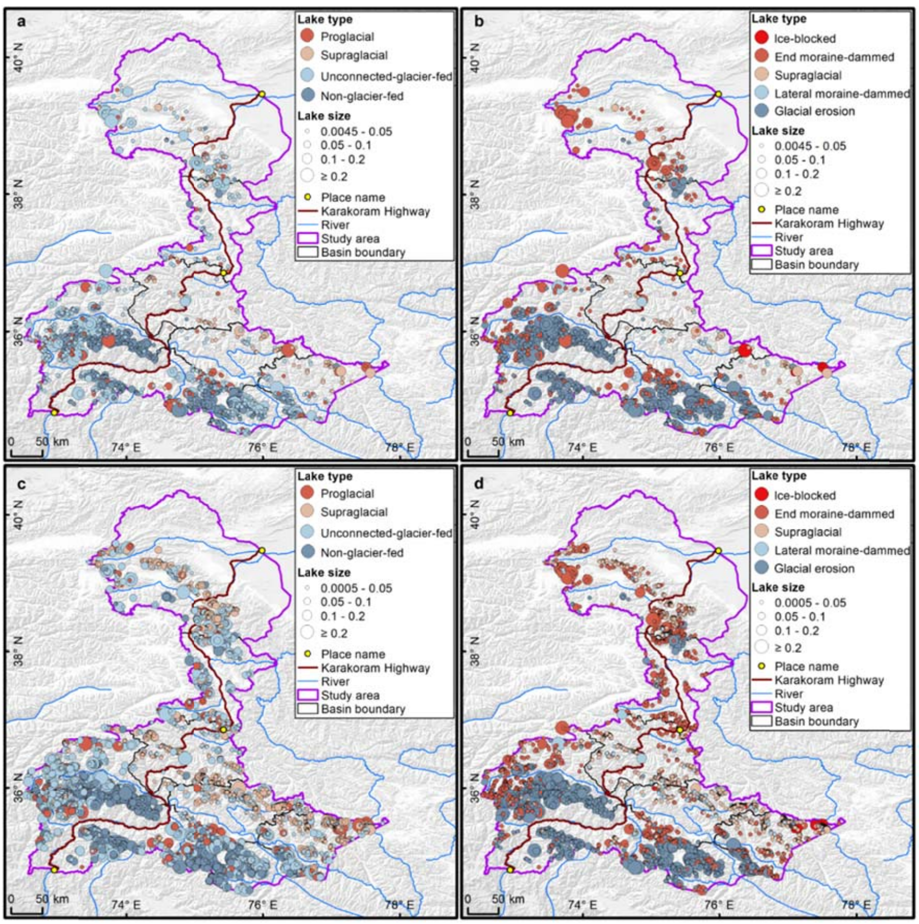


Figure 5. Distribution of glacial lakes in 2020 extracted from Landsat (a, b) and Sentinel (c, d) images.

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

Panels a and b are classified by GLCS1, and GLCS2 for sub-graph c and d.

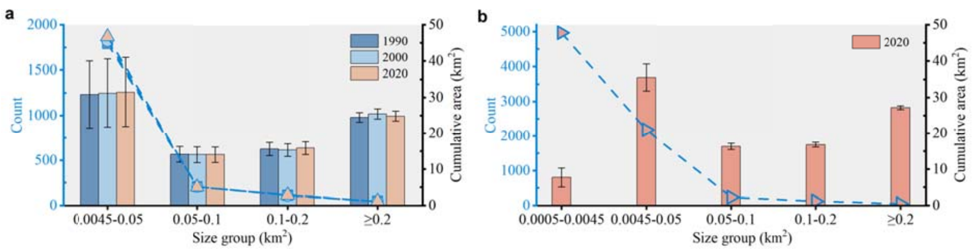


Figure 6. Statistics of different sizes of glacial lakes in the study area from 1990 to 2020. Panels a and b were derived from Landsat and Sentinel images, respectively.

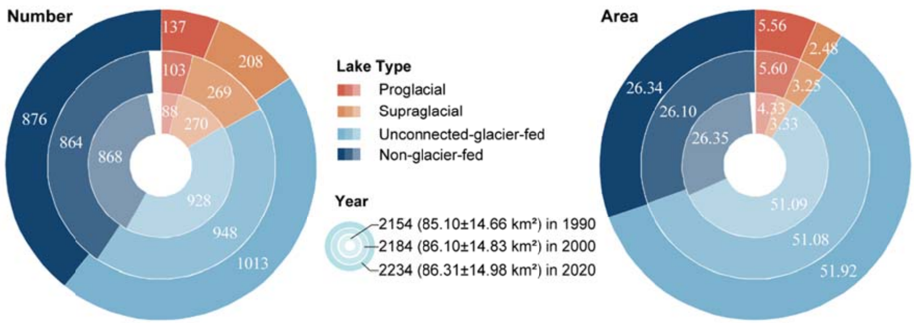


Figure 7. Number and area of different types of glacial lakes classified based on the condition of glacier supply in the study area. The outermost ring represents glacial lake data in 2020, middle ring for 2000 and innermost ring for 1990. Lake number and area in 2020 were selected as reference, meaning a concept of "100 %" for a complete ring. Labeled values are scaled in degrees rather the radius of rings.

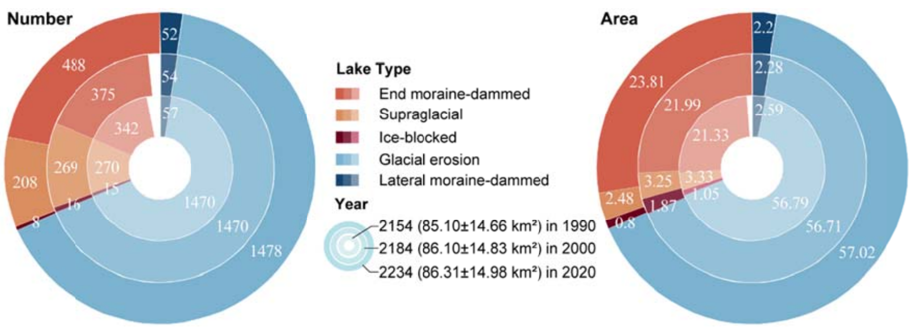


Figure 8. Number and area of different types of glacial lakes classified based on glaciation and nature of dam in the study area. The outermost ring represents glacial lake data in 2020, middle ring for 2000 and innermost ring for 1990. Lake number and area in 2020 were selected as reference, meaning a concept of "100 %" for a complete ring. Labeled values are scaled in degrees rather the radius of rings.

