Abstract. There is currently no glacial lake inventory data set for the entire High Mountain Asia (HMA) area. The definition and classification of glacial lakes remain controversial, presenting certain obstacles to extensive utilization of glacial lake inventory data. This study integrated glacier inventory data and 426 Landsat TM/ETM+/OLI images, and adopted manual visual interpretation to extract glacial lake boundaries within a 10-km buffer from glacier terminals using ArcGIS and ENVI software, normalized difference water index maps, and Google Earth images. The theoretical and methodological basis for all processing steps including glacial lake definition and classification, lake boundary delineation, and error assessment are discussed comprehensively in the paper. Moreover, detailed information regarding the coding, location, perimeter and area, area error, type, time phase, source image information, and sub-regions of the located lakes is presented. It was established that 26,089 and 28,953 glacial lakes in HMA, with sizes of 0.0054–5.83 km², covered a combined area of 1692.74 ± 231.44 and 1955.94 ± 239.68 km² in 1990 and 2018, respectively. The data set now is available at the National Special Environment and Function of Observation and Research Stations Shared Service Platform (China) at http://dx.doi.org/10.12072/casnw.064.2019.db (Wang et al., 2019a).

1 Introduction

Under the background of climate warming and the consequent widespread mass loss of glaciers in alpine regions, increasing volumes of glacial meltwater are being released. This results in glacial lake expansion and extended areas of low-lying terrain (e.g., depressions and troughs) left behind by
retreating glaciers in which water can accumulate and new glacial lakes can form (Clague and Evans, 2000; Mool et al., 2001; Song et al., 2016). As both a water resource and a source of flash flood/debris flow hazards, glacial lakes participate in several natural processes, e.g., regional energy and water cycles (Slemmons et al., 2013), act as both indicators and containers of environmental information (Wang et al., 2016, 2019b), and drive hillslope erosion and landscape evolution (Cook et al., 2018) in the alpine cryosphere. On the one hand, glacial lakes act as temporary storage for the meltwater resource because a considerable amount of meltwater is retained by glacial lake expansion, e.g., approximately 0.2 % a⁻¹ of the total glacial meltwater was reserved in the glacial lakes from 1990 to 2010 in the Tien Shan Mountains in Central Asia (Wang et al., 2013). On the other hand, given the worldwide expansion in lake area in recent decades, the potential will increase for glacial lakes to develop into glacial lake outburst floods and related debris flows that could threaten downstream residents, infrastructure, and regional ecological and environmental security (Huggel et al., 2002; ICMOD, 2011; Bolch et al., 2012; Haebeli et al., 2016). Thus, glacial lakes perform important roles both in the meltwater cycle and in glacier hazard evolution in the cryosphere.

Following the rapid development of remote sensing technology and computer science, remote sensing imagery acquired by various satellites and sensors has been used widely in glacial lake research. In particular, Landsat imagery has become the most important data source for dynamic investigation of glacial lakes because of its wide coverage, continuous and long-term temporal sequence, and accessibility. Based on remote sensing data, both the distribution and the characteristics of change of glacial lakes in the mountains and watersheds in the High Mountain Asia (HMA) region have been widely reported. For example, multi-source remote sensing imagery has been used to compile glacial lake inventories for regions of the Tibetan Plateau (Zhang et al., 2015), Tien Shan Mountains (Wang et al., 2013), Himalaya (Gardelle et al., 2011; Nie et al., 2017), Hengduan Mountains (Wang et al., 2017), Uzbekistan (Petrov et al., 2017), and Pakistan (Senese et al., 2018). These inventories have proved an important data resource both for recognizing the spatiotemporal characteristics of glacial lakes and for understanding the response of glacial lakes to the effects of climate change in these regions.

