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], Jianrong Fan[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

20 **Abstract.** The China-Pakistan Economic Corridor (CPEC) is one of the flagship projects of
21 the One Belt One Road Initiative, which faces threats from water shortage and mountain
22 disasters in the high-elevation region, such as glacial lake outburst floods (GLOFs). An up-to-
23 date high-quality glacial lake dataset with parameters such as lake area, volume, and type,
24 which is fundamental to water resource and flood risk assessments, and predicting glacier-
25 lake evolutions, is still largely absent for the entire CPEC. This study describes a glacial lake
26 dataset for the CPEC using a threshold-based mapping method associated with rigorous
27 visual inspection workflows. This dataset includes (1) multi-temporal inventories for 1990,
28 2000, and 2020 produced from 30 m resolution Landsat images, and (2) a glacial lake
29 inventory for the year 2020 at 10 m resolution produced from Sentinel-2 images. The results
30 show that, in 2020, 2234 lakes were derived from the Landsat images, covering a total area of
31 86.31 ± 14.98 km^2 with a minimum mapping unit of 5 pixels (4500 m^2), whereas 7560 glacial
32 lakes were derived from the Sentinel-2 images with a total area of 103.70 ± 8.45 km^2 with a
33 minimum mapping unit of 5 pixels (500 m^2). The discrepancy shows that Sentinel-2 can
34 detect a significant quantity of smaller lakes than Landsat due to its finer spatial resolution.
35 Glacial lake data in 2020 was validated by Google Earth-derived lake boundaries with a
36 median ($\pm$standard deviation) difference of 7.66 ± 4.96 % for Landsat-derived product and
37 4.46 ± 4.62 % for Sentinel-derived product. The total number and area of glacial lakes from
38 consistent 30 m resolution Landsat images remain relatively stable despite a slight increase
39 from 1990 to 2020. A range of critical attributes has been generated in the dataset, including
40 lake types and mapping uncertainty estimated by an improved Hanshaw's equation. This
41 comprehensive glacial lake dataset has the potential to be widely applied in studies on water
42 resource assessment, glacial lake-related hazards, and glacier-lake interactions, and is freely
43 available at https://doi.org/10.12380/Glaci.msdc.000001 (Lesi et al., 2022).

1 Introduction

45 Glaciers in High-mountain Asia (HMA) play a crucial role in regulating climate, supporting
46 ecosystems, modulating the release of freshwater into rivers, and sustaining municipal water
47 supplies (Wang et al., 2019; Viviroli et al., 2020), agricultural irrigation, and hydropower
48 generation (Pritchard, 2019; Nie et al., 2021). Most HMA glaciers are losing mass in the
49 context of climate change (Brun et al., 2017; Maurer et al., 2019; Shean et al., 2020;
50 Bhattacharya et al., 2021), therefore, unsustainable glacier melt and the passing of peak water
51 are reducing the hydrological role of glaciers (Huss and Hock, 2018) and impacting
52 downstream ecosystem services, agriculture, hydropower and other socioeconomic values
53 (Carrivick and Tweed, 2016; Nie et al., 2021). The present and future glacier changes not
54 only impact the water supply for the downstream area but also alter the frequency and
55 intensity of glacier-related hazards, such as glacier lake outburst floods (GLOFs) (Nie et al.,
56 2018; Rounce et al., 2020; Zheng et al., 2021), and rock and ice avalanches (Shugar et al.,
57 2021). Global glacial lake number and total area both increased between 1990 and 2018 in
58 response to glacier retreat and climate change (Shugar et al., 2020), affecting the allocation of
59 freshwater resources. The Indus is globally the most important and vulnerable water tower
60 unit where glaciers, lakes, and reservoir storage contribute about two-thirds of the water
61 supply (Immerzeel et al., 2020). Ice-marginal lakes store ~1% of total ice discharge in
62 Greenland and accelerate lake-terminating ice velocity by ~25% (Carrivick et al., 2022). An

63   increasing frequency and risk of GLOFs (Nie et al., 2021; Zheng et al., 2021) is threatening
64   the Asian population and infrastructures in the mountain ranges, such as the China-Pakistan
65   Economic Corridor (CPEC), as a flagship component of One Belt One Road Initiative
66   (Battamo et al., 2021; Li et al., 2021). The northern section of the CPEC passes through
67   Pamir, Karakoram, Hindu Kush, and Himalaya mountains where droughts and glacier-related
68   hazards are frequent and severe (Hewitt, 2014; Bhambri et al., 2019; Pritchard, 2019),
69   threatening local people, the existing, under-construction and planned infrastructures, such as
70   highways, hydropower plants, and railways. Understanding the risk posed by water shortage
71   and glacier-related hazards is a critical step toward sustainable development for the CPEC.
72   Glacial lake inventories with a range of attributes benefit water resource assessment and
73   disaster risk assessment related to glacial lakes (Wang et al., 2020; Carrivick et al., 2022),
74   and contribute to predicting glacier-lake evolution and cryosphere-hydrosphere interactions
75   under climate change (Nie et al., 2017; Brun et al., 2019; Maurer et al., 2019; Carrivick et al.,
76   2020; Liu et al., 2020). Remote sensing is the most viable way to map glacial lakes and detect
77   their spatio-temporal changes in the high-elevation zones where in situ accessibility is
78   extremely low (Huggel et al., 2002; Quincey et al., 2007). Studies in glacial lake inventories
79   using satellite observations have been heavily conducted at regional scales recently, such as
80   in the Tibetan Plateau (Zhang et al., 2015), the Himalaya (Gardelle et al., 2011; Nie et al.,
81   2017), the HMA (Wang et al., 2020; Chen et al., 2021), the Tien Shan (Wang et al., 2013),
82   the Alaska (Rick et al., 2022), the Greenland (How et al., 2021) and the northern Pakistan
83   (Ashraf et al., 2017). However, the latest glacial lake mapping in 2020 is still absent along the
84   CPEC. Among existing studies, Landsat archival images are the most widely used due to their
85   multi-decadal record of earth surface observations, reasonably high spatial resolution (30 m),
86   and publicly available distribution (Roy et al., 2014). Freely available Sentinel-2 satellite
87   images show a better potential than Landsat in glacial lake mapping and inventories due to
88   their higher spatial resolution (10 m) and global coverage, but have only been available since
89   late 2015 (Williamson et al., 2018; Paul et al., 2020). Glacial lake inventories using Sentinel-
90   2 images are relatively scarce at regional scales, and studies of the latest glacial lake mapping
91   as well as comparisons of glacial lake datasets derived from Sentinel-2 and Landsat
92   observations are still lacking.
93   Discrepancies between various glacial lake inventories (Zhang et al., 2015; Shugar et al.,
94   2020; Wang et al., 2020; Chen et al., 2021; How et al., 2021) result from differences in
95   mapping methods, minimum mapping units, the definition of glacial lakes, periods, data
96   sources and other factors. For example, the manual vectorization method was widely adopted
97   at the earlier stage for its high accuracy. However, it is time-consuming associated with high
98   labor intensity, and is only practical at regional scales (Zhang et al., 2015; Wang et al., 2020).
99   Automated and semi-automated lake mapping methods, such as multi-spectral index
100 classification (Gardelle et al., 2011; Nie et al., 2017; Zhang et al., 2018; How et al., 2021),
101 have been developed to improve the efficiency of glacial lake inventories using optical
102 images, although manual modification is often unavoidable to assure the quality of lake data
103 impacted by cloud cover, mountain shadows, seasonal snow cover and frozen lake surfaces
104 (Sheng et al., 2016; Wang et al., 2017, 2018). Backscatter images from Synthetic Aperture
105 Radar (SAR) (Wangchuk and Bolch, 2020; How et al., 2021) were used to remove the impact
106 of cloud cover for lake mapping. Besides, other approaches such as hydrological sink

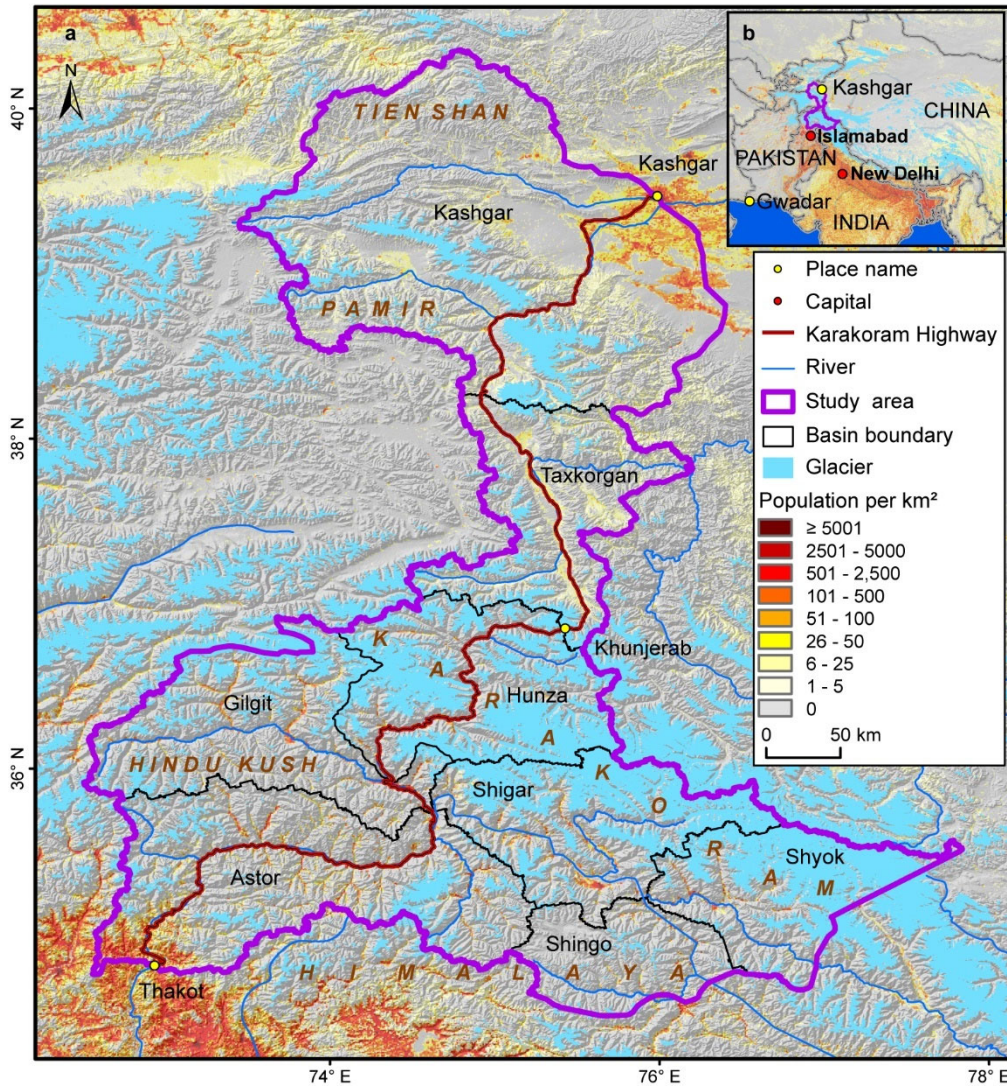
107 detection using DEM (How et al., 2021) and land surface temperature-based detection
108 method (Zhao et al., 2020) were also used for lake inventories. Different classification
109 methods impact the results of lake mapping and monitoring. So far, we are lacking a unified
110 standard for the classification system of glacial lakes (Yao et al., 2018). Existing
111 classification systems are generally used for their research purposes, mainly based on the
112 relative positions of glacial lakes and glaciers, the supply conditions of glaciers, and the
113 attributes of dams. In addition to different classification standards, the same type of glacial
114 lakes may also have different names given by different scholars. For example, ice-marginal
115 (Carrivick and Quincey, 2014; Carrivick et al., 2020), ice-contact (Carrivick and Tweed,
116 2013), and proglacial (Nie et al., 2017) lakes all represent glacial lakes sharing the boundary
117 with glaciers. Glacier lakes in currently available datasets have been traditionally categorized
118 by their spatial relationship with upstream glaciers (Gardelle et al., 2011; Wang et al., 2020;
119 Chen et al., 2021), and classification attributes considering the formation mechanism and the
120 properties of dams are rare or incomplete in the CPEC (Yao et al., 2018; Li et al., 2020).
121 Dam-type classification of glacial lakes provides a crucial attribute for glacier-lake
122 interactions and risk assessment (Emmer and Cuřín, 2021). Therefore, an up-to-date glacial
123 lake dataset with critical, quality-assured parameters (e.g. lake area, volume, and type) is
124 necessary.
125   This study aims to (1) present an up-to-date glacial lake dataset in the CPEC in 2020 using
126 both Landsat 8 and Sentinel-2 images to accurately document its detailed lake distribution;
127 (2) present two historical glacial lake datasets for the CPEC to show the extent in 1990 and
128 2000 using consistent 30-m Landsat images to reveal glacial lake changes at three time
129 periods (1990, 2000 and 2020); and (3) generate a range of critical attributes for glacial lake
130 inventories to benefit studies on water resource evaluation, risk assessment of GLOFs, glacier
131 –lake evolution modeling in the HMA.

2 Study area

Figure 1. Location of the study area associated with the distribution of glaciers (RGI Consortium, 2017),
mountains, basins, and population (Rose et al., 2021) (a), and its location within the CPCE (b).