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

Open Access Earth System Science Data Discussions

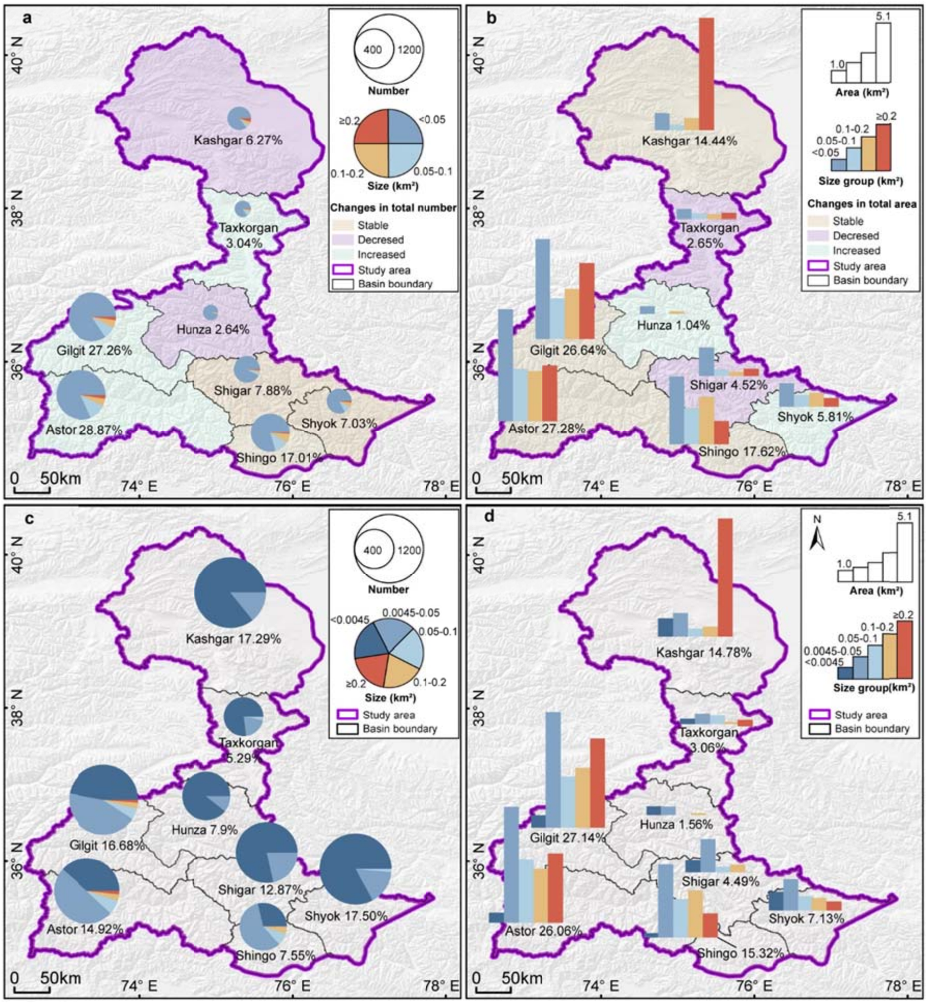


Figure 9. Distributions and changes in count and area of glacial lakes. Percent of glacial lakes in number or area is labeled in each basin. Pie charts present the number of glacial lakes at various size groups between basins (a and c) and bar charts represent total area of glacial lakes at different size groups in each basin (b and d). The background colors represent changes in total number and area between 1990 and 2020 based on Landsat derived dataset (a and b) and distribution of Sentinel derived glacial lakes in 2020 among basins are shown in sub-graphs c and d.

The total number and area of glacial lakes in the study remain relatively stable with a slight increase between 1990 and 2020, and the changes in count and area among various types of glacial lakes vary substantially (Figure 7 and Figure 8). From 1990 to 2020, the total number of glacial lakes increased by 80 or 3.70%, while the area grew by a less extent (1.21 km^2 or 1.42%). Small lakes (<0.05 km^2) continuously increased in number and area, and contributed most in the total lake expansion (Figure 6). Lakes in the size group of 0.05-0.1 km^2 remained stable. The total area of lakes greater than 0.1 km^2 consistently increased.

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

Open Access Earth System Science Data Discussions

In the GLCS1, unconnected-glacier-fed lakes have the largest increase in number, followed by proglacial and non-glacier-fed lakes, whereas supraglacial lakes decreased by 62 in count. Proglacial lakes expanded by 1.24 km^2 (equaling an increase of 26% in proglacial lakes), contributed one third of the total area increase. Supraglacial lakes decreased by 0.85 km^2 in area whereas the areas of unconnected-glacier-fed and non-glacier-fed lakes remained stable as a result of disconnections from glaciers (Figure 7).

In the GLCS2, end moraine-dammed lakes increased by 2.48 km^2 and contributed most of the glacier lake area expansion, whereas supraglacial, ice-blocked and later moraine-dammed lakes decreased slightly in both number and area. Glacial erosion lakes accounted for the maximum percentage (about 66% for both count and area) in each time period and remained stable (Figure 8).

Spatially, glacial lake changes in number and area vary among different mountain ranges and basins between 1990 and 2020 in the study area. Glacial lakes across the west Himalaya and Hindu Kush increased both in number and area between 1990 and 2020 whereas the total number of glacial lakes decreased in the Karakoram, Pamir and Tien Shan of study area (**Table 4**). The total area of glacial lakes continued to increase in the Hindu Kush, but decreased between 1990 and 2000 and increased between 2000 and 2020 in the Himalaya. The total number of glacial lakes continuously decreased in the Pamir and Tien Shan in the past three decades but increased at the first stage and decreased after in the Karakoram. The total area of glacial lakes persistently grew in the Pamir whereas fluctuated in the Tien Shan and Karakoram.

The total numbers of glacial lakes in Shingo, Shigar and Shyok basins were stable (Figure 9a and b); however, the areal changes were less so, including being stable for Shingo, decreasing for Shigar, and increasing for Shyok. The total number of glacial lakes increased in the basins of Astor, Gilgit and Taxkorgan, whereas the total area of glacial lakes remained stable in Astor and Gilgit basins and decreased in Taxkorgan basin. The total numbers of Kashgar and Hunza basins decreased, whereas the total area of glacial lakes remained stable in Kashgar and increased in the Hunza basin.

Table 4. Distributions in count and area (km^2) of glacial lakes among mountain ranges within the study area.