Automatic and semi-automatic glacial lake boundary vectorization approaches have been used most widely in regional glacial lake investigations because of their higher efficiency and objectivity in comparison with manual visual vectorization. In such research, water bodies are usually determined based on the characteristics of different remote sensing bands and computer-dependent algorithms, e.g., the normalized difference water index (NDWI), band ratio, support vector machine, decision tree,
spectral transformation, object-oriented classification, global–local iterative scheme, active contour model, and random forest (Gardelle et al., 2011, Huggel et al., 2002, Li et al., 2011; Veh et al., 2018, Zhang et al., 2018). However, manual post-processing is often required to calibrate the uncertainties that could easily be produced by the above approaches. Furthermore, the labour costs associated with rectification of lake boundary errors increase sharply with increasing complexity of study area terrain (Yang et al., 2019). With consideration of the accuracy, efficiency, and time overheads associated with the various vectorization approaches, a manual vectorization approach was adopted for investigation of the glacial lakes on the Tibetan Plateau (Zhang et al., 2015) despite the labour requirements and anticipated additional errors produced by individual subjectivity (Nie et al., 2017; Yang et al., 2019; Song et al., 2014).

Controversies and knowledge gaps remain regarding available glacial lake inventories for different alpine cryosphere regions, which present certain obstacles to extensive utilization of glacial lake inventory data. The main problems relate to regional differences in lake development, inconsistent specifications of lake definition, and the adoption of various approaches to lake interpretation (Yao et al., 2018). There is no existing comprehensive glacial lake inventory for the entire HMA and knowledge regarding the spatiotemporal characteristics of glacial lakes in this region is incomplete. The objectives of this study were to fill this knowledge gap by producing a glacial lake inventory data set for HMA derived from a Landsat imagery series, and to provide fundamental data for water resource evaluation, assessment of glacial lake outburst floods, and glacier hydrology research in the mountain cryosphere region.

2 Study area

The HMA area mainly comprises the Tibetan Plateau and surrounding alpine ranges. The area is divided into 13 sub-regions in version 5 of the Randolph Glacier Inventory (RGI 5.0), i.e., the Himalaya area (Western Himalaya, Central Himalaya, and Eastern Himalaya), Hengduan Mountains, Southern and Eastern Tibet, Inner Tibetan Plateau, Karakoram and Western Kun Lun, Qilian Shan and Eastern Kun Lun, Hindu Kush, Pamir, Alay and Western Tien Shan, Eastern Tien Shan, and Altay and Sayan (Arendt et al., 2014; Pfeffer et al., 2014). The boundaries of the 13 sub-regions and outlines of the glaciers in HMA derived from RGI 5.0 are shown in Figure . This region covers an area that extends approximately 40° in longitude and 27° in latitude, which is characterized by tremendously complex topographic conditions with widespread distribution of mountain glaciers. According to the Climatic Research Unit Time Series v4.02 data set
The air temperature of the different sub-regions in HMA increased at an average annual rate of 0.002–0.054 °C a⁻¹ during 1990–2018 (Fig. 1). The annual rate of change of precipitation in HMA during 1990–2018 varied from −9.9 to 4.2 mm a⁻¹ with a small average rate of increase of 0.3 mm a⁻¹.

The HMA area has the largest surviving glaciers of any region other than Antarctica. As reported in RGI 5.0, there were 97,974 modern glaciers in our study area, covering a total area of approximately 98,768.86 km². Together, these glaciers produced an average meltwater volume of 110–150 km³ a⁻¹, which was the primary source of both lake basin formation and water supply for the development of glacial lakes. Over recent decades, glaciers in most areas of HMA appear to have experienced widespread mass wastage and area shrinkage (Bolch et al., 2012; Yao et al., 2012; Kääb, et al., 2012; Brun, et al., 2017). However, the so-called “Karakoram Anomaly” refers to a region that is a prominent exception, which is characterized by glaciers with stable or positive mass balance (Hewitt, 2005; Gardelle et al., 2013; Kääb, et al., 2015).
Figure 1. Location of sub-regions, rate of change of air temperature (1990–2018), and buffer area within 10 km of glacier terminals for glacial lake inventory of High Mountain Asia.