The northern part of the CPEC is selected as the study area (Figure 1). The CPCE, originating
from Kashgar of the Xinjiang Uygur Autonomous region, China and extending to Gwadar Port,
Pakistan (Ullah et al., 2019; Yao et al., 2020), is connecting China and Pakistan via the only
Karakoram Highway. The study area covers all the drainage basins along Karakoram Highway
starting from Kashgar and ending at Thakot, with a total area of ~125,000 km². The upper Indus
basins beyond the Pakistani-administrated border are excluded from this study due to the spatial
coverage of the CPCE. The entire study area is divided into eight sub-basins, covering most of
the Karakoram with the highest elevation up to 8611 m, western Himalaya and Tien Shan,
eastern Hindu Kush, and the Pamir Mountains. The 9710 glaciers in the study area cover a total
area of 17,447 km² and nearly 60% of glaciers are distributed in the Karakoram (5818 glaciers
with a total area of 14,067.52 km²) (RGI Consortium, 2017). Most glaciers in the western
Himalaya and eastern Hindu Kush are losing mass in the context of climate change (Kääb et

149 al., 2012; Yao et al., 2012; Brun et al., 2017; Shean et al., 2020; Hugonnet et al., 2021), whereas
150 the glaciers in the eastern Karakoram and Pamir have shown unusually little changes, including
151 unchanged, retreated, advanced and surged glaciers (Hewitt, 2005; Kääb et al., 2012; Bolch et
152 al., 2017; Brun et al., 2017; Shean et al., 2020; Nie et al., 2021). The spatially heterogeneous
153 distribution and changes of glaciers are primarily explained as a result of differences in the
154 dominant precipitation-bearing atmospheric circulation patterns that include the winter
155 westerlies the Indian summer monsoon, their changing trends, and their interactions with local
156 extreme topography (Yao et al., 2012; Azam et al., 2021; Nie et al., 2021).

3 Data sources

158 Both Landsat and Sentinel-2 images have been employed to map glacial lakes between 1990
159 and 2020 in the CPEC (Figure 2). A total number of 71 Landsat Thematic Mapper (TM),
160 Thematic Mapper Plus (ETM+), and Landsat 8 Operational Land Imager (OLI) images with a
161 consistent spatial resolution of 30 m were downloaded from the United States Geological
162 Survey Global Visualization Viewer (GloVis, https://glovis.usgs.gov/) to be used to create
163 glacial lake inventories in 1990, 2000 and 2020. High-quality Landsat-5 images around 2010
164 are insufficient to cover the entire study area, so we were unable to map lakes in 2010 due to
165 Landsat-7's scan-line corrector errors and significant cloud covers. In addition, 39 Sentinel-2
166 images (23 scenes in 2020) were downloaded from Copernicus Open Access Hub
167 (https://scihub.copernicus.eu/) to produce the 10-m resolution glacial lake inventory in 2020.
168 All images used in this study have been orthorectified before download, but we still find that
169 one Sentinel-2 image was not well matched with Landsat images, leading to the discrepancy
170 between the two glacial lake datasets. We manually georeferenced the shifted image to
171 minimize the difference between Sentinel- and Landsat-derived glacial lakes.
172   Cloud and snow covers heavily affect the usability of optical satellite images (Wulder et
173 al., 2019) and their availability in the entire study area, so we took advantage of the images
174 acquired before and after each of the baseline years 1990, 2000 and 2020 to construct the
175 glacial lake inventories. Only 4 images in 1990 (the largest covering the study area), 16
176 images in 2000, and 23 images in 2020 were used for matching baseline year. Spatially, high-
177 quality images in given baseline years were preferentially chosen, or we selected one or more
178 alternative images acquired in adjacent years to delineate glacial lakes by removing the effect
179 of cloud and snow covers. To minimize the impact of intra-annual changes on glacial lakes,
180 most of the used images (82% for Sentinel-2 and 75% for Landsat) were acquired from
181 August to October in the given baseline year with cloud coverage of <20% for each image.
182 For some specific scenes where cloud cover exceeded the threshold of 20%, we selected
183 more than one image to remedy the effect of cloud contamination (Nie et al., 2010, 2017;
184 Jiang et al., 2018).
185   Other datasets used include the Randolph Glacier Inventory version 6.0 (Pfeffer et al.,
186 2014; RGI Consortium, 2017) and the Glacier Area Mapping for Discharge from the Asian
187 Mountains (GAMDAM) glacier inventory (Sakai, 2019). These two glacier datasets were
188 used to determine glacial lake types, such as ice-contact, ice-dammed, and unconnected-
189 glacier-fed lakes. The Shuttle Radar Topography Mission Digital Elevation Model (SRTM
190 DEM) at a 1-arc second (30 m) resolution (Jarvis et al., 2008) was employed to extract the
191 altitudinal characteristics of the glacial lakes. The absolute vertical accuracy of the SRTM

192 DEM is 16 m (90%) (Rabus et al., 2003; Farr et al., 2007). We also applied other published
193 glacial lake datasets for comparative analysis. They include the glacial lake inventories of
194 HMA in 1990 and 2018 downloaded from http://doi.org/10.12072/casnw.064.2019.db (Wang
195 et al., 2020), the Third Pole region in 1990, 2000, and 2010 publicly shared at
196 http://en.tpedatabase.cn/ (Zhang et al., 2015), the Tibet Plateau from 2008 to 2017 accessed at
197 https://doi.org/10.5281/zenodo.3700282 (Chen et al., 2021), and the entire world in 1990,
198 2000 and 2015 provided at https://nsidc.org/data /HMA_ GLI/versions/1 (Shugar et al.,
199 2020). In addition, field survey data collected between 2017 and 2018 were also used to assist
200 in lake mapping and glacial lake type classification.
201

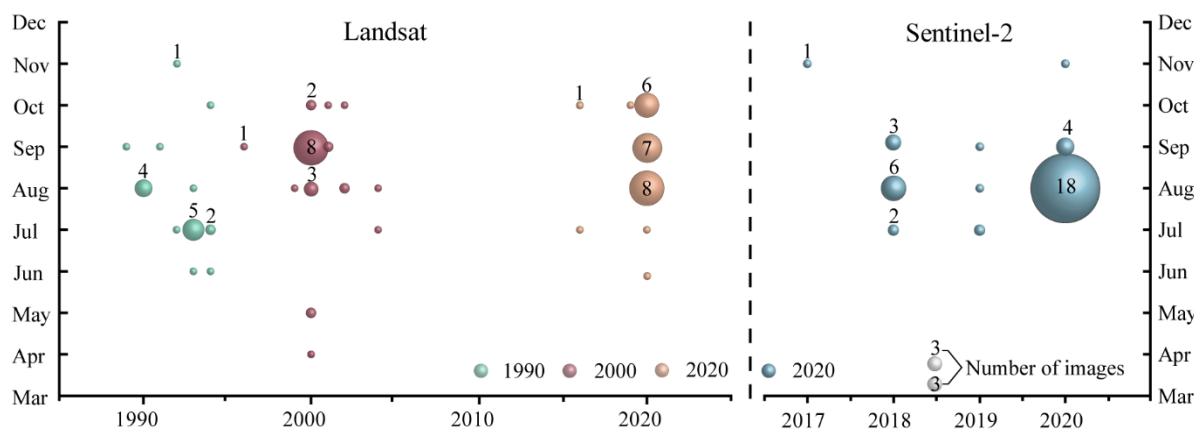



Figure 2. Acquisition of years and months of Landsat and Sentinel-2 images selected for glacial lake
inventories. The bubble size indicates the available high-quality image number.

4 Glacial lake inventory methods

4.1 Definition of glacial lakes

207 We consider a glacial lake as one that formed as a result of modern or ancient glaciation.
208 Contemporary glacial lakes are easily recognized using a combination of glacier inventories
209 and remote sensing images. Ancient glacial lakes can be identified from periglacial
210 geomorphological characteristics, including moraine remnants and U-shaped valleys that are
211 discernible from satellite observations (Post and Mayo, 1971; Westoby et al., 2014; Nie et al.,
212 2018; Martín et al., 2021). A 10-km buffering distance of RGI 6.0 glacier boundaries that has
213 been widely used in previous studies (Zhang et al., 2015; Wang et al., 2020), was created to
214 help map glacial lakes. A few glacial lakes in the study area (a total of 84 lakes for the
215 Sentinel-2 dataset and 55 lakes for the Landsat dataset in 2020) beyond the buffering zone,
216 located near buffering boundaries, were intentionally included due to clear evidence of
217 glaciation (Figure 3). Landslide-dammed lakes (Chen et al., 2017) in the buffering zone were
218 excluded from our inventories because of their irrelevance to glaciation. All glacial lakes in
219 the study area were mapped according to our definition. We were able to implement this
220 definition by carefully leveraging the spectral properties of glacial lakes and the periglacial
221 geomorphological features that are often evident in remote sensing images (see more in
222 sections 4.3 and 4.4).
223

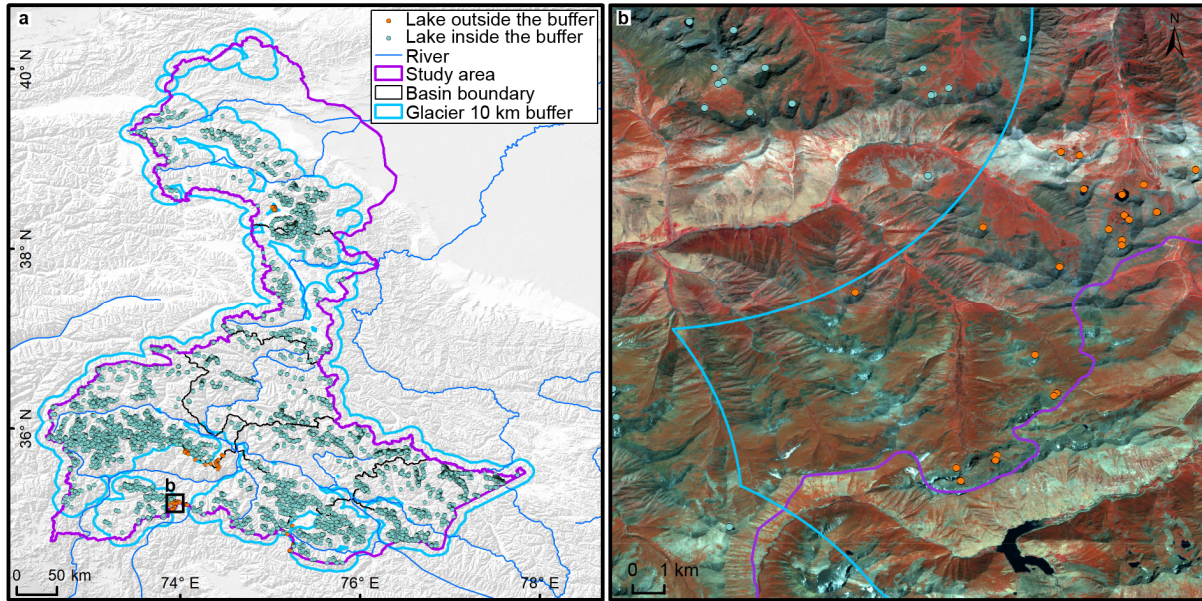



Figure 3. The 10-km buffer zone of RGI 6.0 glacier boundaries (a) and Sentinel-derived glacial lakes located near buffering boundary within the study area (b).

4.2 Interactive lake mapping

A human-interactive and semi-automated lake mapping method (Wang et al., 2014; Nie et al., 2017, 2020) was adopted to accurately extract glacial lake extents using Landsat and Sentinel-2 images, based on the Normalized Difference Water Index (NDWI) (Mcfeeters, 1996). The NDWI uses the green and near-infrared bands and is calculated by the following equation:

$$NDWI = \frac{Band_{Green} - Band_{NIR}}{Band_{Green} + Band_{NIR}} \tag{1}$$

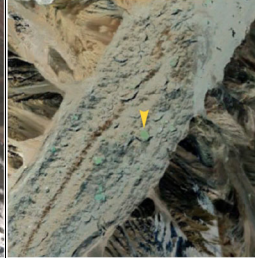
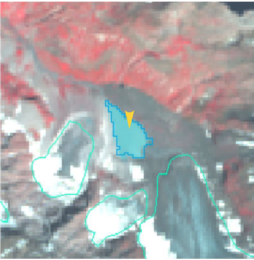
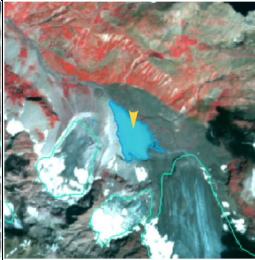
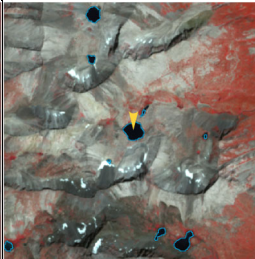
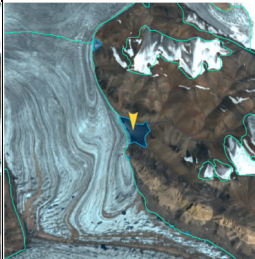
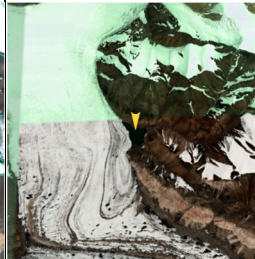
where the green band and near-infrared band were provided by both Landsat and Sentinel multispectral images.

Specifically, the method calculated the NDWI histogram based on the pixels with each user-defined and manually-drawn region of interest. The NDWI threshold that separates the lake surface from the land was interactively determined by screening the NDWI histogram against the lake region in the imagery (Wang et al., 2014; Nie et al., 2020). This way, the determined NDWI threshold can be well-tuned to adapt to various spectral conditions of the studied glacier lakes. The raster lake extents segmented by the thresholds were then automatically 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), 6 pixels (Wang et al., 2020), or 9 pixels (Chen et al., 2021) for a regional scale, or 55 pixels (Shugar et al., 2020) for a global scale. While a smaller threshold leads to a large number of lakes mapped, it also generates larger mapping noises or uncertainties.