Source and year	Tien Shan	Karakoram	Pamir	Hindu Kush	Himalaya	Total
Landsat in 1990	10 (0.12)	370 (11.11)	178 (13.73)	780 (28.33)	816 (31.81)	2154 (85.10)
Landsat in 2000	7 (0.11)	393 (11.76)	163 (13.96)	792 (28.50)	829 (31.77)	2184 (86.10)
Landsat in 2020	5 (0.17)	334 (10.10)	182 (14.14)	835 (29.25)	878 (32.65)	2234 (86.31)
Sentinel in 2020*	11 (0.21)	479 (11.69)	262 (15.71)	880 (34.96)	959 (33.39)	2591 (95.96)

*Note: Glacial lake greater than 4500 m^2 are calculated for Sentinel-2 derived dataset in order to be in line with Landsat derived dataset.

5.2 Glacier lake distribution observed from Sentinel-2

Sentinel-derived results shows that there are 7,560 glacial lakes (103.70±8.45 km^2) in 2020 across the entire CPEC (Table 5) under a minimum mapping unit of 5 pixels (500 m^2). Similar to the pattern from Landsat mapping, the lake abundance extracted from Sentinel images is inversely related to lake size (following a typical Pareto distribution). The smallest size class (0.0005-0.0045 km^2) contains the maximum lake count (4,969) but the least lake

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

area (7.73±2.62 km^2) (Table 5), which is not available in the Landsat-derived lake data due to a coarser spatial resolution. In each size class, there are also a higher number of larger glacial lakes from Sentinel than that from Landsat images. The discrepancy is mainly attributed to inconsistency of image acquisition dates and spatial resolutions.

Table 5. Count and area of glacial lakes mapped from Sentinel and Landsat images in 2020 between various size classes

Lake size km^2	Glacial lakes from Sentinel count (km^2)	Glacial lakes from Landsat count (km^2)	Overlap % (%)
0.0005-0.0045	4969 (7.73±2.62)	—	—
0.0045-0.05	2182 (35.52±3.72)	1870 (31.47±9.57)	85.70 (88.60)
0.05-0.1	237 (16.37±0.89)	204 (14.07±2.18)	86.08 (85.95)
0.1-0.2	122 (16.88±0.68)	115 (15.91±1.83)	94.26 (94.25)
⩾0.2	50 (27.20±0.54)	45 (24.86±1.40)	90.00 (91.40)
Total	7560 (103.70±8.45)	2234 (86.31±14.98)	—

Compared with our Landsat-based product, glacial lakes from Sentinel-2 have similar distribution characteristics (Figure 9c and d) among mountain ranges, basins, types and altitudinal locations (Figure 10); meanwhile, a larger quantity of glacier lakes, with more accurate boundaries and a greater total lake area, were generated from Sentinel-2 images. Taking altitudinal distribution for example, the number and size of glacial lakes in the study area appear follow a normal distribution against elevation for both Sentinel-2 and Landsat derived products (Figure 10). The elevation of all glacial lakes mapped in 2020 based on Sentinel-2 images ranged from 2500 m to 5750 m (a.s.l.), with 89.58% between 3600 m and 5100 m and a mean altitude of 4421 m. The peak number appears between 4500 m and 4550 m whereas the maximum area emerges between 4250 m and 4300 m. The anomalously large area between 3600 and 3650 m shows up in Fig. 10b because of several disproportionally large lakes. Although Landsat derived lakes show a similar distribution pattern to Sentinel derived lakes, the lake count and area in each altitudinal band are greater in the Sentinel product due to the improved spatial resolution and image quality.

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

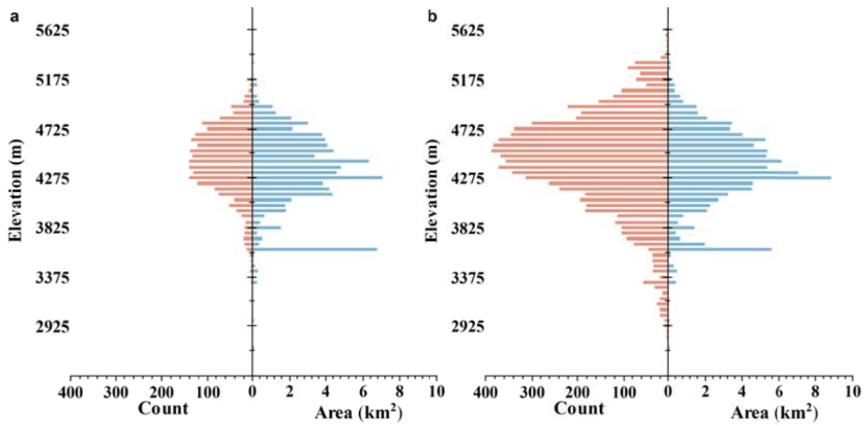


Figure 10. Altitudinal distribution of glacial lakes in 2020 derived from Landsat (a) and Sentinel images (b)

6 Discussions

6.1 Comparison of Sentinel-2 and Landsat derived products

Glacial lakes from Landsat and Sentinel images have a high consistency in number and area with overlap rates from 85.7% to 94.26% for all lakes greater than 0.0045 km^2 approximately (Table 5), implying a good potential for coordinated utility with Landsat archived observation (Figure 11). Lake extents extracted from Landsat and Sentinel images match well for various types and sizes (Table 4). The best consistency rate reaches 94% for the glacial lakes between 0.1 km^2 and 0.2 km^2. The difference in area of glacial lakes extracted from Landsat and Sentinel images generally lies within the uncertainty ranges.

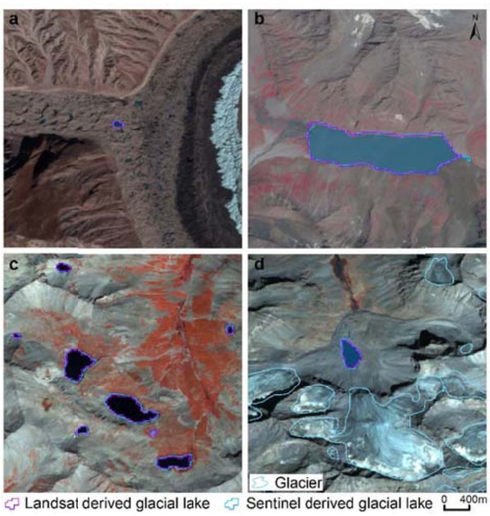


Figure 11. High consistency of lake extents extracted from Landsat and Sentinel images. Lake types

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

Open Access Earth System Science Data Discussions

shown include supraglacial (a), glacier-fed moraine-dammed (b), unconnected glacial erosion lake without glacier melt supply (c) and glacier-fed moraine-dammed (d).