3 Data source

We developed our glacial lake inventory of HMA based on 426 high-quality images selected from more than 1800 Landsat images with 30-m spatial resolution derived from the websites of the United States Geological Survey (https://www.usgs.gov/) and Geospatial Data Cloud (http://www.gscloud.cn/). To ensure the accuracy of glacial lake boundary extraction, the following criteria were applied to imagery selection. First, the cloud coverage in an image had to be <10%. Second, for areas with no eligible or only low-quality imagery (because of snow or shadows) in the given year, acceptable images from years closest to the given year were chosen as replacements (Fig. 2). Third, images acquired in summer or autumn (July–November), when lake areas were believed near or at their maximal extent, were set as optimal choices to minimize the impact produced by seasonal area changes of the glacial lakes (Fig. 2). Based on the above criteria, 204 and 222 Landsat images were selected to represent 1990 and 2018, respectively, which completely covered the buffer area within 10 km of the glaciers terminals acquired from the Second Chinese Glacier Inventory (http://westdc.westgis.ac.cn) and RGI 5.0 (https://www.glims.org/RGI/rgi50_dl.html). Among the selected images, those acquired during July–November accounted for 91% of the total number of selected images, while those acquired during September–October accounted for 87% of the total number. In addition, a Shuttle Radar Topography Mission digital elevation model with spatial resolution of 1″ (http://imagico.de/map/demsearch.php) was used to derive the elevation of the glacial lakes.
4 Glacial lake inventory methods

4.1 Outline of workflow
The methods and workflow adopted in this study to produce a glacial lake inventory mainly included collation of knowledge and formulation of the specifications of the glacial lake inventory, data pre-processing, manual vectorization of glacial lakes, interactive checking and error controlling, and attribute database assignment (Fig. 3).

1 (1) Collation of available knowledge regarding glacial lake inventories. As much literature as possible relevant to the investigation and recording of glacial lakes was collected. The various definitions and classifications of glacial lakes, as well as the methods adopted previously for glacial lake boundary extraction and assessment of the extent of glacial lake distributions, were summarized and normative rules formulated for the HMA glacial lake inventory, as explained further in Sect. 3.2.

2 (2) Formulation of the specifications of lake identification. First, a working group of four leading experts in the field was founded in 2014 to discuss and formulate the specifications of the glacial lake inventory. Current knowledge regarding identification of lakes from Landsat imagery (e.g., pixel colour, lake shape, and lake background features) and specifications of vectorization (e.g., viewing scale on a computer screen vectorization of mixed pixels) was discussed and unified operating criteria were
compiled to guide the glacial lake inventory operatives. Novice vectorization operatives were trained until their vectorization results met the pre-specifications of the inventory.

(3) Pre-processing of remote sensing data. Pre-processing of the Landsat imagery included false colour compositing and calculation of NDWI maps. The false colour composite images were based on combinations of the operating bands of 7, 5, and 2 or 4, 3, and 2 for Landsat TM/ETM/ETM+ images and 5, 4, and 3 for Landsat OLI images. The preliminary lake extent was extracted automatically from each image over the entire HMA area using the NDWI based on the near infrared band (NIR) and green band (GREEN), which represent the minimum and maximum water reflectance, respectively (McFeeters, 1996; Zhai et al., 2015; Li et al., 2016; Zhang et al., 2018):

\[
\text{NDWI} = \frac{B_{\text{GREEN}} - B_{\text{NIR}}}{B_{\text{GREEN}} + B_{\text{NIR}}}
\]

where \(B_i\) is the spectral band of Landsat imagery. The NDWI maps were calculated for each selected Landsat image using different region-specific thresholds, which typically were in the range −0.10 to 0.20 for lake surfaces. When manual vectorization was performed on a false colour composite image, the NDWI maps of potential glacial lakes were overlaid to assist in glacial lake identification.

(4) Manual vectorization and entering of attribute data. The inventory work was performed during 2014–2019. Seven groups were formed to conduct lake boundary vectorization of the 13 HMA sub-regions. After vectorization of a glacial lake, it was required that manual attribute items (e.g., data source and lake type) be input concurrently.