252 Considering this signal-noise balance and our focus on identifying prominent glacier lake
253 dynamics in the study area, we opted to use 5 pixels as the MMU for both Landsat and
254 Sentinel-2 images.
255   Several procedures were taken to assure the quality assurance and quality control for lake
256 mapping, including 1) visual inspection and modification using the threshold-based mapping
257 method for each lake according to Landsat, Sentinel-2, and Google Earth high-resolution
258 images overlaying preliminarily lake boundary extraction at the given period; 2) time series
259 check for Landsat-derived glacial lake datasets from 1990 and 2020, and cross-check
260 between Landsat and Sentinel-2-derived lake dataset in 2020 to reduce errors of omission and
261 commission; 3) topological validation of glacial lake mapping, such as repeated removal,
262 elimination of small sliver polygons; and 4) logical check for lake types between two
263 classification systems of glacial lakes. False lake extents resulting from cloud or snow cover,
264 lake ice, and topographic shadows (Nie et al., 2017, 2020) were modified using the previous
265 semi-automated mapping method based on alternative images acquired in adjacent years.
266 Those procedures were time-consuming but helped to minimize the effect of cloud and snow
267 covers, and lake mapping errors, and to maximize the quality of the produced lake product
268 and the derived glacial lake changes.

269 4.3 Classification of glacial lakes

270 Two glacial lake classification systems (GLCS) have been established based on the
271 relationship of interaction between glacial lakes and glaciers as well as lake formation
272 mechanism and dam material properties. In the first GLCS (GLCS1), glacial lakes were
273 classified into four types based on their spatial relationship to upstream glaciers: supraglacial,
274 ice-contact, unconnected-glacier-fed lakes, and non-glacier-fed lakes according to Gardelle et
275 al. (2011) and Carrivick et al. (2013). Alternatively, combining the formation mechanism of
276 glacial lakes and the properties of natural dam features, glacial lakes were classified into five
277 categories (herein named GLCS2) modified from Yao's classification system (2018):
278 supraglacial, end-moraine-dammed, lateral-moraine-dammed, glacial-erosion lakes and ice-
279 dammed lakes. Subglacial lakes were excluded due to the mapping challenge from spectral
280 satellite images alone. Characterization and examples for each type are provided in Table 1
281 and Table 2. Individual glacial lakes were categorized into the specific types for each GLCS
282 according to available glacier inventory data, and geomorphological and spectral
283 characteristics interpreted from Landsat, Sentinel, and Google Earth images. The synergy of
284 these two GLCSs is beneficial to predicting glacier-lake evolutions and providing
285 fundamental data for water resource and glacial lake disaster risk assessment.
286

287   **Table 1.** A classification system of glacial lake types (GLCS1) according to the relationship between

288   glacial lakes and glaciers (© Google Earth 2019). Glacier outlines are from RGI 6.0 (RGI Consortium,

289   2017), and the yellow marker represents the target lake.

Lake types	Characteristics	Landsat	Sentinel-2	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			
Ice-contact	Lakes are dammed by moraine, ice, or bedrock, supplied by glacial meltwater, and shared boundary with glaciers. Case location: 39°09'32.40" N 73°43'12.00" E			
Unconnected-glacier-fed	Lakes are currently supplied by upstream glacial meltwater but disconnected from 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			

290

Table 2. A classification system of glacial lake types (GLCS2) according to the formation mechanism of glacial lakes and dam material properties (© Google Earth 2019). The glacier outlines from RGI 6.0 (RGI Consortium, 2017), and the yellow marker represents the target lake.

Lake types	Characteristics	Landsat	Sentinel-2	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 are dammed by debris, different from an ice-dammed 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-dammed	Lakes formed behind glaciers, dammed by glacier ice (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 18 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, lake water volume 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. The 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-2 image associated
with 5 digit code of military grid reference system. SceneID indicated identifying information
of image source for Landsat or Sentinel-2, consisting of the orbital number, sensor ID, and
acquisition date (YYYYMMDD) for Landsat image, or the orbital number and acquisition date
(YYYYMMDD) for Sentinel-2 image. Area and perimeter were automatically calculated based
on glacial lake extents. Lake water volume was estimated by an area-volume empirical
equation (Cook and Quincey, 2015). 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 the appendix tutorial. Located
country, sub-basin, and the mountain range of each glacial lake were identified by overlapping
the geographic boundaries of countries, basins, and mountain ranges.

Table 3. Attributes of glacial lake dataset.

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	Elevation of the centroid of glacial lake polygon	Unit: meter above sea level
SceneID	String	Scene ID of image source for Landsat or Sentinel-2	PxxxRxxx_xxxDYYYYMMDD or Txxxxx_YYYYMMDD
ACQDATE	String	The acquisition date of the source image	YYYYMMDD
GLCS1	String	The first classification system of glacial lakes based on the relationship of interaction between glacial lakes and glaciers	Supraglacial, Ice-contact, Unconnected-glacier-fed, and None-glacier-fed
GLCS2	String	The second classification system of glacial lakes is based on lake formation mechanism and dam material properties	Supraglacial, End-moraine-dammed, Lateral-moraine-dammed, Glacial-erosion and Ice-dammed
Basin	String	Basin name where the glacial lake locates in	

Field Name	Type	Description	Note
Mountain	String	Mountain name where the glacial lake locates in	
Country	String	Country name where the glacial lake locates in	
Perimeter	Double	The perimeter of the glacial lake boundary	Unit: meter
Area	Double	Area of glacial lake coverage	Unit: square meter
AreaUncer	Double	Area uncertainty of glacial lake mapping estimated based on modified Hanshaw's equation (2014)	Unit: square meter
Operator	String	The operator of the glacial lake dataset	Muchu, Lesi
Examiner	String	Examiner of glacial lake dataset	Yong, Nie
Volume	Double	The water volume of a glacial lake estimated by an area-volume empirical equation	Unit: cubic meter

316

4.5 Error and uncertainty assessment

4.5.1 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 (Carrivick and Quincey, 2014; Hanshaw and Bookhagen, 2014; Wang et al., 2020). 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 below:

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

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

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 coefficient of 0.6872 (1σ), which means nearly 69% of the edge pixels are subject to errors (Hanshaw and Bookhagen, 2014), 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

333   individual glacial lake.

334   In the original equation 2, the number of edge pixels varies by the shape of the lake and is

335   indicated by $\frac{P}{G}$. However, the pixels in the corner are double-counted (Figure 4). The total

336   number of repeatedly calculated edge pixels equals the number of inner nodes. Therefore, we

337   adjusted the calculation of the actual number of edge pixels as the maximum of edge pixels

338    $(\frac{P}{G})$ subtracting the number of inner nodes. Accordingly, the equation of uncertainty

339   estimation for lake mapping is modified as below:

340
$$Error(1\sigma) = (\frac{P}{G} - N_{Inner}) \times 0.6872 \times \frac{G^2}{2} \tag{4}$$

341   Where N_{Inner} is the number of inner nodes (inflection points) of each lake. The modified

342   equation is also suitable for lakes with islands (as illustrated in Figure 4b).

343   For polygons without islands (Figure 4a), use the following equation:

344
$$N_{Inner} = (\frac{N_{Total} - 4 - 1}{2}) \tag{5}$$

345   N_{Total} is the total number of nodes, including both the outer and inner. N_{Total} is calculated by

346   the "Field Calculator" in ArcGIS, in some cases, it is necessary to remove the redundant

347   nodes before calculating the total number of nodes (See the Appendix for more details). An

348   inner node is a polygon vertex where the interior angle surrounding it is greater than 180

349   degrees. An outer node is the opposite of the inner node, where the interior angle is less than

350   180 degrees. We found that the outer nodes are usually four more than the inner nodes in our

351   glacial lake dataset. The total nodes in ArcGIS contain one overlapping node to close the

352   polygon, meaning the endpoint is also the start point. This extra count was deleted from the

353   calculation (equation 5).

354   For polygons with island (Figure 4b) use the following equation:

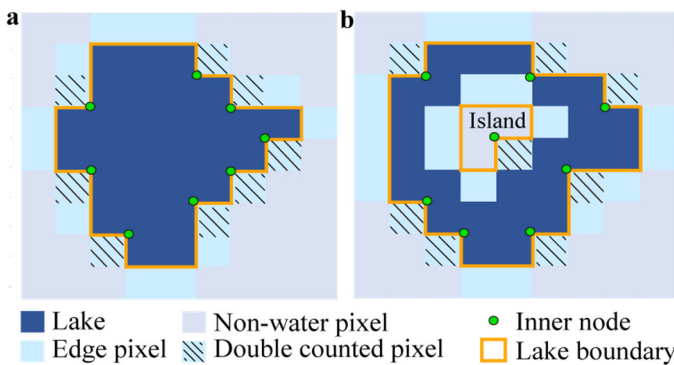
355
$$N_{Inner} = (\frac{N_{Total} - (N_{Island} + 1) \times 5}{2}) \tag{6}$$

356   N_{Island} is the number of islands within each polygon. A calculation method of N_{Island} is

357   given in the Appendix.

358

    ■ Lake   Non-water pixel   • Inner node
    Edge pixel   Double counted pixel   □ Lake boundary

359

360   **Figure 4.** Sketch of estimating the actual edge pixels for uncertainty calculation of individual glacial lakes

361   (with (a) and without islands (b)).

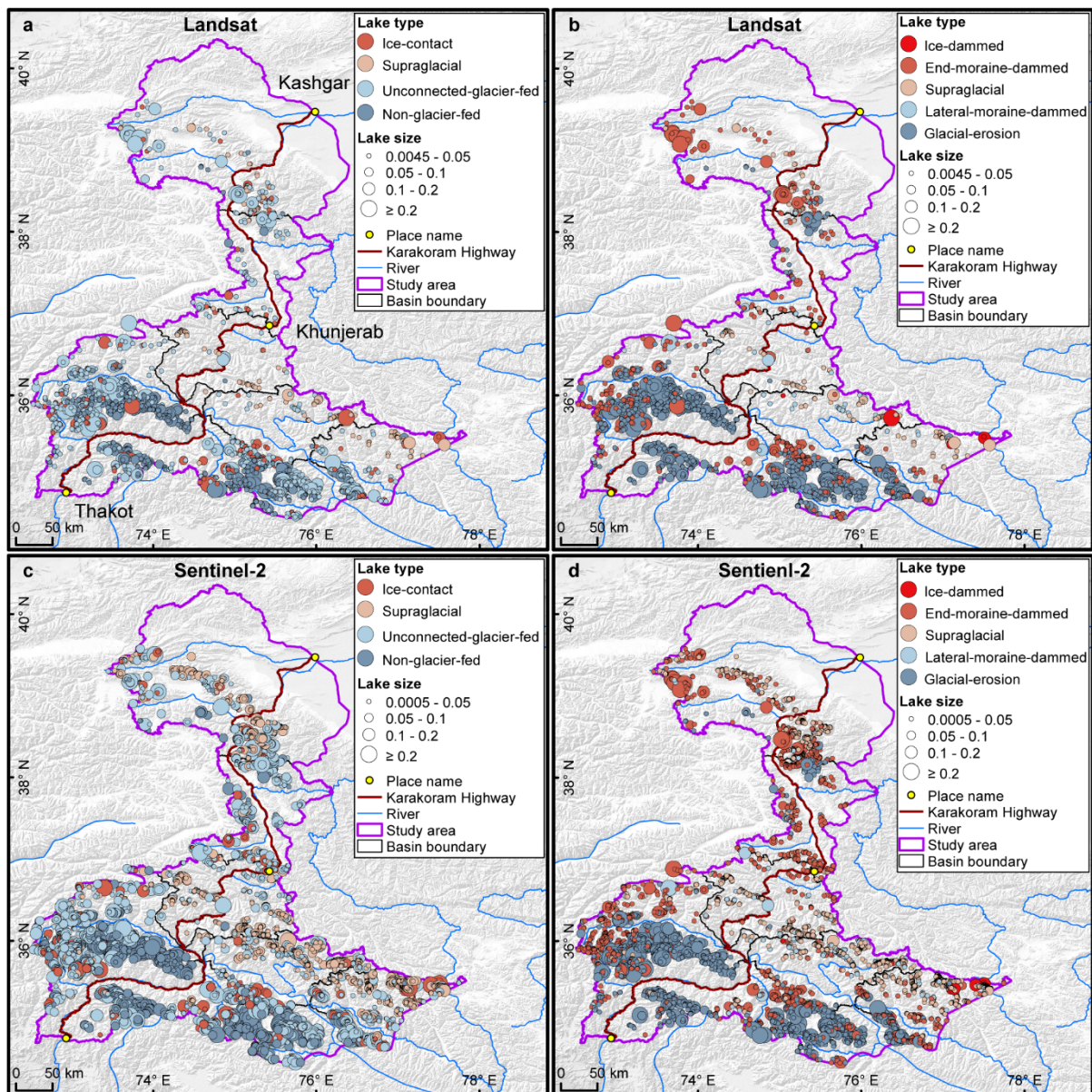
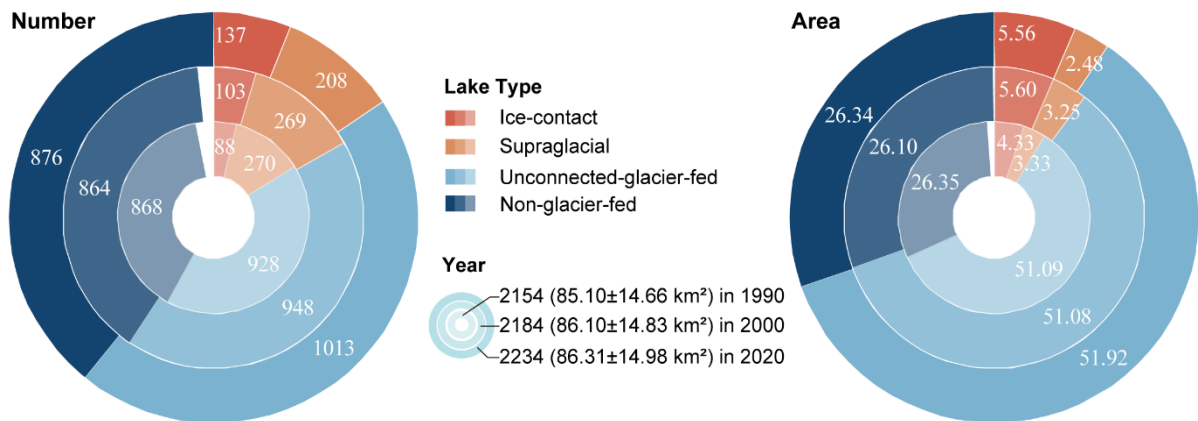
362