Spatial resolution of satellite images plays a primary role in the discrepancies in count and area of glacial lakes extracted from Landsat (30 m) and Sentinel (10 m) observations. Due to a finer spatial resolution, Sentinel images can extract more glacial lakes and more accurate extents than those from Landsat images. We set the same 5 pixels as the minimum mapping unit for both Landsat and Sentinel images, which corresponds to a minimum area of 0.0045 km^2 and 0.0005 km^2, respectively. The minimum mapping area results in generating nearly 5000 more lakes from Sentinel images than from Landsat images, causing the greatest discrepancy in number of the two glacial lake products (Table 5), such as Figure 12a. Meanwhile, Sentinel images are able to depict boundaries of glacial lake with a lower uncertainty (Figure 12b-d). For example, some small islands and narrow channels (Figure 12b and c) were mapped from Sentinel imagery that are unable to be detected in Landsat imagery.

Different acquisition dates between Sentinel and Landsat images also contribute to the discrepancy of those two glacial lake datasets. Acquiring same-day images from the two sensors were not always possible due to the impacts of cloud contaminations, topographic shadows, snow cover and revisit periods (Williamson et al., 2018; Paul et al., 2020). Glacial lakes are changing temporally in the context of climate and glacier changes, taking supraglacial lakes for example that evolve dramatically in a short period (Figure 12e). Despite our efforts of leveraging all available high-quality images, the overlap of acquisition dates between Landsat and Sentinel images for the same location is relatively low in this study area, and the consequential temporal gaps led to a difference in the number and area of the derived glacial lakes.

Displacement between images also resulted in a certain degree of discrepancy between Landsat and Sentinel derived glacial lakes. All images used in this study have been orthorectified, but we still find that a few Sentinel images were not well matched with Landsat images, leading to the discrepancy between the two glacial lake datasets (Figure 12f). We manually georeferenced the shifted images to minimize the difference between Sentinel and Landsat derived glacial lakes (Figure 12f). Original geo-referencing accuracy is approximate half of one pixel for Landsat and Sentinel image, and this displacement likely contributes a minor error to glacial lake changes at various time periods. Although we could not eliminate this intrinsic error, the error has been considered in the uncertainty assessment of our glacial lake mapping.

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

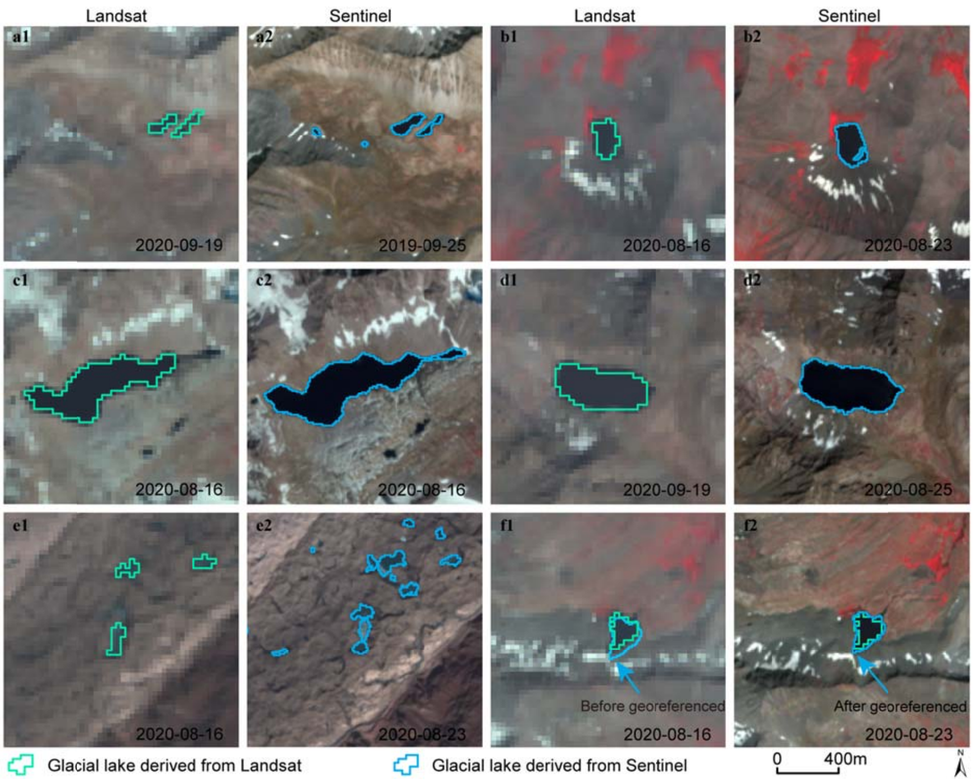


Figure 12. Discrepancy of lake extents extracted from Landsat and Sentinel images.

6.2 Comparison with other datasets

Glacial lake datasets play a fundamental role in GLOF risk evaluation, glacier change prediction, and water resource availability. An increasing number of glacier lake datasets have been released over the past years, and most of them were produced from long-term Landsat archives. Glacial lake datasets using Sentinel images are so scarce that we are unable to compare our product with other existing ones in the study area. Here we selected four available glacial lake datasets to compare with our Landsat-derived dataset.

Our study provides the latest glacial lake dataset (in 2020) and the most long-term Landsat observation (1990 to 2020) for this study, with a range of critical attributes including two types of classification systems. Within the same study area, our 2020 glacial lakes appear to be closest to the 2018 dataset produced by Wang et al (2020), with the highest overlap of greater than 74% in both number and area (Table 6). In Wang et al. (2020) , the minimum mapping unit is 6 pixels so their dataset has a smaller lake quantity. However, their dataset contains all lakes within 10 km of glacier boundaries, including many large landslide-dammed lakes that are excluded in our glacial lake mapping. As a result, their total glacier lake area is greater than ours. The overlapping rates between Wang's glacial lakes (2020) in 1990 and ours are more than 69% in both number and area. However, their results show a distinct increase of glacial lakes in number and area between 1990 and 2018 (Wang et

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

al., 2020) whereas our data show a more stable change between 1990 and 2020. One possible reason is that manually delineating glacial lakes twice by different operators during Wang's lake mapping (2020) exacerbates the errors of mapping. Another reason is that their data contains landslide-dammed lakes that fluctuate greatly with time and expanded recently. One example is the Attabad Lake (Located at 36°18'22.33"N, 74°49'34.36"E).

Table 6. Comparison of different glacial lake datasets sourced from Landsat images in the study area.