(5) Interactive checking and accuracy control. First, glacial lakes were discerned via human–computer interaction, i.e., potential glacial lakes were revealed by the NDWI maps or identified visually from the false colour composite images. Second, glacial lake boundary vectorization results were checked interactively by another vectorization operative to eliminate misclassified areas of shadow and ice and to add areas of glacial lakes evidently omitted in the boundary extraction process. This checking process also minimized the subjective judgment errors of the operatives. Third, attribute items such as glacial lake classification, new/disappeared lakes, and separated/coalesced lakes were checked interactively. In this process, Google Earth imagery was used as an important auxiliary reference data source for error examination.
4.2 Illustration of key methods

4.2.1 Definition of glacial lakes

The definition of a glacial lake determines the type of cryosphere water body that will be recorded as a glacial lake. There are multiple definitions of a glacial lake based on different perspectives (Mool et al., 2001; Yao et al., 2018). When glacial lake inventories are undertaken, most emphasize the elementary role of glaciation in the formation of glacial lakes (Clague and Evans, 2000; Qin et al., 2016; Mool et al., 2001). The remarkable difference is whether the period of glaciation or the supply source of glacial lakes is given greatest attention. Some studies that focused on the former proposed that a glacial lake is a natural water body formed by alpine glacier movement since the Last Glacial...
Maximum, i.e., ancient or modern glaciers (Liu et al., 1988; Costa and Schuster, 1988). However, other studies emphasized the relation of glacial lakes to meltwater in glaciated areas (Wang et al., 2013; Weicai et al., 2014; Zhang et al., 2015). The glacial lake inventory data compiled in this study are intended for use both in water source evaluation and in assessment of environmental change in the alpine cryosphere. Thus, lakes related to glaciers or to glaciation in the alpine cryosphere were all recorded as glacial lakes.

Most Quaternary glaciers have disappeared and the remaining relics are incomplete, which makes it difficult to recover a continuous and complete glaciation range in alpine regions. Thus, it is of great importance to ensure the range of glaciation in an alpine region when conducting a glacial lake inventory based on remote sensing data. The most practical approach might be to specify an indicator threshold to define the glaciation extent according to relevant findings of existing glacier relics in a typical region. On the one hand, the glaciation frontier can usually be indicated by a specified lowest elevation threshold, which is generally closely related to the regional climatic context caused by the elevation effect. However, the lowest elevation threshold might vary enormously with respect to different regions because regional climatic settings differ. For instance, the lowest elevations of 1700 m in Austria (Buckel et al., 2018), 2000 m in Pakistan (Senese et al., 2018), 3000 m in Nepal and Bhutan (Mool et al., 2001), and 3500 m in Peru (Hanshaw and Bookhagen, 2014) were used as specified elevation thresholds to record glacial lakes. On the other hand, defining glaciation extent within a specific distance from glacier terminals could be more suitable for the establishment of a glacial lake inventory in relatively large-scale regions with complex regional climate, because the differing climate within large-scale regions can be indicated approximately by the lowest elevation of individual glacier terminals. Some studies adopted distances of 2, 3, or 10 km from modern glacier terminals as thresholds with which to define areas of glacial lakes (Petrov et al., 2017; Veh et al., 2018; Wang et al., 2012, 2013). Distances of 2, 5, 10, and 20 km were considered by Zhang et al. (2015). They found that a distance of 10 km from a modern glacier terminal might be a reasonable guide to glaciation extent and a threshold suitable for a glacial lake inventory of the Tibetan Plateau. This was supported by the finding that the most distant glacierized boundary of the Little Ice Age was up to 10 km from the modern glaciers in the Himalaya area (Wang et al., 2012, Nie et al., 2017). Additionally, to record glacial lakes more precisely, combined distance and elevation thresholds have been used simultaneously to define areas of glacial lakes in special small regions, e.g., lakes at elevations above 1500 m and within 2 km of modern glaciers were recorded as glacial lakes in Uzbekistan (Petrov et al., 2017). In this study, given
the large scale of the HMA region with its complex climatic context and extremely varied terrain, the
data set compiled included glacial lakes within a buffer zone of 10 km from modern glacier terminals,
which covered an area of approximately $1.19 \times 10^6$ km$^2$ according to the Second Glacier Inventory of
China and RGI 5.0 (Fig. 1).