363   4.5.2 Validation of glacial lake mapping

364   A total of 89 glacial lakes were selected by stratified random sampling and manually digitized
365   based on the Google Earth images in circa 2020 with a spatial resolution of ～2 m acquired
366   from WorldView, GeoEye, Pleiades, etc. satellites to further validate the absolute error of our
367   glacial lake products in 2020 due to lacking field measurements for glacial lakes in the study
368   area. During the sampling, we set a minimum lake area to be 4500 m^2 and a relative
369   difference between Landsat- and Sentinel-derived lake areas of less than 18% (nearly
370   equaling the average relative error of ±17.36% for Landsat lake mapping) to minimize the
371   effect of lake changes from multi-temporal satellite observations in circa 2020. The 89
372   sample lakes range from 0.005 km^2 to 0.802 km^2 with a median (standard deviation) size of
373   0.047±0.134 km^2 and a total area of 8.033 km^2 for Landsat-derived dataset, and range from
374   0.005 km^2 to 0.849 km^2 with a median (standard deviation) size of 0.045±0.144 km^2 and a
375   total area of 8.447 km^2 for Sentinel-derived dataset.
376

5 Results

378   5.1 Glacier lake distribution and changes observed from Landsat

379   We mapped 2,234 glacial lakes for 2020 across the studied CPEC from Landsat-8 images,
380   with a total area of 86.31±14.98 km^2 (Figure 5a and b). Unconnected-glacier-fed lakes are
381   dominant in the first classification system, followed by non-glacier-fed lakes (Figure 6)
382   whereas glacial-erosion lakes dominate at both number (1478) and area (57.02 km^2) in the
383   second classification system (Figure 7), followed by end-moraine-dammed lakes and
384   supraglacial lakes. Among the classified lakes, 137 are ice-contact lakes and cover an area of
385   5.56 km^2, implying a higher mean size of ice-contact lakes than supraglacial lakes.
386

387
388 **Figure 5.** Distribution of glacial lakes in 2020 extracted from Landsat (a, b) and Sentinel-2 (c, d) images.
389 Panels a and c are classified by GLCS1 and GLCS2 for sub-graph b and d.
390

391
392 **Figure 6.** The number and area of different types of glacial lakes are classified based on the condition of

glacier supply in the study area (GLCS 1). The outermost ring represents glacial lake data in 2020, the middle ring for 2000, and the innermost ring for 1990. Lake number and area in 2020 were selected as references, meaning a concept of "100 %" for a complete ring. Labeled values are scaled in degrees rather than the radius of rings.

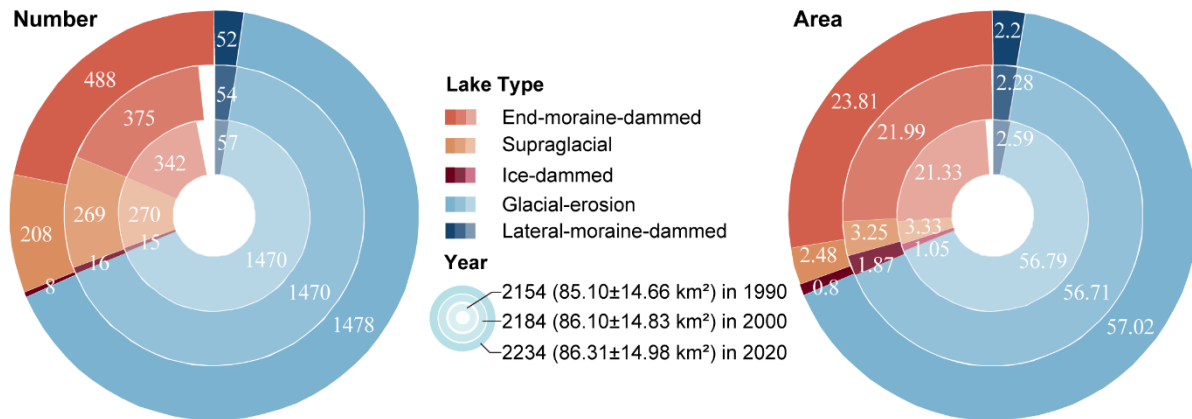



Figure 7. The number and area of different types of glacial lakes are classified based on glaciation and the nature of the dam in the study area (GLCS 2). The outermost ring represents glacial lake data in 2020, the middle ring for 2000, and the innermost ring for 1990. Lake number and area in 2020 were selected as references, meaning a concept of "100 %" for a complete ring. Labeled values are scaled in degrees rather than the radius of rings.

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 6 and Figure 7). From 1990 to 2020, the total number of glacial lakes increased by 80 or 3.70%, while the area grew by 1.21 km^2 (or 1.42%). In GLCS1, unconnected-glacier-fed lakes have the largest increase in number, followed by ice-contact and non-glacier-fed lakes, whereas supraglacial lakes decreased by 62 in count. Ice-contact lakes expanded by 1.24 km^2 (equaling an increase of 26% in ice-contact lakes), contributing 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 6). In GLCS2, end-moraine-dammed lakes increased by 2.48 km^2 and contributed most of the glacier lake area expansion, whereas supraglacial, ice-dammed, and lateral-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 period and remained stable (Figure 7).

5.2 Glacier lake distribution observed from Sentinel-2

Sentinel-derived results show that there are 7,560 glacial lakes (103.70±8.45 km^2) in 2020 across the entire CPEC under an MMU of 5 pixels (500 m^2). Compared with Landsat-derived product, glacial lakes from Sentinel-2 have similar spatial distribution characteristics (Figure 5); meanwhile, a larger quantity of glacier lakes, with more accurate boundaries and a greater total lake area, were generated from Sentinel-2 images (Table 4). The smallest size class

426 (0.0005-0.0045 km^2) contains the maximum lake number (4,969) but the least lake area
427 (7.73±2.62 km^2), which is not available in the Landsat-derived lake data due to a coarser
428 spatial resolution. In each size class, the overlap ratios are greater than 85% in count and
429 area, and there are also a higher number and larger area of glacial lakes from Sentinel than
430 that from Landsat images. Sentinel-2 images (10 m) with a finer spatial resolution produce
431 more glacial lakes than those from Landsat images (30 m). The discrepancy is mainly
432 attributed to the inconsistency of spatial resolutions and image acquisition dates, as discussed
433 in section 6.2.
434
435 **Table 4.** Count and area of glacial lakes mapped from Sentinel-2 and Landsat images in 2020 in various
436 size classes.

Lake size km^2	Glacial lakes from Sentinel-2 count (km^2)	Glacial lakes from Landsat count (km^2)	Overlap % (%)
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	2591 (95.97±5.83)	2234 (86.31±14.98)	86.22 (89.93)

437 Note: Second column excludes 4969 (7.73±2.62 km^2) lakes in the 0.0005 to 0.0045 km^2 range. Overlap % (%) represents the
438 ratios between our Landsat-derived dataset and Sentinel-derived product in count and area, respectively.

6 Discussions

440 6.1 Uncertainty and error of lake mapping

441 The uncertainty estimated from our improved equation shows that the relative error of
442 individual glacial lakes decreases when lake size increases or the cell size of remote sensing
443 images reduces (Lyons et al., 2013; Carrivick and Quincey, 2014) (Figure 8). Total area
444 errors of glacial lakes in the study area are approximate ±14.98 km^2 and ±8.45 km^2 in 2020
445 for Landsat and Sentinel-2 datasets, respectively, and the average relative errors are ±17.36%
446 and ±8.15%. Generally, small lakes have greater relative errors. For example, the mean
447 relative error is 35.38% for Landsat-derived glacial lakes between 0.0045 and 0.1 km^2 and
448 10.63% for glacial lakes greater than 0.1 km^2. The mean area error of Sentinel-derived glacial
449 lakes is almost one-third of that extracted from Landsat images for glacial lakes of all or
450 specific size groups. Because the relative error was estimated as a function of satellite image
451 spatial resolution and lake perimeter, the calculated error for a large lake is proportionally
452 smaller than that of a small lake (Salerno et al., 2012) and the error for Landsat-derived lake
453 is naturally greater than that of Sentinel-derived lake at the same size group.

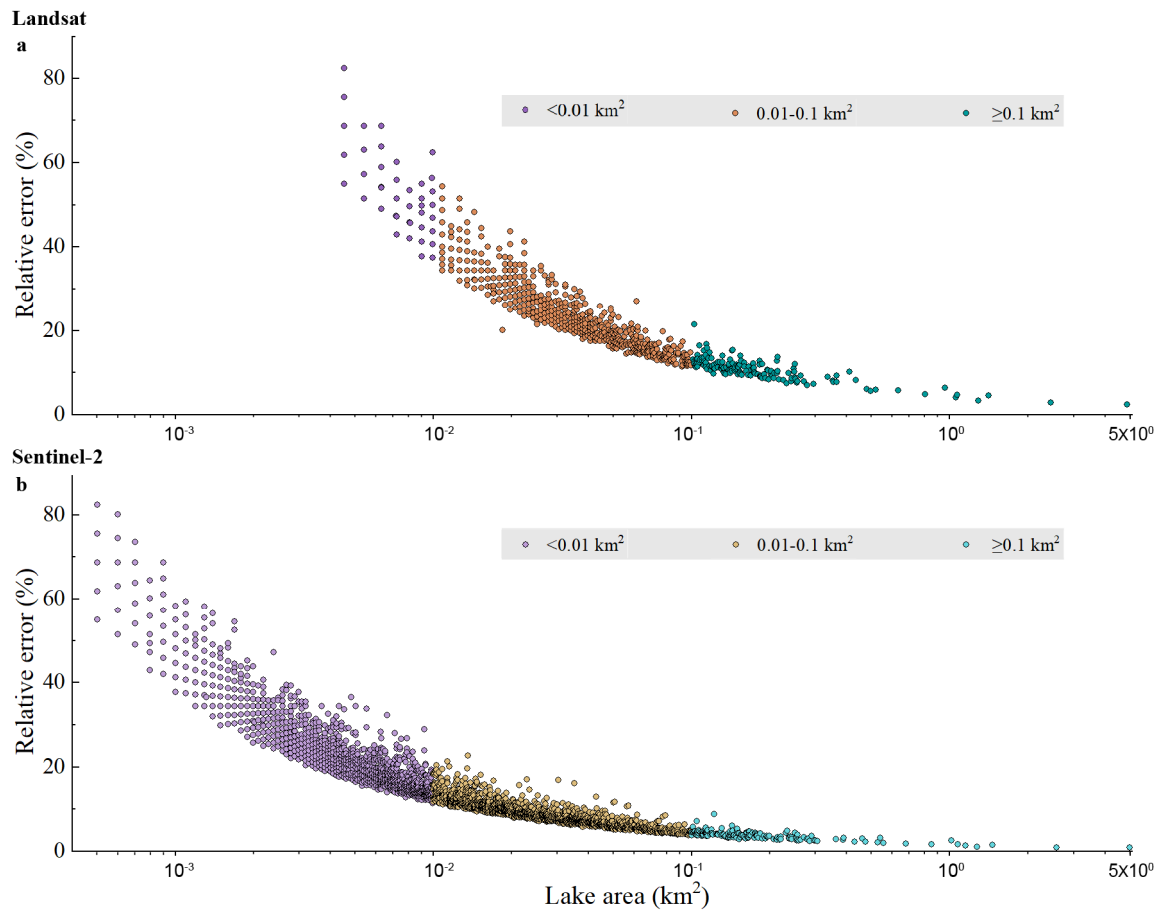
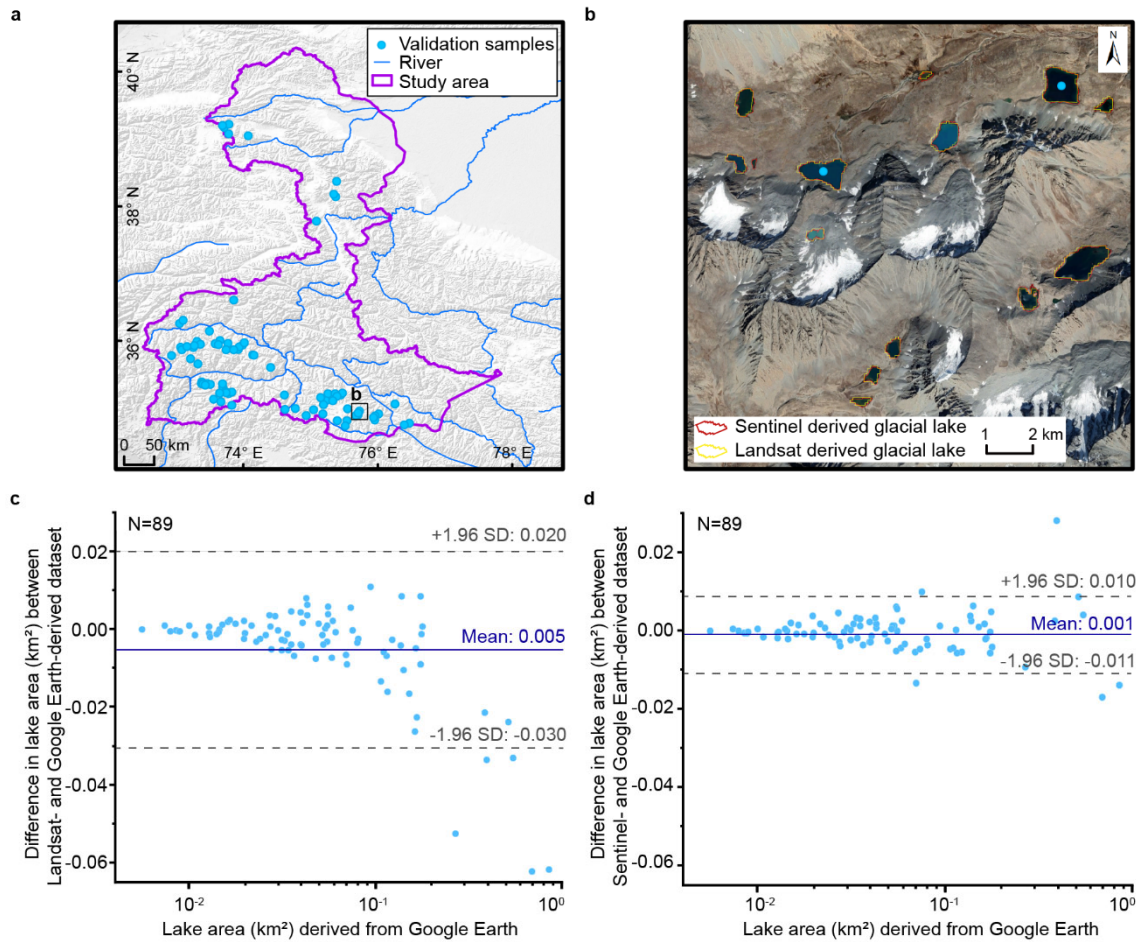



Figure 8. The estimated relative error for glacial lakes of all or specific size ranges in the study area. Error estimation is based on the modified equation and lake data extracted from Landsat (a) and Sentinel-2 images (b).