Acquisition year (period)	Method	MMU m^2 (pixels)	Count (km^2)	Overlap % (%)	Reference
1990 (1988-1993)	Manual	5400 (6)	1720 (89.68±13.69)	69.17 (76.33)	Wang et al., 2020
1990 (1990-1999)	Automated	50000 (55)	145 (20.28)	6.27 (21.66)	Shugar et al., 2020
1990 (1989-1992)	Manual	2700 (3)	622 (51.93±10.15)	27.72 (39.94)	Zhang et al., 2015
1990 (1989-1994)	*Automated & Manual*	*4500 (5)*	*2154 (85.10±14.66)*	—	*This study*
2000 (1999-2001)	Manual	2700 (3)	724 (61.41±11.91)	31.91 (46.97)	Zhang et al., 2015
2000 (2000-2004)	Automated	50000 (55)	155 (22.35)	6.78 (23.72)	Shugar et al., 2020
2008	Automated & Manual	8100 (9)	1067 (65.45)	44.14 (53.58)	Chen et al., 2021
2000 (1996-2004)	*Automated & Manual*	*4500 (5)*	*2184 (86.10±14.83)*	—	*This study*
2018 (2017-2018)	Manual	5400 (6)	1956 (102.46±15.48)	74.57 (85.63)	Wang et al., 2020
2015 (2015-2018)	Automated	50000 (55)	148 (21.45)	6.27 (22.97)	Shugar et al., 2020
2017	Automated & Manual	8100 (9)	1063 (63.23)	45.21 (57.78)	Chen et al., 2021
2020 (2016-2020)	*Automated & Manual*	*4500 (5)*	*2234 (86.31±14.98)*	—	*This study*

Note: MMU represents minimum mapping units.

The second highest overlapping rate is approximate 55% in area with Chen's data in 2008 and 2017 (Chen et al., 2021). However, the overlapping rate in number is nearly 45% due to their larger minimum mapping unit (9 pixels). Similarly, a minimum mapping unit of 55 pixels (50000 m^2) in Shugar et al.'s, dataset (2020) led to the lowest overlap with less than 24% in area. Zhang's dataset shows fewer glacial lakes in 1990 and 2000 even with a smaller minimum mapping unit of 3 pixels (Zhang et al., 2015). By inspecting their dataset, we attributed this anomalous discrepancy to a range of glacial lakes that were missed during their manual delineation as a result of insufficient high quality images in the earlier Landsat era. Our Landsat derived glacial lake dataset has been visually cross-checked over three time periods after the step of object-based automated lake mapping, and also been visually validated by Sentinel-2 derived glacial lakes. Through this series of quality assurance, we aim at delivering one of the most reliable multi-decadal glacial lake products for this study area.

Other factors, such as minimum mapping units, definition of glacial lakes and study areas, image quality and acquisition dates, mapping methods and quality assurance workflow, might also lead to the discrepancies between the glacial lake datasets. Despite such discrepancies, an increasing number of publically-shared datasets benefit potential users to select the most suitable one for their objectives. Herein, we provide an up-to-date glacial lake dataset derived from both Landsat and Sentinel observations, which further promoted the capacity of GLOFs risk assessment and predicting glacier evolutions in the context of climate change.

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

Open Access Earth System Science Data Discussions

6.3 Limitation and updating plan

We would like to acknowledge several limitations of our glacier lake dataset, largely due the availability of high quality satellite images in the study area and inadequate field survey data (Wang et al., 2020; Chen et al., 2021). First, it is unlikely to collect enough good-quality images within one calendar year for the entire study area due to high possibility of cloud or snow covers. Even though an capacity of repetitive observations for Landsat8 OLI and Sentinel-2 increased (Williamson et al., 2018; Paul et al., 2020; Roy et al., 2014; Wulder et al., 2019), the 2020 glacial lake dataset has to employ images acquired in other years besides 2020. Most images used from Landsat and Sentinel platforms were imaged in autumn, and some images taken between April and July and in November also were employed. Distribution and changes in glacial lakes primarily represent the characteristics between August and October. Glacial lakes evolve with time and space (Nie et al., 2017), and subtle inter- and intra-annual changes (Liu et al., 2020) in glacial lake dataset of each time period were ignored. Second, field investigation data are limited due to low accessibility of high mountain environment in the study area, which restrained the accuracy in classifying the glacial lake types. Although very high-resolution Google Earth images were utilized to assist in lake type interpretation, occasional misclassification was inevitable. We implemented two types of classification systems based on a careful utilization of glacier data, DEM, geomorphological features and expert knowledge. However, the lack of in situ survey prohibited a thorough validation of the glacial lake types.

7 Data availability

Our glacial lake dataset extracted from Sentinel-2 images in 2020 and Landsat observation between 1990 and 2020 are available online via the Mountain Science Data Center, the Institute of Mountain Hazards and Environment, the Chinese Academy of Sciences at https://doi.org/10.12380/Glaci.msdc.000001 (Lesi et al., 2022). The glacial lake dataset is provided in both ESRI shapefile format (total size of 22.6 MB) and the Geopackage format (version 1.2.1) with a total size of 9.2MB, which can be opened and further processed by open-source geographic information system software such as QGIS. The glacial lake dataset will be updated using newly collected Landsat and Sentinel images at a five-year interval or modified according to user feedbacks. The updated glacial lake dataset will continue to be released freely and publicly on the Mountain Science Data Center sharing platform.

8 Conclusions

Glacial lake inventories of the entire China-Pakistan Economic Corridor in 2020 were completed based on Landsat and Sentinel-2 images using a human-interactive and semi-automated mapping method. Both Landsat and Sentinel derived glacial lake datasets show similar characteristics in spatial distribution and in the statistics of count and area. By contrast, glacial lake dataset derived from Sentinel-2 images with a spatial resolution of 10 m has a lower mapping error and more accurate lake boundary than those from 30 m spatial resolution Landsat images whereas Landsat imagery is more suitable to analyze spatial-temporal changes at longer time scale due to its long-term archived observation at a consistent spatial resolution of 30 m started from around 1990.

Glacial lakes in the study area remain relatively stable with a slight increase in number and area between 1990 and 2020 according to Landsat observations. Our dataset reveals that 2154 glacial lakes in 1990 covering 85.1±14.66km^2 increased to 2234 lakes with a total area of 86.31±14.98km^2. The same mapping method and rigorous workflow of quality assurance and quality control used in this study reduced the error in multi-temporal changes of glacial lakes.

The Hanshaw's error estimation method for automated lake mapping was improved by removing repeatedly calculated edge pixels that vary with lake shape. Therefore, the newly proposed method reduces the estimated value of uncertainty from satellite observations.

Our glacial lake dataset contains a range of critical parameters that maximize their potential utility for GLOFs risk evaluation and glacier-lake evolution projection. The dual classification systems of glacial lake types were developed and are very likely to attract broader researchers and scientists to use our datasets. In comparison with other existing glacial lake datasets, our products were created through a thorough consideration of lake types, cross checks and rigorous quality assurance, and will be updated and released continuously in the data center of mountain science. As such, we expect that our glacial lake dataset will have significant values for cryospheric-hydrology research, assessment of glacier-related hazards and engineering project construction in the CPEC.