4.2.2 Classification of glacial lakes

In glaciation regions, the characteristics of glacial lakes, which include the phase of lake formation,
lake basin topography, dam material constituents, geometrical relationship with modern glaciers, and
source of water supply (or combinations thereof), have been employed as the basis for glacial lake
classification systems (Huggel et al., 2002; Liu et al., 1988; Mool et al., 2001; Yao et al., 2018). For
instance, based on lake basin topography, lakes in an inventory of the Hindu Kush–Himalaya region
were classified as erosion lakes, valley trough lakes, cirque lakes, blocked lakes, lateral and end
moraine-dammed lakes, and supraglacial lakes (Liu et al., 1988; Mool et al., 2001). Recently, Yao et al.
(2018) presented a reasonably complete classification schema for glacial lake inventory and study of
glacial lake hazards that included six classes and eight sub-classes based mainly on the mechanism of
glacial lake formation, lake basin topography, and the geometrical relationship with modern glaciers.

Generally, it is a little difficult to distinguish glacial lake type in terms of material properties,
topographic features, and phase of lake formation using remote sensing imagery. Moreover, most of the
standards mentioned above were found inapplicable in previous studies of glacial lake classification in
large-scale regions such as HMA because of the lack of sufficient amounts of remote sensing data with
appropriate resolution. In this study, the hydrologic relationship between glacial lakes and modern
glaciers was adopted as a classification criterion because the present data set is intended to provide
fundamental data for water resource evaluation and glacier hazard assessment. Consequently, glacial
lakes were divided into just two types: glacier-fed lakes and non-glacier-fed lakes. This classification
was based on whether the hydrological flow of the modern glacier and topographic features of the lake
basin allowed a lake to receive meltwater from the modern glacier. To achieve reliable classification
results, glacial lakes were distinguished with the assistance of 3D digital terrain imagery from Google
Earth, a Shuttle Radar Topography Mission digital elevation model, and glacier outlines from RGI 5.0.
4.2.3 Extraction of lake boundary

This study adopted automatic glacial lake extraction and manual glacial lake boundary vectorization to determine glacial lake boundaries. In the NDWI-based automatic lake boundary extraction approach, two bands were selected to facilitate a ratio calculation to maximize the difference between water and non-water objects in the remote sensing imagery based on a given threshold. The given threshold was determined subjectively with consideration of how much detailed information of the lake water bodies was captured precisely. The given threshold was varied to account for various factors such as the differences in Landsat sensors (i.e., TM, ETM/ETM+, and OLI), time phase of images, quality of images, and complexity of surface features. To achieve the optimal threshold for lake water body recognition, the candidate threshold was debugged iteratively for each image. In practice, because the area of the glacial lakes was usually small (see next paragraph) and the spectral features of the lake water bodies were varied, the threshold had to be set to allow capture of the greatest number of water body pixels, which consequently resulted in simultaneous acquisition of more non-lake-water-body noise information. It also resulted in more effort in the subsequent manual modification to reduce noise information using methods such as algorithms to eliminate mountain shadows (Gardelle et al., 2011).

Manual visual vectorization distinguishes lake boundaries by identifying the unique texture, colour, and other characteristics of glacial lakes in false colour composite images based on available professional knowledge and accumulated experience in vectorization operations. Even though it was regarded a time-consuming and labour-intensive process, it was also considered an attractive approach because of its consistency, high level of quality control, and reasonably simple operational procedure, given the varied quality of Landsat images for the large-scale HMA region. In this study, the manual visual vectorization process was generally found more suitable in terms of effort and precision for generating a glacial lake inventory data set of the HMA region in comparison with automatic glacial lake extraction. Therefore, manual visual vectorization in conjunction with NDWI maps was the main method adopted to extract glacial lake boundaries to minimize the deficiencies produced by individual subjectivity of the operatives.