Our Landsat- and Sentinel-derived glacial lake dataset match well lake boundaries in Google Earth higher resolution images (Figure 9). The mean difference in area is 0.005 km² between Landsat- and Google Earth-derived lakes and 0.001 km² between Sentinel- and Google Earth-derived lakes, and major validation samples (84/89) are within the confidence interval of 95%, indicating high accuracy in lake mapping (Figure 9c and d). The error of 89 sample lakes is 5.48% in the total area between Landsat- and Google Earth-derived data, and 0.61% for Sentinel- and Google Earth-derived data. The median (±standard deviation) in a discrepancy of the individual lake area is 7.66±4.96 % for Landsat- and Google Earth-derived data, and 4.46±4.62 % for Sentinel- and Google Earth-derived data. Our glacial lake dataset shows satisfactory mapping accuracy, although Sentinel-derived lake data performs more accurately than those from Landsat images. We also validated the sampling of Landsat-derived 89 lakes by the existing Landsat-extracted lake data produced by Wang et al. (2020). A total of 83 lakes are available in Wang's data with a mean difference of 0.005 km² in the lake area (Figure A8). This also shows an improvement in our lake product in contrast to the existing dataset.

474

475 **Figure 9.** Distribution of the validation sample (a), visual comparison of glacial lakes derived from
476 Landsat and Sentinel-2 images overlaying Google Earth imagery (© Google Earth 2019) in a zoomed site
477 (b), and differences between our glacial lake product (mapped from Landsat and Sentinel-2 images) and
478 the validation reference (digitized from Google Earth images) (c and d).

479 6.2 Comparison of Sentinel- and Landsat-derived products

480 Glacial lakes from Landsat and Sentinel-2 images have high consistency in number and area
481 with overlap rates from approximately 86% to 94% for all lakes greater than 0.0045 km^2
482 (Table 4), indicating a good potential for coordinated utility with Landsat archived
483 observation (Figure 10). Lake extents extracted from Landsat and Sentinel images match well
484 for various types and sizes (Figure 10 and Figure 11, Table 4). The best consistency rate
485 reaches 94% for the glacial lakes between 0.1 km^2 and 0.2 km^2. The difference in the area of
486 glacial lakes extracted from Landsat and Sentinel-2 images generally lies within the
487 uncertainty ranges.

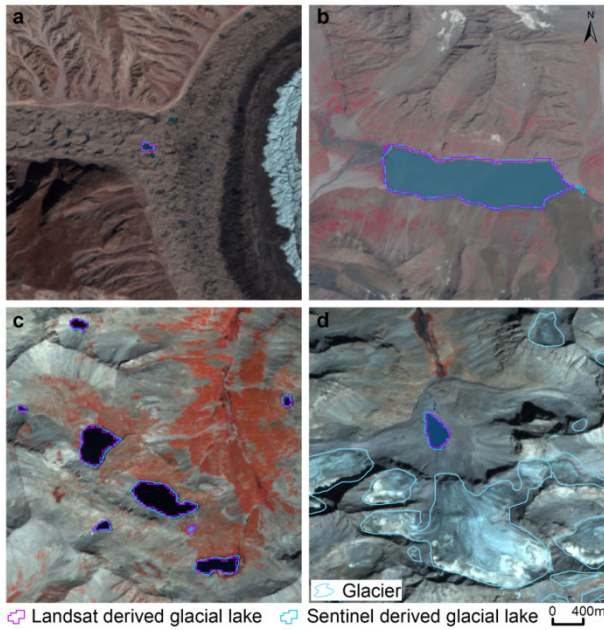
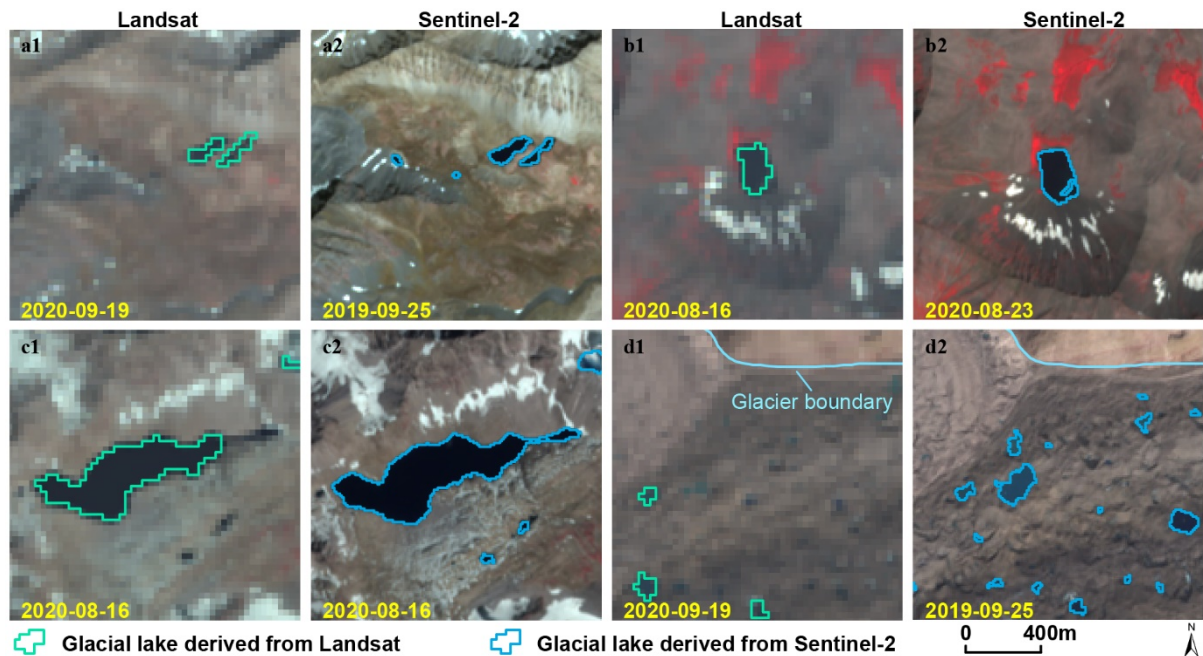



Figure 10. High consistency of lake extents extracted from Landsat and Sentinel-2 images. Lake types shown include supraglacial (a), glacier-fed moraine-dammed (b), unconnected glacial-erosion lake without glacier melt supply (c), and glacier-fed moraine-dammed lakes (d).

The 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-2 (10 m) observations. Due to a finer spatial resolution, Sentinel-2 images can extract more glacial lakes and more accurate extents than those from Landsat images. We set the same 5 pixels as the MMU for both Landsat and Sentinel-2 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-2 images than from Landsat images, causing the greatest discrepancy in number, such as Figure 11. Small lakes such as supraglacial lakes play an important role in analyzing glacier evolution and supraglacial drainage systems (Liu and Mayer, 2015; Miles et al., 2018), implying a potential of our dataset to be applied in studies of glacier-lake evolutions. Meanwhile, Sentinel-2 images can depict boundaries of glacial lakes with lower uncertainty, as some small islands and narrow channels (Figure 11b and c) were mapped from Sentinel-2 imagery that was unable to be detected in Landsat imagery.

In addition to the difference in image resolution, different acquisition dates between Sentinel-2 and Landsat images can also contribute to the discrepancy between those two glacial lake datasets. The total number of supraglacial lakes and ice-dammed glacial lakes are less than 300, but those lakes are controlled by glacier movement and temperature changes (Liu and Mayer, 2015; Miles et al., 2018), which vary faster with time than relatively stable glacial-erosion and moraine-dammed lakes. Acquiring same-day images from the two sensors was 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). Despite our efforts of leveraging all available high-quality images, the overlap of acquisition dates between Landsat and Sentinel-2 images for the same location is relatively low (only 7 scenes of Sentinel-2 images or 112 glacial lakes in 2020) in this study area, and the consequential

517 temporal gaps led to a difference in the number and area of the derived glacial lakes. As
518 exemplified in Figure 11d, the mapped supraglacial lakes in the same location exhibit a
519 considerable discrepancy, which is likely a joint consequence of both sensor difference and
520 glacier lake evolution.
521

522
523 **Figure 11.** The discrepancy of lake extents extracted from Landsat and Sentinel-2 images.
524

525 6.3 Comparison with the previous similar dataset

526 An increasing number of glacier lake datasets have been released over the past years, and
527 most of them were produced from long-term Landsat archives. Regional glacial lake datasets
528 using Sentinel images are scarce. The lack of Sentinel-derived glacial lake data in the study
529 area makes it impossible to compare. Here we selected four available glacial lake datasets to
530 compare with our Landsat-derived dataset at the same MMU and study area.
531   We provide the latest glacial lake dataset (in 2020) and the most long-term 30-m Landsat
532 observation (1990 to 2020) for this study, with a range of critical attributes including two
533 types of classification systems. Within the same study area, our 2020 glacial lakes appear to
534 be closest to the 2018 dataset produced by Wang et al. (2020), with the highest overlap of
535 greater than 91% in count at the minimum mapping unit of 5400 m^2 or 6 pixels (Table 5).
536 Wang's dataset (2020) contains many large landslide-dammed lakes that are excluded in our
537 glacial lake mapping, so their total glacier lake area is greater than ours. The overlapping
538 rates between Wang's glacial lakes (2020) in 1990 and ours are more than 83% in count.
539 However, their results show a distinct increase of glacial lakes in number and area between
540 1990 and 2018 (Wang et al., 2020) whereas our data show a more stable change between
541 1990 and 2020. One possible reason is that manually delineating glacial lakes twice by
542 different operators during Wang's lake mapping (2020) exacerbates the errors of mapping.
543 Another reason is that their data contains landslide-dammed lakes that fluctuate greatly with
544 time and expanded recently. One example is Attabad Lake (Located at 36°18'22.33"N,

545   74°49'34.36"E).

546

547   **Table 5.** Comparison between our Landsat-based mapping and other third-party Landsat-based glacial lake
548   datasets in the study area.

Baseline year (them/us)	Method (them/us)	MMU m^2 (pixels)	Count (them)	Count (us)	Ratio (%)	Reference
1990/1990	Manual/Semi-automated	5400 (6)	1720	2069	83.13	Wang et al., 2020
1990/1990	Automated/Semi-automated	50000 (55)	145	363	39.94	Shugar et al., 2020
1990/1990	Manual/Semi-automated	4500 (5)*	622	2154	28.88	Zhang et al., 2015
2000/2000	Manual/Semi-automated	4500 (5)*	724	2184	33.15	Zhang et al., 2015
2000/2000	Automated/Semi-automated	50000 (55)	155	361	42.94	Shugar et al., 2020
2008/2000	Automated & Manual/Semi-automated	8100 (9)	1067	1800	59.28	Chen et al., 2021
2015/2020	Automated/Semi-automated	50000 (55)	148	364	40.66	Shugar et al., 2020
2017/2020	Automated & Manual/Semi-automated	8100 (9)	1063	1813	58.63	Chen et al., 2021
2018/2020	Manual/Semi-automated	5400 (6)	1956	2149	91.02	Wang et al., 2020

549   Note: MMU represents the minimum mapping unit that is possible to enable a valid comparison between our product and each
550   of the third-party datasets. * The MMU in the dataset of Zhang et al. (2015) is 3 pixels, finer than 5 pixels in our product, so
551   an MMU threshold of 5 pixels was used for this comparison.