Supplement. The supplement related to this article is available online.

Author contributions. ML and YN conceived the study, ML, YN and XD performed data processing and analysis of the glacial lake inventory data, JW contributed to tool development and mapping methods, ML and YN wrote the manuscript. All authors reviewed and edited the manuscript before submission.

Competing interests. The authors declare no conflict of interest.

Acknowledgements.
This study was supported by the National Natural Science Foundation of China (Grant Nos. 42171086, 41971153), the International Science & Technology Cooperation Program of China (No. 2018YFE0100100), the Chinese Academy of Sciences "Light of West China" and Natural Sciences and Engineering Research Council of Canada (Grant No. DG-2020-04207).

References

Ashraf, A., Naz, R., Iqbal, M.B.: Altitudinal dynamics of glacial lakes under changing climate in the Hindu Kush, Karakoram, and Himalaya ranges. Geomorphology, 283: 72-79, https://doi.org/10.1016/j.geomorph.2017.01.033, 2017.

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

Azam, M.F., Kargel, J.S., Shea, J.M., Nepal, S., Haritashya, U.K., Srivastava, S., Maussion, F., Qazi, N., Chevallier, P., Dimri, A.P., Kulkarni, A.V., Cogley, J.G., Bahuguna, I.: Glaciohydrology of the Himalaya-Karakoram. Science, 373: eabf3668, https://doi.org/10.1126/science.abf3668, 2021.

Battamo, A.Y., Varis, O., Sun, P., Yang, Y., Oba, B.T., Zhao, L.: Mapping socio-ecological resilience along the seven economic corridors of the Belt and Road Initiative. J. Clean. Prod., 309: 127341, https://doi.org/10.1016/j.jclepro.2021.127341, 2021.

Bhambri, R., Hewitt, K., Kawishwar, P., Kumar, A., Verma, A., Snehmani, Tiwari, S., Misra, A.: Ice-dams, outburst floods, and movement heterogeneity of glaciers, Karakoram. Global Planet. Change, 180: 100-116, https://doi.org/10.1016/j.gloplacha.2019.05.004, 2019.

Bhattacharya, A., Bolch, T., Mukherjee, K., King, O., Menounos, B., Kapitsa, V., Neckel, N., Yang, W., Yao, T.: High Mountain Asian glacier response to climate revealed by multi-temporal satellite observations since the 1960s. Nat. Commun., 12: 4133, https://doi.org/10.1038/s41467-021-24180-y, 2021.

Bolch, T., Pieczonka, T., Mukherjee, K., Shea, J.: Brief communication: Glaciers in the Hunza catchment (Karakoram) have been nearly in balance since the 1970s. The Cryosphere, 11: 531-539, https://doi.org/10.5194/tc-11-531-2017, 2017.

Brun, F., Berthier, E., Wagnon, P., Kääb, A., 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.

Brun, F., Wagnon, P., Berthier, E., Jomelli, V., Maharjan, S.B., Shrestha, F., Kraaijenbrink, P.D.A.: Heterogeneous Influence of Glacier Morphology on the Mass Balance Variability in High Mountain Asia. J. Geophys. Res-Earth, 124: 1331-1345, https://doi.org/10.1029/2018JF004838, 2019.

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

Chen, X., Cui, P., You, Y., Cheng, Z., Khan, A., Ye, C., Zhang, S.: Dam-break risk analysis of the Attabad landslide dam in Pakistan and emergency countermeasures. Landslides, 14: 675-683, https://doi.org/10.1007/s10346-016-0721-7, 2017.

Emmer, A., Cuřín, V.: Can a dam type of an alpine lake be derived from lake geometry? A negative result. J. Mt. Sci.-Engl., 18: 614-621, https://doi.org/10.1007/s11629-020-6003-9, 2021.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D.: The Shuttle Radar Topography Mission. Rev. Geophys., 45: RG2004, https://doi.org/10.1029/2005RG000183, 2007.

Gardelle, J., Arnaud, Y., Berthier, E.: Contrasted evolution of glacial lakes along the Hindu Kush Himalaya mountain range between 1990 and 2009. Global Planet. Change, 75: 47-55, https://doi.org/10.1016/j.gloplacha.2010.10.003, 2011.

Hanshaw, M.N., 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.

Hewitt, K.: The Karakoram Anomaly? Glacier Expansion and the ‘Elevation Effect,’ Karakoram Himalaya. Mt. Res. Dev., 25: 332-340, https://doi.org/10.1659/0276-4741(2005)025[0332:TKAGEA]2.0.CO;2, 2005.

Hewitt, K., 2014. Glaciers of the Karakoram Himalaya： Glacial Environments, Processes, Hazards and Resources. Springer, Dordrecht.

Huggel, C., Kääb, A., Haeberli, W., Teysseire, P., Paul, F.: Remote sensing based assessment of hazards from glacier lake outbursts: a case study in the Swiss Alps. Can. Geotech. J., 39: 316-330, https://doi.org/10.1139/t01-099, 2002.

Hugonnet, R., Mcnabb, R., Berthier, E., Menounos, B., Nuth, C., Girod, L., Farinotti, D., Huss, M., Dussaillant, I., Brun, F., 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.

Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E., 2008. Hole-filled seamless SRTM data V4. 2008, International Centre for Tropical Agriculture (CIAT), available from http://srtm.csi.cgiar.org.

Jiang, S., Nie, Y., Liu, Q., Wang, J., Liu, L., Hassan, J., Liu, X., Xu, X.: Glacier Change, Supraglacial Debris Expansion and Glacial Lake Evolution in the Gyirong River Basin, Central Himalayas, between 1988 and 2015. Remote Sens.-Basel, 10: 986, https://doi.org/10.3390/rs10070986, 2018.

Kääb, A., Berthier, E., Nuth, C., Gardelle, J., Arnaud, Y.: Contrasting patterns of early twenty-first-century glacier mass change in the Himalayas. Nature, 488: 495-498, https://doi.org/10.1038/nature11324, 2012.

Lesi, M., Nie, Y., Shugar, D.H., Wang, J., Deng, Q., Chen, H.: Landsat and Sentinel-derived glacial lake dataset in the China-Pakistan Economic Corridor from 1990 to 2020. Mountain Science Data Center, https://doi.org/10.12380/Glaci.msdc.000001, 2022.

Li, Z., Deng, X., Zhang, Y.: Evaluation and convergence analysis of socio-economic vulnerability to natural hazards of Belt and Road Initiative countries. J. Clean. Prod., 282: 125406, https://doi.org/10.1016/j.jclepro.2020.125406, 2021.