The minimum number of pixels used to extract a glacial lake water body was found inconsistent in the available literature. For example, arbitrary threshold areas of 0.0027 km² (three lake water body pixels) (Zhang et al., 2015) and 0.0081 km² (nine lake water body pixels) (Nie et al., 2017) have been used in earlier glacial lake investigations. Moreover, minimum threshold areas of 0.01 km²
(approximately 10 lake water body pixels), 0.02 km$^2$ (approximately 22 lake water body pixels) and 0.1 km$^2$ (~111 lake water body pixels) have also been set to evaluate the level of risk of glacial lake outburst floods in Himalaya and the Tien Shan Mountains (Petrov et al., 2017, Wang et al., 2013; Worni et al., 2013; Bolch et al., 2011; Allen et al., 2019). Theoretically, one pure pixel of a lake water body could be recorded as a glacial lake. However, a glacial lake is generally not represented by one pure pixel unless perfectly aligned with the raster grid; usually, it would be surrounded partly or fully by 1–8 mixed lake water body pixels (Fig. 4a, b). Consequently, manual delineation was required for approximately 1/2, 1/4, or 3/4 of the peripheral mixed pixels surrounding pure lake water body pixels (Fig. 4c). In this study, the minimum glacial lake area recorded was set at 0.0054 km$^2$ (e.g., 3–4 pure lake water body pixels with approximately 12 mixed boundary pixels) because a lake area covering fewer than three pure lake water pixels could possibly have an error of >100% (see Sect. 5).

![Figure 4](https://doi.org/10.5194/essd-2019-212)

4.2.4 Input of attribute items

Eight attribute items were input into the HMA glacial lake inventory: lake coding, location (longitude, latitude, and elevation), perimeter, area, type, area error, time phase, source image
information, and sub-region of located lake. (1) We encoded each glacial lake based on its central location using the same coding format used by the National Snow and Ice Data Centre to encode glaciers. The code can be expressed as “GLNmmmmmmEEnnnnnN”, where m and n represent the results of the longitude and latitude of each glacial lake centroid multiplied by 1000, respectively, GL is the abbreviation of glacial lake, and E and N represent eastings and northing, respectively. (2) The location information of each glacial lake was labelled as the geographic coordinates of the centroid of the shape of each glacial lake, calculated using ArcMap software. The lake elevation was defined as the average elevation of a buffer zone of 30-m radius centred on the glacial lake centroid, which was derived from the Shuttle Radar Topography Mission digital elevation model. (3) The area and perimeter of each lake were calculated using ArcMap based on the unified geography coordinate system of GCS_WGS_1984 and the Asia_North_Albers_Equal_Area_Conic projection system, respectively, to avoid errors caused by projection deformation. (4) The error of lake area was calculated using Eqs. (2) and (3) (Sect. 5). (5) Lake type, which was input manually, was defined as either glacier-fed lake or non-glacier-fed lake (see Sect. 4.2.2). (6) Lake time phase was the acquisition date of the original Landsat image, which was recorded as the month and year for each lake. (7) Source image information referred to the image number of the Landsat images used to extract the glacial lake boundary. (8) The sub-region to which each lake belonged identified the regional location within the HMA area. Each lake was assigned based on shp. file data of the boundaries of the 13 HMA sub-regions, obtained from the National Snow and Ice Data Centre, using the ArcMap spatial analysis tool.

5 Error assessment

The errors associated with glacial lake extraction from remote sensing imagery using manual visual delineation are generally related to components of the quality of the images (e.g., spatiotemporal resolution, cloud coverage, and mountain shadows), experience, operative subjectivity, and the threshold area of the inventory (Gardelle et al., 2011; Hall et al., 2003; Paul et al., 2004; Salerno et al., 2012; Zhang et al., 2015). It has been reported that the area error of glacial lake boundary extraction based on remote sensing images can be approximately ±0.5 pixels depending on the quality of the imagery (Fujita et al., 2009; Salerno et al., 2012). Furthermore, the area error of glacial lake delineation attributable to manual delineation can be assumed to follow a Gaussian distribution (Hanshaw and Bookhagen, 2014). Hence, the theoretical maximum area error of glacial lake boundary extraction is the half-area of the edge pixels.
because pure lake water body pixels are usually surrounded by mixed pixels (Fig. 4b, c). The lake area error within one standard deviation (1σ) can be expressed as follows (Hanshaw and Bookhagen, 2014):

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

$$E = \frac{\text{Error}(1\sigma)}{A} \times 100\%$$

where $P$ is the perimeter of the glacial lake (m), $G$ is the spatial resolution of the remote sensing imagery (30 m in this data set), 0.6872 is the revised coefficient under 1σ (i.e., approximately 69 % of peripheral pixels are subjected to errors), $E$ is the relative error of the glacial lake, and $A$ is the total area of the glacial lake.