552

553   The second highest overlapping rate is approximate 59% for 2008 and 58% for 2017 in
554   count comparing with Chen's data (Chen et al., 2021). Similarly, the overlapping rate between
555   Shugar's dataset (2020) and ours is lower than 43% in count at the minimum mapping unit of
556   50000 m^2. The dataset from Zhang et al. (2015) shows fewer glacial lakes in 1990 and 2000 at
557   the same MMU of 5 pixels. Our product has more lakes than each of the other 4 products at 9
558   time periods. By inspecting their dataset, we attributed this anomalous discrepancy to a range
559   of glacial lakes that were missing due to a lack of thorough cross-check quality assurance
560   during their lake mapping over a larger study area. And those more glacial lakes show an
561   improvement of our product in contrast to the previous similar datasets. Our Landsat-derived
562   glacial lake dataset has been visually cross-checked over three time periods after the step of
563   threshold-based semi-automated lake mapping and has also been visually validated by
564   Sentinel-derived glacial lakes. Through this series of quality assurance, we aim at delivering
565   one of the most reliable multi-decadal glacial lake products for this study area.
566   Other factors, such as image quality and acquisition dates, mapping methods, and quality
567   assurance workflow, might also lead to discrepancies between the glacial lake datasets. Despite
568   such discrepancies, an increasing number of publically-shared datasets benefit potential users
569   to select the most suitable one for their objectives. Herein, we provide an up-to-date glacial
570   lake dataset derived from both Landsat and Sentinel-2 observations, which further increased
571   the availability of glacial lake dataset for water resource and GLOFs risk assessment, predicting
572   glacier-lake evolutions (Carrivick et al., 2020) in the context of climate change.

573

574   6.4 Limitation and updating plan

575   We would like to acknowledge several limitations of our glacier lake dataset, largely due to
576   the availability of high-quality satellite images in the study area and inadequate field survey
577   data (Wang et al., 2020; Chen et al., 2021). First, it is unlikely to collect enough good-quality
578   images within one calendar year for the entire study area due to the high possibility of cloud

or snow cover. Even though the capacity of repeat observations for Landsat-8 OLI and Sentinel-2 increased (Roy et al., 2014; Williamson et al., 2018; Wulder et al., 2019; Paul et al., 2020), the 2020 glacial lake dataset has to employ images acquired in adjacent years besides 2020. Most images used from Landsat and Sentinel-2 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) for each period were ignored. Second, field investigation data are limited due to the low accessibility of the 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 unavoidable. 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 surveys prohibited a thorough validation of the glacial lake types. Third, the rigorous quality assurance and cross-check after semi-automated lake mapping assures the quality of our lake dataset but are still time and cost-prohibitive. State-of-the-art mapping methods, such as deep learning method (Wu et al., 2020), Google Earth Engine cloud-computing (Chen et al., 2021), and synergy of SAR and optical images (Wangchuk and Bolch, 2020; How et al., 2021), would be used in the future to balance product accuracy and time cost.

The glacial lake dataset will be updated using newly collected Landsat and Sentinel images at a five-year interval or modified according to user feedback. The updated glacial lake dataset will continue to be released freely and publicly on the Mountain Science Data Center sharing platform.

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.

8 Conclusions

Glacial lake inventories of the entire China-Pakistan Economic Corridor in 2020 were provided based on Landsat and Sentinel-2 images using a threshold-based semi-automated mapping method. Both Landsat and Sentinel-2 derived glacial lake dataset show similar characteristics in spatial distribution and the statistics of count and area. By contrast, the 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 a longer time scale due to its long-term archived observations at a

620   consistent 30 m spatial resolution starting from the late 1980s.
621   Glacial lakes in the study area remain relatively stable with a slight increase in number and
622   area between 1990 and 2020 according to Landsat observations. Our dataset reveals that 2154
623   glacial lakes in 1990 covering 85.1 ± 14.66 km^2 increased to 2234 lakes with a total area of
624   86.31 ± 14.98 km^2. The same mapping method and rigorous workflow of quality assurance
625   and quality control used in this study reduced the error in multi-temporal changes of glacial
626   lakes.
627   Hanshaw's error estimation method for pixel-based lake mapping was improved by
628   removing repeatedly calculated edge pixels that vary with lake shape. Therefore, the newly
629   proposed method reduces the estimated value of uncertainty from satellite observations. The
630   average relative error is ±17.36% for the Landsat-derived dataset and ±8.15% for the product
631   from Sentinel-2.
632   Our glacial lake dataset contains a range of critical parameters that maximize their
633   potential utility for water resource and GLOFs risk evaluation, cryosphere-hydrological, and
634   glacier-lake evolution projection. The dual classification systems of glacial lake types were
635   developed and are very likely to attract broader researchers and scientists to use our datasets.
636   In comparison with other existing glacial lake datasets, our products were created through a
637   thorough consideration of lake types, cross-checks, and rigorous quality assurance, and will
638   be updated and released continuously in the Mountain Science Data Center. As such, we
639   expect that our glacial lake dataset will have significant value to cryospheric-hydrology
640   research, the assessment of water resources, and glacier-related hazards in the CPEC.
641
642   **Appendix.** The appendix related to this article is available online.
643

644   **Author contributions.** ML and YN conceived the study, ML, YN and XD performed data
645   processing and analysis of the glacial lake inventory data, JW contributed to tool
646   development and mapping methods, and ML and YN wrote the manuscript. All authors
647   reviewed and edited the manuscript before submission.
648
649   **Competing interests.** The authors declare no conflict of interest.
650
651   **Acknowledgments.**
652   We are grateful to the chief editor (ice) Kenneth Mankoff and three anonymous referees for
653   their constructive comments that greatly help us to improve this manuscript. This study was
654   supported by the second Tibetan Plateau Scientific Expedition and Research Program (grant
655   2019QZKK0603), the National Natural Science Foundation of China (Grant Nos. 42171086,
656   41971153), the International Science & Technology Cooperation Program of China (No.
657   2018YFE0100100), the Chinese Academy of Sciences "Light of West China" and Natural
658   Sciences and Engineering Research Council of Canada (Grant No. DG-2020-04207).

659
660

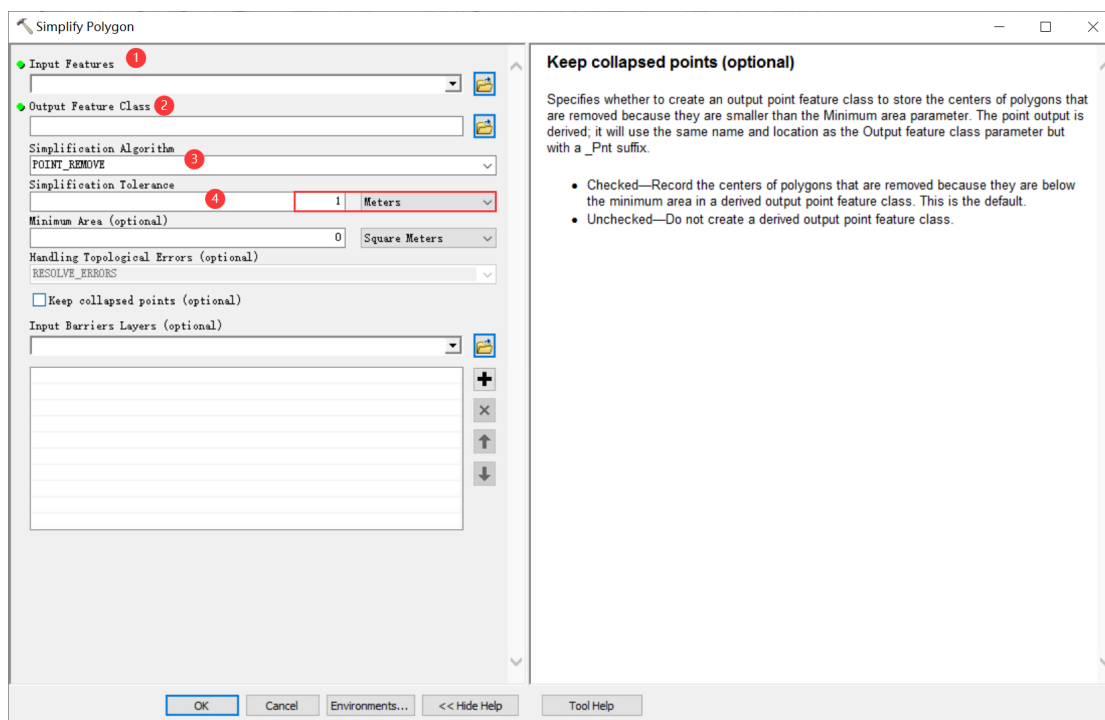
661 **Appendix**

662 **Tutorial for Improved Uncertainty Estimating Method**

663

664 Hanshaw's equation was originally proposed for pixelated polygons (such as a polygon
665 directly extracted from a remote sensing image), and performed more robustly than manually
666 digitized polygons (where vertices do not necessarily follow the pixel edges). Our improved
667 method also performs better for pixelated polygons. This tutorial is dedicated to helping
668 implement our improved uncertainty estimation method.

669

670 **The procedure of uncertainty estimating method (using ArcGIS** (© ESRI) **for example)**
671 1. Removing redundant nodes (optional)
672 We found that a small proportion (~1%) of the pixelated lake polygons (directly extracted
673 from satellite images) have redundant nodes, which affects the value of inner nodes. If no
674 redundant nodes exist, this step can be skipped. Or, we recommend using the "Simplify
675 Polygon" tool in ArcGIS to remove those nodes (Figure A1).
676  In the Simplify Polygon panel
677 • Input your dataset.
678 • Set the output path and output file name.
679 • Choose the simplification algorithm. We recommended "POINT_REMOVE".
680 • Set the tolerance of the simplification algorithm. In this step, we need to ensure that the
681  polygon boundaries remain unchanged after deleting redundant nodes. Generally, a
682  tolerance of 1 meter will suffice, or you can adjust the threshold until your satisfaction.

683

684 **Figure A1.** Input and option for Simplify Polygon in ArcGIS.

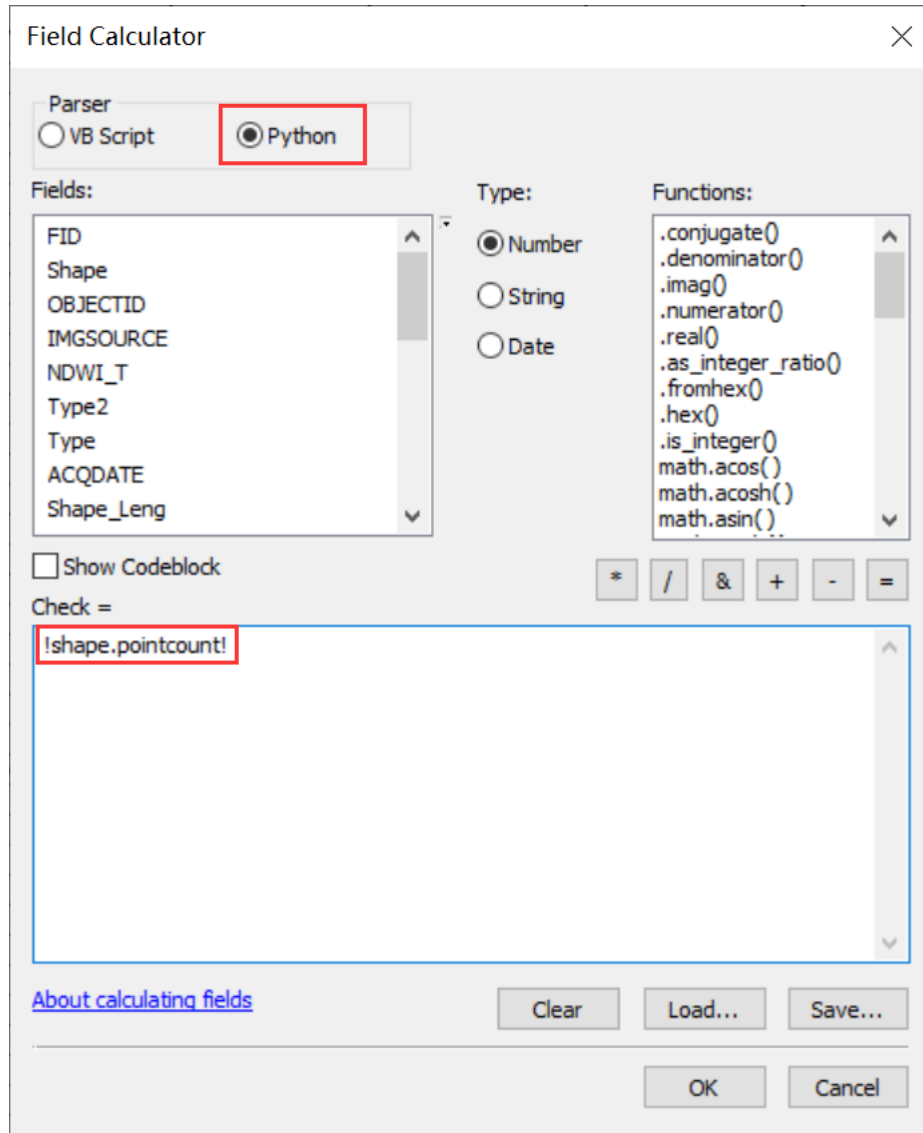
685

686   2. Calculating the total number of nodes using ArcGIS (Figure A2):

687   •   Add a new field in the attribute table of the dataset.

688   •   Open Field Calculator.

689   •   Switch the parser to python-mode, and enter the following code "!shape.pointcount!" in

690   the blue box to calculate the total number of nodes for each glacial lake boundary.

691

692   **Figure A2.** Total node calculation in ArcGIS.