Liu, Q., Mayer, C., Wang, X., Nie, Y., Wu, K., Wei, J., Liu, S.: Interannual flow dynamics driven by frontal retreat of a lake-terminating glacier in the Chinese Central Himalaya. Earth Planet. Sc. Lett., 546: 116450, https://doi.org/10.1016/j.epsl.2020.116450, 2020.

Lutz, A.F., Immerzeel, W.W., Shrestha, A.B., Bierkens, M.F.P.: Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation. Nat. Clim. Change, 4: 587-592, https://doi.org/10.1038/nclimate2237, 2014.

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

Lyons, E.A., Sheng, Y., Smith, L.C., Li, J., Hinkel, K.M., Lenters, J.D., Wang, J.: Quantifying sources of error in multitemporal multisensor lake mapping. Int. J. Remote Sens., 34: 7887-7905, https://doi.org/10.1080/01431161.2013.827343, 2013.

Martín, C.N.S., Ponce, J.F., Montes, A., Balocchi, L.D., Gorza, C., Andrea, C.: Proglacial landform assemblage in a rapidly retreating cirque glacier due to temperature increase since 1970, Fuegian Andes, Argentina. Geomorphology, 390: 107861, https://doi.org/10.1016/j.geomorph.2021.107861, 2021.

Maurer, J.M., Schaefer, J.M., Rupper, S., Corley, A.: Acceleration of ice loss across the Himalayas over the past 40 years. Sci. Adv., 5: eaav7266, https://doi.org/10.1126/sciadv.aav7266, 2019.

Mcfeeters, S.K.: The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int. J. Remote Sens., 17: 1425 - 1432 1996.

Nie, Y., Zhang, Y., Liu, L., Zhang, J.: Glacial change in the vicinity of Mt. Qomolangma (Everest), central high Himalayas since 1976. J. Geogr. Sci., 20: 667-686, https://doi.org/10.1007/s11442-010-0803-8, 2010.

Nie, Y., Sheng, Y., Liu, Q., Liu, L., Liu, S., Zhang, Y., 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., Liu, Q., Wang, J., Zhang, Y., Sheng, Y., Liu, S.: An inventory of historical glacial lake outburst floods in the Himalayas based on remote sensing observations and geomorphological analysis. Geomorphology, 308: 91-106, https://doi.org/10.1016/j.geomorph.2018.02.002, 2018.

Nie, Y., Liu, W., Liu, Q., Hu, X., Westoby, M.J.: Reconstructing the Chongbaxia Tsho glacial lake outburst flood in the Eastern Himalaya: Evolution, process and impacts. Geomorphology, 370: 107393, https://doi.org/10.1016/j.geomorph.2020.107393, 2020.

Nie, Y., Pritchard, H.D., Liu, Q., Hennig, T., Wang, W., Wang, X., Liu, S., Nepal, S., Samyn, D., Hewitt, K.,

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

Chen, X.: Glacial change and hydrological implications in the Himalaya and Karakoram. Nature Reviews Earth & Environment, 2: 91-106, https://doi.org/10.1038/s43017-020-00124-w, 2021.

Paul, F., Rastner, P., Azzoni, R.S., Diolaiuti, G., Fugazza, D., Le Bris, R., Nemec, J., Rabatel, A., Ramusovic, M., Schwaizer, G., Smiraglia, C.: Glacier shrinkage in the Alps continues unabated as revealed by a new glacier inventory from Sentinel-2. Earth System Science Data, 12: 1805-1821, https://doi.org/10.5194/essd-12-1805-2020, 2020.

Pfeffer, W.T., Arendt, A.A., Bliss, A., Bolch, T., Cogley, J.G., Gardner, A.S., Hagen, J., Hock, R., Kaser, G., Kienholz, C., Miles, E.S., Moholdt, G., Mölg, N., Paul, F., Radić, V., Rastner, P., Raup, B.H., Rich, J., Sharp, M.J.: The Randolph Glacier Inventory: a globally complete inventory of glaciers. J. Glaciol., 60: 537-552, https://doi.org/10.3189/2014JoG13J176, 2014.

Post, A., Mayo, L.R., 1971. Glacier dammed lakes and outburst floods in Alaska: U.S. Geological Survey Hydrologic Investigations Atlas 455, U.S. Geological Survey.

Pritchard, H.D.: Asia's shrinking glaciers protect large populations from drought stress. Nature, 569: 649-654, https://doi.org/10.1038/s41586-019-1240-1, 2019.

Quincey, D.J., Richardson, S.D., Luckman, A., Lucas, R.M., Reynolds, J.M., Hambrey, M.J., Glasser, N.F.: Early recognition of glacial lake hazards in the Himalaya using remote sensing datasets. Global Planet. Change, 56: 137-152, https://doi.org/10.1016/j.gloplacha.2006.07.013, 2007.

Rabus, B., Eineder, M., Roth, A., Bamler, R.: The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar. ISPRS J. Photogramm., 57: 241-262, https://doi.org/10.1016/S0924-2716(02)00124-7, 2003.

RGI Consortium: Randolph Glacier Inventory – A Dataset of Global Glacier Outlines: Version 6.0: Technical Report, https://doi.org/10.7265/N5-RGI-60, 2017.

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

Rounce, D.R., Hock, R., Shean, D.E.: Glacier Mass Change in High Mountain Asia Through 2100 Using the Open-Source Python Glacier Evolution Model (PyGEM). Front. Earth Sci, 7: 331, https://doi.org/10.3389/feart.2019.00331, 2020.

Roy, D.P., Wulder, M.A., Loveland, T.R., C. E., W., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T.A., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalskyy, V., Lee, Z.P., Lymburner, L., Masek, J.G., Mccorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H., Zhu, Z.: Landsat-8: Science and product vision for terrestrial global change research. Remote Sens. Environ., 145: 154-172, https://doi.org/10.1016/j.rse.2014.02.001, 2014.

Sakai, A.: Brief communication: Updated GAMDAM glacier inventory over high-mountain Asia. The Cryosphere, 13: 2043-2049, https://doi.org/10.5194/tc-13-2043-2019, 2019.

Shean, D.E., Bhushan, S., Montesano, P., Rounce, D.R., Arendt, A., Osmanoglu, 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.

Sheng, Y., Song, C., Wang, J., Lyons, E.A., Knox, B.R., Cox, J.S., Gao, F.: Representative lake water extent mapping at continental scales using multi-temporal Landsat-8 imagery. Remote Sens. Environ., 185: 129-141, https://doi.org/10.1016/j.rse.2015.12.041, 2016.