The resulting calculated error indicated that the total absolute area error of HMA glacial lakes was approximately ±231.44 and ±259.68 km² and the average relative error was ±13.8 and ±13.3 % in 1990 and 2018, respectively. The relative area errors of each lake varied from 2–85 %, and a significant power exponential relationship was found between the relative area error and the sizes of the glacial lakes ($E = 0.050A^{-0.45}$, $R^2 = 0.96, \sigma < 0.001$) (Fig. 5a). Small-sized lakes (i.e., area ≤ 0.01 km², which accounted for 2 % of the total lake area in HMA) had the largest average relative area error of 44.6 % (Fig. 5b). Medium-sized lakes (i.e., area of 0.01–0.1 km², which accounted for 34 % of the total lake area in HMA) had an average relative area error of 22.1 % (Fig. 5c). Large-sized lakes (i.e., area ≥ 0.1 km², which accounted for 64 % of the total lake area in HMA) had the smallest average relative area error of 4.1 % (Fig. 5d). In summary, smaller glacial lakes in the HMA region had larger relative area errors, and vice versa.

To further verify the accuracy of the manual delineation of glacial lake boundaries, nine lakes located within the HMA region were surveyed using a portable GPS device (Trimble GeoXH6000) during July–August 2018 (Fig. 6). The lakes selected for field survey covered areas of 0.01–2.97 km². The field-based lake boundaries were compared with those obtained via manual delineation (i.e., derived from Landsat OLI imagery acquired during 2018). It was found that the area error (i.e., the percentage difference of the absolute area encircled by the manually delineated lake boundary and that derived by the GPS survey) varied from 5.5–25.5 %. Moreover, it was determined that the average horizontal distance deviation between the two types of boundary varied from 4.5–33.5 m (Table 1). Overall, the horizontal deviations were largely confined to one pixel, and the average accuracy of the delineation of glacial lake boundaries was within ±0.5 pixels (±15 m).
Figure 5. Relationships of relative area error against size of glacial lakes in HMA: (a) relationship for glacial lakes of all sizes and (b)–(d) relationships for glacial lakes of specific size.

Table 1. Horizontal deviations between lake boundaries obtained by manual delineation and field survey using a portable GPS device (Trimble GeoXH6000)

<table>
<thead>
<tr>
<th>Name (labelled in Figure 6)</th>
<th>Lake ID</th>
<th>Lake size (km²)</th>
<th>Horizontal deviations of delineation boundary (m)</th>
<th>Area error(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qiongyong Cuo (A)</td>
<td>GL090225E28890N</td>
<td>0.08</td>
<td>minimum -7.6, maximum 9.1, average 4.5</td>
<td>5.5</td>
</tr>
<tr>
<td>Passu Lake (B)</td>
<td>GL074878E36457N</td>
<td>0.15</td>
<td>minimum -10.9, maximum 12.0, average 6.0</td>
<td>6.8</td>
</tr>
<tr>
<td>Longbas Lake (C)</td>
<td>GL088071E27950N</td>
<td>1.49</td>
<td>minimum -22.7, maximum -8.4, average 12.4</td>
<td>/</td>
</tr>
<tr>
<td>Zongge Cuo (D)</td>
<td>GL087654E28113N</td>
<td>1.48</td>
<td>minimum -26.8, maximum 24.9, average 13.5</td>
<td>6.1</td>
</tr>
<tr>
<td>Unnamed (E)</td>
<td>GL088151E28010N</td>
<td>0.01</td>
<td>minimum -4.7, maximum 4.8, average 3.2</td>
<td>16.3</td>
</tr>
<tr>
<td>Unnamed (F)</td>
<td>GL088257E28011N</td>
<td>0.58</td>
<td>minimum -12.9, maximum 12.4, average 4.6</td>
<td>/</td>
</tr>
<tr>
<td>Unnamed (G)</td>
<td>GL088240E28005N</td>
<td>0.40</td>
<td>minimum -20.8, maximum 15.9, average 7.1</td>
<td>/</td>
</tr>
<tr>
<td>Large Laigu Lake (H)</td>
<td>GL096818E29298N</td>
<td>2.97</td>
<td>minimum -36.7, maximum -6.4, average 15.3</td>
<td>/</td>
</tr>
<tr>
<td>Small laigu Lake (I)</td>
<td>GL096832E29294N</td>
<td>1.02</td>
<td>minimum -32.8, maximum 17.6, average 9.8</td>
<td>/</td>
</tr>
</tbody>
</table>