693

694   3. Calculating the number of inner nodes:

695

696   For polygons without islands (Figure A3), use equation 5. An inner node is a polygon vertex

697   where the interior angle surrounding it is greater than 180 degrees. An outer node is the

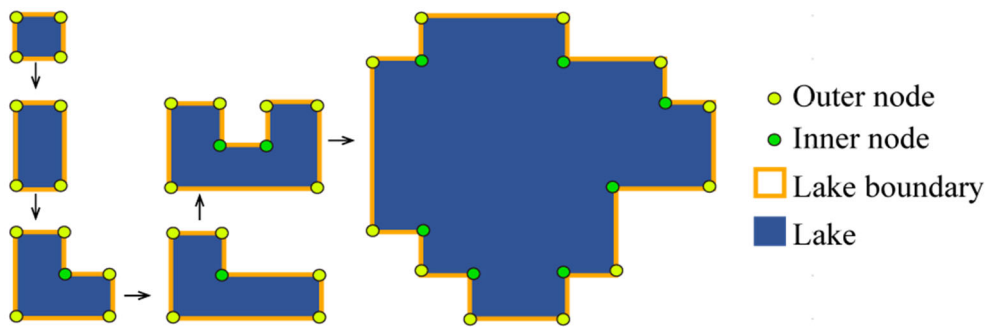
698   opposite of the inner node, where the interior angle is less than 180 degrees. We found that

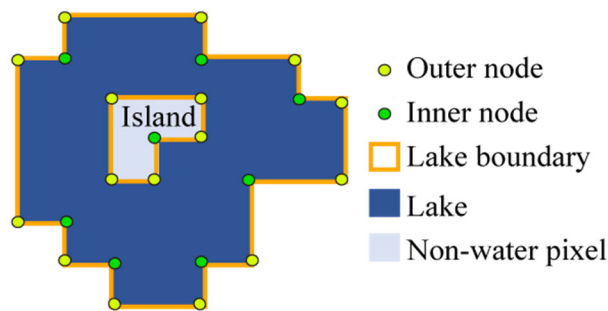
699   the outer nodes are usually four more than the inner nodes in our glacial lake dataset. The

700   total nodes in ArcGIS contain one overlapping node to close the polygon, meaning the

701   endpoint is also the start point. This extra count was deleted from the calculation (equation

702   5).

703

704
705
Figure A3. Sketch of outer and inner nodes of various glacial lakes without island.
706

707   For polygons with islands (Figure A4) use equation 6.
708

709
710   **Figure A4.** Sketch of outer and inner nodes for a glacial lake with an island.
711

712   We further specify the steps below to help implement equation 6.
713

714   Sept 1: detect the number of islands within each polygon.
715   ● Convert the initial lake polygon to a polyline using the "Feature To Line" tool (Figure
716   A5).

717

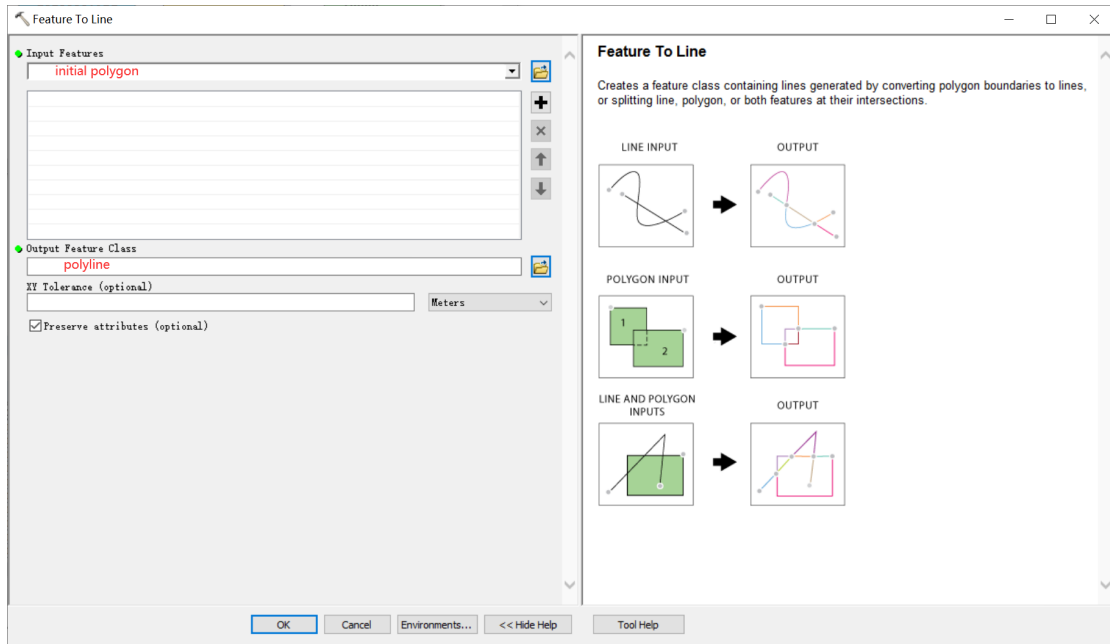
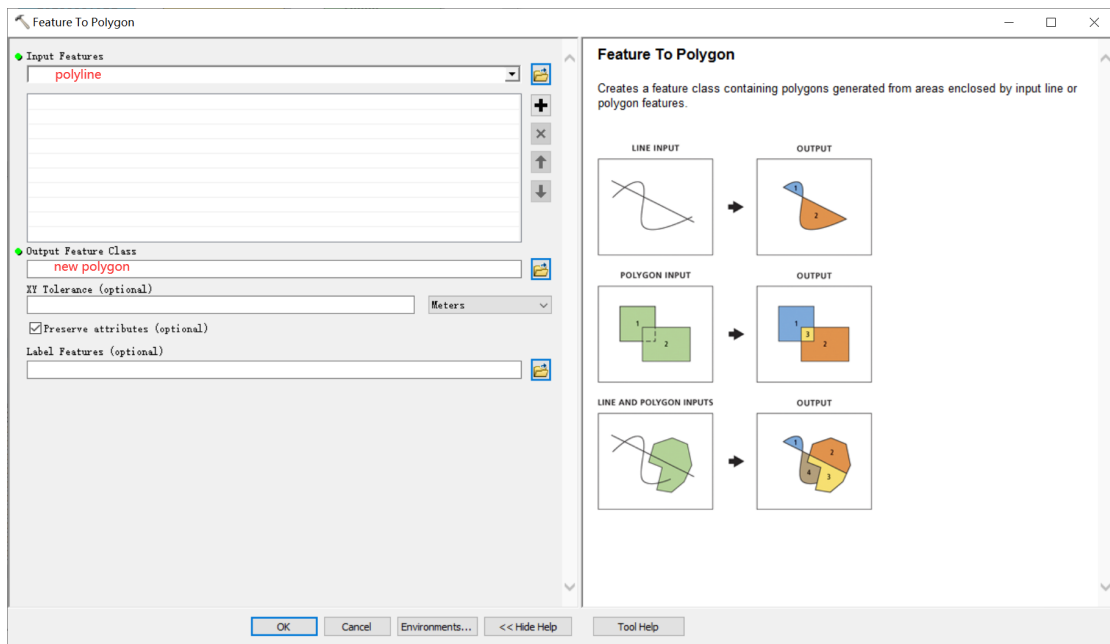



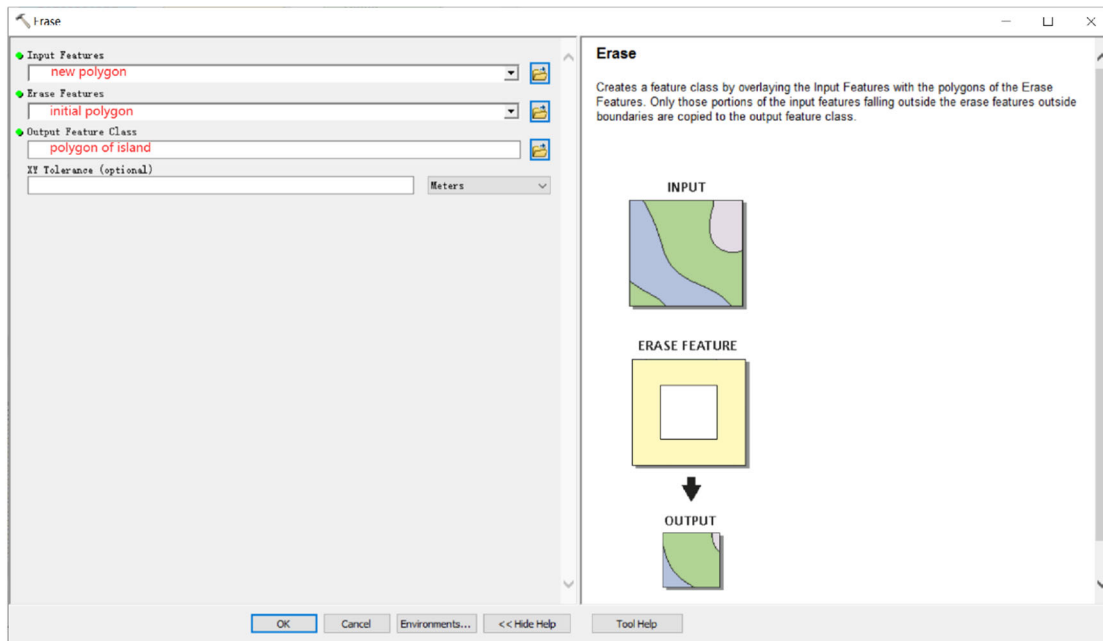
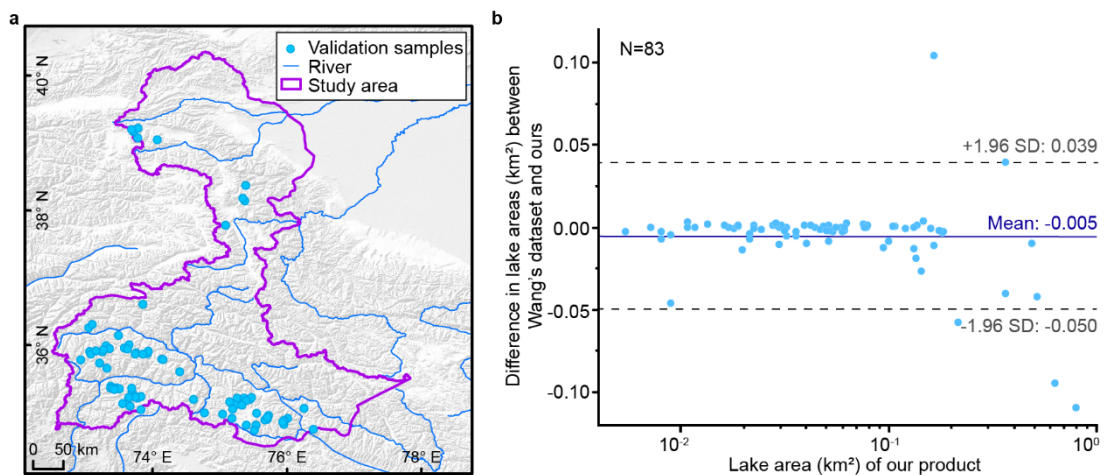
Figure A5. Feature To Line tool in ArcGIS

719

720 ● Convert the polyline to generate a new polygon (Figure A6).

721

722 **Figure A6.** Feature To Polygon tool in ArcGIS

723

724 ● Erase the new polygon by the initial polygon, which outputs the islands. Then we can
725 count how many islands there are in each lake (Figure A7).

726
727 **Figure A7.** Erase tool in ArcGIS.
728
729 Step 2: calculate the number of inner nodes for each polygon with an island or islands using
730 equation 6.
731
732 4. Calculating the uncertainty of lake mapping using equation 4.
733
734

735
736 **Figure A8.** Distribution of validation samples (a) and comparison of glacial lakes (b) derived from our
737 Landsat product in 2020 and Wang's lake data in 2018.
738

739 # References:

740 [1]. Lesi, M., et al., Landsat and Sentinel-derived glacial lake dataset in the China-Pakistan Economic Corridor
741 from 1990 to 2020. Mountain Science Data Center, 2022.
742 [2]. Wang, X., S. Liu and J. Zhang, A new look at roles of the cryosphere in sustainable development. Advances

743 in Climate Change Research, 2019. 10(2): p. 124-131.

744 [3]. Viviroli, D., et al., Increasing dependence of lowland populations on mountain water resources. Nature
745 Sustainability, 2020. 3(11): p. 917-928.

746 [4]. Pritchard, H.D., Asia's shrinking glaciers protect large populations from drought stress. Nature, 2019.
747 569(7758): p. 649-654.

748 [5]. Nie, Y., et al., Glacial change and hydrological implications in the Himalaya and Karakoram. Nature
749 Reviews Earth & Environment, 2021. 2(2): p. 91-106.

750 [6]. Brun, F., et al., A spatially resolved estimate of High Mountain Asia glacier mass balances from 2000 to
751 2016. Nature Geoscience, 2017. 10(9): p. 668-673.

752 [7]. Shean, D.E., et al., A Systematic, Regional Assessment of High Mountain Asia Glacier Mass Balance.
753 Frontiers in Earth Science, 2020. 7: p. 363.

754 [8]. Bhattacharya, A., et al., High Mountain Asian glacier response to climate revealed by multi-temporal satellite
755 observations since the 1960s. Nature Communications, 2021. 12(1): p. 4133.

756 [9]. Maurer, J.M., et al., Acceleration of ice loss across the Himalayas over the past 40 years. Science Advances,
757 2019. 5(6): p. eaav7266.

758 [10]. Huss, M. and R. Hock, Global-scale hydrological response to future glacier mass loss. Nature Climate
759 Change, 2018. 8(2): p. 135-140.

760 [11]. Carrivick, J.L. and F.S. Tweed, A global assessment of the societal impacts of glacier outburst floods. Global
761 and Planetary Change, 2016. 144: p. 1-16.

762 [12]. Nie, Y., et al., An inventory of historical glacial lake outburst floods in the Himalayas based on remote
763 sensing observations and geomorphological analysis. Geomorphology, 2018. 308: p. 91-106.

764 [13]. Zheng, G., et al., Increasing risk of glacial lake outburst floods from future Third Pole deglaciation. Nature
765 Climate Change, 2021. 11(5): p. 411-417.

766 [14]. Rounce, D.R., R. Hock and D.E. Shean, Glacier Mass Change in High Mountain Asia Through 2100 Using
767 the Open-Source Python Glacier Evolution Model (PyGEM). Frontiers in Earth Science, 2020. 7: p. 331.

768 [15]. Shugar, D.H., et al., A massive rock and ice avalanche caused the 2021 disaster at Chamoli, Indian Himalaya.
769 Science, 2021. 373(6552): p. 300-306.