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

Shugar, D.H., Jacquemart, M., Shean, D., Bhushan, S., Upadhyay, K., Sattar, A., Schwanghart, W., Mcbride, S., de Vries, M., Mergili, M., Emmer, A., Deschamps-Berger, C., Mcdonnell, M., Bhambri, R., Allen, S., Berthier, E., Carrivick, J.L., Clague, J.J., Dokukin, M., Dunning, S.A., Frey, H., Gascoin, S., Haritashya, U.K., Huggel, C., Kaab, A., Kargel, J.S., Kavanaugh, J.L., Lacroix, P., Petley, D., Rupper, S., Azam, M.F., Cook, S.J., Dimri, A.P., Eriksson, M., Farinotti, D., Fiddes, J., Gnyawali, K.R., Harrison, S., Jha, M., Koppes, M., Kumar, A., Leinss, S., Majeed, U., Mal, S., Muhuri, A., Noetzli, J., Paul, F., Rashid, I., Sain, K., Steiner, J., Ugalde, F., Watson, C.S., Westoby, M.J.: A massive rock and ice avalanche caused the 2021 disaster at Chamoli, Indian Himalaya. Science, 373: 300-306, https://doi.org/10.1126/science.abh4455, 2021.

Ullah, S., You, Q., Ali, A., Ullah, W., Jan, M.A., Zhang, Y., Xie, W., Xie, X.: Observed changes in maximum and minimum temperatures over China- Pakistan economic corridor during 1980–2016. Atmos. Res., 216: 37-51, https://doi.org/10.1016/j.atmosres.2018.09.020, 2019.

Viviroli, D., Kummu, M., Meybeck, M., Kallio, M., Wada, Y.: Increasing dependence of lowland populations on mountain water resources. Nature Sustainability, 3: 917-928, https://doi.org/10.1038/s41893-020-0559-9, 2020.

Wang, J., Sheng, Y., Hinkel, K.M., Lyons, E.A.: Drained thaw lake basin recovery on the western Arctic Coastal Plain of Alaska using high-resolution digital elevation models and remote sensing imagery. Remote Sens. Environ., 119: 325-336, https://doi.org/10.1016/j.rse.2011.10.027, 2012.

Wang, J., Sheng, Y., Tong, T.S.D.: Monitoring decadal lake dynamics across the Yangtze Basin downstream of Three Gorges Dam. Remote Sens. Environ., 152: 251-269, https://doi.org/10.1016/j.rse.2014.06.004, 2014.

Wang, J., Sheng, Y., Wada, Y.: Little impact of the Three Gorges Dam on recent decadal lake decline across China's Yangtze Plain. Water Resour. Res., 53: 3854-3877, https://doi.org/10.1002/2016WR019817, 2017.

Wang, J., Song, C., Reager, J.T., Yao, F., Famiglietti, J.S., Sheng, Y., Macdonald, G.M., Brun, F., Schmied, H.M., Marston, R.A., Wada, Y.: Recent global decline in endorheic basin water storages. Nat. Geosci., 11: 926-932, https://doi.org/10.1038/s41561-018-0265-7, 2018.

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

Wang, X., Ding, Y., Liu, S., Jiang, L., Wu, K., Jiang, Z., Guo, W.: Changes of glacial lakes and implications in Tian Shan, Central Asia, based on remote sensing data from 1990 to 2010. Environ. Res. Lett., 8: 44052, https://doi.org/10.1088/1748-9326/8/4/044052, 2013.

Wang, X., Liu, S., Zhang, J.: A new look at roles of the cryosphere in sustainable development. Advances in Climate Change Research, 10: 124-131, https://doi.org/10.1016/j.accre.2019.06.005, 2019.

Wang, X., Guo, X., Yang, C., Liu, Q., Wei, J., Zhang, Y., Liu, S., Zhang, Y., Jiang, Z., Tang, Z.: Glacial lake inventory of high-mountain Asia in 1990 and 2018 derived from Landsat images. Earth System Science Data, 12: 2169-2182, https://doi.org/10.5194/essd-12-2169-2020, 2020.

Westoby, M.J., Glasser, N.F., Brasington, J., Hambrey, M.J., Quincey, D.J., Reynolds, J.M.: Modelling outburst floods from moraine-dammed glacial lakes. Earth-Sci. Rev., 134: 137-159, https://doi.org/10.1016/j.earscirev.2014.03.009, 2014.

Williamson, A.G., Banwell, A.F., Willis, I.C., Arnold, N.S.: Dual-satellite (Sentinel-2 and Landsat 8) remote sensing of supraglacial lakes in Greenland. The Cryosphere, 12: 3045-3065, https://doi.org/10.5194/tc-12-3045-2018, 2018.

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., Zhu, Z.: Current status of Landsat program, science, and applications. Remote Sens. Environ., 225: 127-147, https://doi.org/https://doi.org/10.1016/j.rse.2019.02.015, 2019.

Yao, C., Wang, X., Zhao, X., Wei, J., Zhang, Y.: Temporal and Spatial Changes of Glacial Lakes in the China-Pakistan Economic Corridor from 1990 to 2018. Journal of Glaciology and Geocryology, 42: 33-42,

https://doi.org/10.5194/essd-2021-468
Preprint. Discussion started: 1 February 2022

https://doi.org/https://doi.org/10.7522/j.issn.1000-0240.2020.0009, 2020.

Yao, T., Thompson, L., Yang, W., Yu, W.S., Gao, Y., Guo, X.J., Yang, X.X., Duan, K.Q., Zhao, H.B., Xu, B.Q., Pu, J.C., Lu, A.X., Xiang, Y., Kattel, D.B., Joswiak, D.: Different glacier status with atmospheric circulations in Tibetan Plateau and surroundings. Nat. Clim. Change, 2: 663-667, https://doi.org/10.1038/NCLIMATE1580, 2012.

Yao, X., Liu, S., Han, L., Sun, M., Zhao, L.: Definition and classification system of glacial lake for inventory and hazards study. J. Geogr. Sci., 28: 193-205, https://doi.org/10.1007/s11442-018-1467-z, 2018.

Zhang, G., Yao, T., Xie, H., Wang, W., Yang, W.: 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., Tian, B.: An automated method for glacial lake mapping in High Mountain Asia using Landsat 8 imagery. J. Mt. Sci.-Engl., 15: 13-24, https://doi.org/10.1007/s11629-017-4518-5, 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., Stoffel, M.: Increasing risk of glacial lake outburst floods from future Third Pole deglaciation. Nat. Clim. Change, 11: 411-417, https://doi.org/10.1038/s41558-021-01028-3, 2021.