770 [16]. Shugar, D.H., et al., Rapid worldwide growth of glacial lakes since 1990. Nature Climate Change, 2020.
771 10(10): p. 939-945.

772 [17]. Immerzeel, W.W., et al., Importance and vulnerability of the world's water towers. Nature, 2020.
773 577(7790): p. 364-369.

774 [18]. Carrivick, J.L., et al., Ice-Marginal Proglacial Lakes Across Greenland: Present Status and a Possible Future.
775 Geophysical Research Letters, 2022. 49(12): p. e2022GL099276.

776 [19]. Li, Z., X. Deng and Y. Zhang, Evaluation and convergence analysis of socio-economic vulnerability to
777 natural hazards of Belt and Road Initiative countries. Journal of Cleaner Production, 2021. 282(125406): p.
778 125406.

779 [20]. Battamo, A.Y., et al., Mapping socio-ecological resilience along the seven economic corridors of the Belt
780 and Road Initiative. Journal of Cleaner Production, 2021. 309(127341): p. 127341.

781 [21].Hewitt, K., Glaciers of the Karakoram Himalaya:Glacial Environments, Processes, Hazards and Resources.
782 2014, Dordrecht: Springer.

783 [22]. Bhambri, R., et al., Ice-dams, outburst floods, and movement heterogeneity of glaciers, Karakoram. Global
784 and Planetary Change, 2019. 180(9): p. 100-116.

785 [23]. Wang, X., et al., Glacial lake inventory of high-mountain Asia in 1990 and 2018 derived from Landsat
786 images. Earth System Science Data, 2020. 12(3): p. 2169-2182.

787 [24]. Nie, Y., et al., A regional-scale assessment of Himalayan glacial lake changes using satellite observations
788 from 1990 to 2015. Remote Sensing of Environment, 2017. 189(2): p. 1-13.

789 [25]. Brun, F., et al., Heterogeneous Influence of Glacier Morphology on the Mass Balance Variability in High
790 Mountain Asia. Journal of Geophysical Research: Earth Surface, 2019. 124(6): p. 1331-1345.

791 [26]. Liu, Q., et al., Interannual flow dynamics driven by frontal retreat of a lake-terminating glacier in the Chinese
792 Central Himalaya. Earth and Planetary Science Letters, 2020. 546: p. 116450.

793 [27]. Carrivick, J.L., et al., Toward Numerical Modeling of Interactions Between Ice-Marginal Proglacial Lakes
794 and Glaciers. Frontiers in Earth Science, 2020. 8.

795 [28]. Huggel, C., et al., Remote sensing based assessment of hazards from glacier lake outbursts: a case study in
796 the Swiss Alps. Canadian Geotechnical Journal, 2002. 39(2): p. 316-330.

797 [29]. Quincey, D.J., et al., Early recognition of glacial lake hazards in the Himalaya using remote sensing datasets.
798 Global and Planetary Change, 2007. 56(1-2): p. 137-152.

799 [30]. Zhang, G., et al., An inventory of glacial lakes in the Third Pole region and their changes in response to
800 global warming. Global and Planetary Change, 2015. 131: p. 148-157.

801 [31]. Gardelle, J., Y. Arnaud and E. Berthier, Contrasted evolution of glacial lakes along the Hindu Kush Himalaya
802 mountain range between 1990 and 2009. Global and Planetary Change, 2011. 75(1-2): p. 47-55.

803 [32]. Chen, F., et al., Annual 30 m dataset for glacial lakes in High Mountain Asia from 2008 to 2017. Earth
804 System Science Data, 2021. 13(2): p. 741-766.

805 [33]. Wang, X., et al., Changes of glacial lakes and implications in Tian Shan, Central Asia, based on remote
806 sensing data from 1990 to 2010. Environmental research letters, 2013. 8(4): p. 44052.

807 [34]. Rick, B., et al., Dam type and lake location characterize ice-marginal lake area change in Alaska and
808 NW Canada between 1984 and 2019. The Cryosphere, 2022. 16(1): p. 297-314.

809 [35]. How, P., et al., Greenland-wide inventory of ice marginal lakes using a multi-method approach. Scientific
810 Reports, 2021. 11(1): p. 4481.

811 [36]. Ashraf, A., R. Naz and M.B. Iqbal, Altitudinal dynamics of glacial lakes under changing climate in the Hindu
812 Kush, Karakoram, and Himalaya ranges. Geomorphology, 2017. 283: p. 72-79.

813 [37]. Roy, D.P., et al., Landsat-8: Science and product vision for terrestrial global change research. Remote
814 Sensing of Environment, 2014. 145: p. 154-172.

815 [38]. Williamson, A.G., et al., Dual-satellite (Sentinel-2 and Landsat 8) remote sensing of supraglacial lakes in
816 Greenland. The Cryosphere, 2018. 12(9): p. 3045-3065.

817 [39]. Paul, F., et al., Glacier shrinkage in the Alps continues unabated as revealed by a new glacier inventory from
818 Sentinel-2. Earth System Science Data, 2020. 12(3): p. 1805-1821.

819 [40]. Zhang, M., F. Chen and B. Tian, An automated method for glacial lake mapping in High Mountain Asia
820 using Landsat 8 imagery. Journal of Mountain Science, 2018. 15(1): p. 13-24.

821 [41]. Sheng, Y., et al., Representative lake water extent mapping at continental scales using multi-temporal
822 Landsat-8 imagery. Remote Sensing of Environment, 2016. 185: p. 129-141.

823 [42]. Wang, J., Y. Sheng and Y. Wada, Little impact of the Three Gorges Dam on recent decadal lake decline
824 across China's Yangtze Plain. Water Resources Research, 2017. 53(5): p. 3854-3877.

825 [43]. Wang, J., et al., Recent global decline in endorheic basin water storages. Nature Geoscience, 2018. 11(12):
826 p. 926-932.

827 [44]. Wangchuk, S. and T. Bolch, Mapping of glacial lakes using Sentinel-1 and Sentinel-2 data and a random
828 forest classifier: Strengths and challenges. Science of Remote Sensing, 2020. 2: p. 100008.

829 [45]. Zhao, W., et al., Lake area monitoring based on land surface temperature in the Tibetan Plateau from 2000
830 to 2018. ENVIRONMENTAL RESEARCH LETTERS, 2020. 15(0840338).

831 [46]. Yao, X., et al., Definition and classification system of glacial lake for inventory and hazards study. Journal

832 of Geographical Sciences, 2018. 28(2): p. 193-205.

833 [47]. Carrivick, J.L. and D.J. Quincey, Progressive increase in number and volume of ice-marginal lakes on the

834 western margin of the Greenland Ice Sheet. Global and Planetary Change, 2014. 116: p. 156-163.

835 [48]. Carrivick, J.L. and F.S. Tweed, Proglacial lakes: character, behaviour and geological importance. Quaternary

836 Science Reviews, 2013. 78: p. 34-52.

837 [49]. Li, D., D. Shangguan and M.N. Anjum, Glacial Lake Inventory Derived from Landsat 8 OLI in 2016–2018

838 in China–Pakistan Economic Corridor. ISPRS international journal of geo-information, 2020. 9(5): p. 294.

839 [50]. Emmer, A. and V. Cuřín, Can a dam type of an alpine lake be derived from lake geometry? A negative result.

840 Journal of Mountain Science, 2021. 18(3): p. 614-621.

841 [51]. RGI Consortium, Randolph Glacier Inventory – A Dataset of Global Glacier Outlines: Version 6.0:

842 Technical Report. 2017: Global Land Ice Measurements from Space, Colorado, USA.

843 [52]. Rose, A., et al., LandScan Global 2020. 2021, Oak Ridge National Laboratory: Oak Ridge, TN.

844 [53]. Yao, C., et al., Temporal and Spatial Changes of Glacial Lakes in the China-Pakistan Economic Corridor

845 from 1990 to 2018. Journal of Glaciology and Geocryology, 2020. 42(01): p. 33-42.

846 [54]. Ullah, S., et al., Observed changes in maximum and minimum temperatures over China- Pakistan economic

847 corridor during 1980–2016. Atmospheric Research, 2019. 216: p. 37-51.

848 [55]. Kääb, A., et al., Contrasting patterns of early twenty-first-century glacier mass change in the Himalayas.

849 Nature, 2012. 488(7412): p. 495-498.

850 [56]. Yao, T., et al., Different glacier status with atmospheric circulations in Tibetan Plateau and surroundings.

851 NATURE CLIMATE CHANGE, 2012. 2(9): p. 663-667.

852 [57]. Hugonnet, R., et al., Accelerated global glacier mass loss in the early twenty-first century. Nature, 2021.

853 592(7856): p. 726-731.

854 [58]. Hewitt, K., The Karakoram Anomaly? Glacier Expansion and the 'Elevation Effect,' Karakoram

855 Himalaya. Mountain Research and Development, 2005. 25(4): p. 332-340.

856 [59]. Bolch, T., et al., Brief communication: Glaciers in the Hunza catchment (Karakoram) have been nearly in

857 balance since the 1970s. The Cryosphere, 2017. 11(1): p. 531-539.

858 [60]. Azam, M.F., et al., Glaciohydrology of the Himalaya-Karakoram. Science, 2021. 373(6557): p. eabf3668.

859 [61]. Wulder, M.A., et al., Current status of Landsat program, science, and applications. Remote Sensing of

860 Environment, 2019. 225: p. 127-147.

861 [62]. Nie, Y., et al., Glacial change in the vicinity of Mt. Qomolangma (Everest), central high Himalayas since

862 1976. Journal of Geographical Sciences, 2010. 20(5): p. 667-686.

863 [63]. Jiang, S., et al., Glacier Change, Supraglacial Debris Expansion and Glacial Lake Evolution in the Gyirong

864 River Basin, Central Himalayas, between 1988 and 2015. Remote Sensing, 2018. 10(7): p. 986.

865 [64]. Pfeffer, W.T., et al., The Randolph Glacier Inventory: a globally complete inventory of glaciers. Journal of

866 Glaciology, 2014. 60(221): p. 537-552.

867 [65]. Sakai, A., Brief communication: Updated GAMDAM glacier inventory over high-mountain Asia. The

868 Cryosphere, 2019. 13(7): p. 2043-2049.

869 [66]. Jarvis, A., et al., Hole-filled seamless SRTM data V4. 2008. 2008, International Centre for Tropical

870 Agriculture (CIAT): available from http://srtm.csi.cgiar.org.

871 [67]. Farr, T.G., et al., The Shuttle Radar Topography Mission. Reviews of Geophysics, 2007. 45(2): p. RG2004.

872 [68]. Rabus, B., et al., The shuttle radar topography mission—a new class of digital elevation models acquired by

873 spaceborne radar. ISPRS Journal of Photogrammetry and Remote Sensing, 2003. 57(4): p. 241-262.

874 [69]. Post, A. and L.R. Mayo, Glacier dammed lakes and outburst floods in Alaska: U.S. Geological Survey

875   Hydrologic Investigations Atlas 455. 1971, U.S. Geological Survey. p. 1-10.

876   [70]. Martín, C.N.S., et al., Proglacial landform assemblage in a rapidly retreating cirque glacier due to

877   temperature increase since 1970, Fuegian Andes, Argentina. Geomorphology, 2021. 390: p. 107861.

878   [71].Westoby, M.J., et al., Modelling outburst floods from moraine-dammed glacial lakes. Earth-Science Reviews,

879   2014. 134: p. 137-159.

880   [72]. Chen, X., et al., Dam-break risk analysis of the Attabad landslide dam in Pakistan and emergency

881   countermeasures. Landslides, 2017. 14(2): p. 675-683.

882   [73]. Wang, J., Y. Sheng and T.S.D. Tong, Monitoring decadal lake dynamics across the Yangtze Basin

883   downstream of Three Gorges Dam. Remote Sensing of Environment, 2014. 152(0): p. 251-269.

884   [74]. Nie, Y., et al., Reconstructing the Chongbaxia Tsho glacial lake outburst flood in the Eastern Himalaya:

885   Evolution, process and impacts. Geomorphology, 2020. 370(12): p. 107393.

886   [75]. McFeeters, S.K., The use of the Normalized Difference Water Index (NDWI) in the delineation of open

887   water features. International Journal of Remote Sensing, 1996. 17(7): p. 1425 - 1432.

888   [76]. Cook, S.J. and D.J. Quincey, Estimating the volume of Alpine glacial lakes. Earth Surf. Dynam., 2015. 3(4):

889   p. 559-575.

890   [77]. Hanshaw, M.N. and B. Bookhagen, Glacial areas, lake areas, and snow lines from 1975 to 2012: status of

891   the Cordillera Vilcanota, including the Quelccaya Ice Cap, northern central Andes, Peru. The Cryosphere, 2014.

892   8(2): p. 359-376.

893   [78]. Lyons, E.A., et al., Quantifying sources of error in multitemporal multisensor lake mapping. International

894   Journal of Remote Sensing, 2013. 34(22): p. 7887-7905.

895   [79]. Salerno, F., et al., Glacial lake distribution in the Mount Everest region: Uncertainty of measurement and

896   conditions of formation. GLOBAL AND PLANETARY CHANGE, 2012. 92-93: p. 30-39.

897   [80]. Miles, E.S., et al., Glacial and geomorphic effects of a supraglacial lake drainage and outburst event, Everest

898   region, Nepal Himalaya. The Cryosphere, 2018. 12(12): p. 3891-3905.

899   [81]. Liu, Q. and C. Mayer, Distribution and interannual variability of supraglacial lakes on debris-covered

900   glaciers in the Khan Tengri-Tumor Mountains, Central Asia. Environmental Research Letters, 2015. 10(1): p.

901   014014.

902   [82]. Wu, R., et al., A Deep Learning Method for Mapping Glacial Lakes from the Combined Use of Synthetic-

903   Aperture Radar and Optical Satellite Images. Remote Sensing, 2020. 12(24): p. 4020

904   .

905