Note: ‘/’ indicates the sample lake boundary was only partly surveyed using the handheld GPS device.
Figure 6. Glacial lakes in the HMA region surveyed in summer 2018 (the background maps of surveyed lakes are the Landsat OLI images; red lines denote lake boundary obtained by field survey; blue lines denote outlines derived from satellite imagery).

6 Distribution and changes of HMA glacial lakes

As indicated by the achieved HMA glacial lake inventory, 28,953 (1955.94 ± 259.68 km²) glacial lakes were identified in 2018 and their distribution had considerable spatial heterogeneity (Fig. 7). The greatest concentration of glacial lakes was in Eastern Himalaya (308.23 ± 37.29 km², which accounted for 15.8 % of the total area of glacial lakes) and Inner Tibet (299.67 ± 38.75 km², which accounted for 15.3 % of the total area of glacial lakes). Relatively few glacial lakes were found distributed in Eastern Kun Lun and Qilian Shan (38.85 ± 6.04 km², which accounted for 2.0 % of the total area of glacial lakes) and Eastern Tien Shan (40.55 ± 6.98 km², which accounted for 2.1 % of the total area of glacial lakes). The HMA glacial lakes are located within the elevation range of 1600–6300 m. An approximate normal distribution was presented both for the lakes of the entire HMA region and for the lakes in most sub-regions. More than 46 % of the HMA lake area has survived within the vertical range of 4500–5400 m, with the peak lake area of 256.13 ± 34.25 km² (accounting for 13.1 % of the total area) in the range of 5000–5200 m. The elevation band of peak lake area in the different sub-regions varied from 2400–2600 m in Altay and Sayan to 5200–5400 m in Central Himalaya, Karakoram, and Western Kun Lun.
Figure 7. Distribution and change of glacial lake area in the entire HMA and its 13 sub-regions from 1990–2018.

The HMA glacial lakes experienced widespread areal expansion during 1990–2018 with an average rate of increase in area of 15.5 % (Fig. 7). The rate of change of area varied widely between different sub-regions and different 200-m elevation bands. The glacial lakes in Altay and Sayan experienced the most rapid expansion in area during 1990–2018 with an average rate of increase of 45.6 %, whereas the rate of change was only 7.2 % in Pamir. Glacial lakes have tended to develop to higher elevations during recent decades with the maximum distribution elevation of 6078 m in 1990 rising to 6247 m in 2018. The rate of change of glacial lakes in the different 200-m elevation bands presented a large average trend against elevation, rising as a whole during 1990–2018 (Fig. 7). The lake area expanded most from the elevation of approximately 5000 m and the rate of expansion reached approximately 35 % above 5800 m in the entire HMA region, although it differed between different sub-regions. Lake area showed a notable rate of increase with elevation in most sub-regions, e.g., Hissar Alay and Western Tien Shan, Hindu Kush, Eastern Himalaya, Hengduan Shan, Eastern Tien Shan, and Kun Lun.
Shan, and Altay and Sayan. Although the rate of expansion varied violently, no observable trends were discovered in Karakoram and Western Kun Lun, Western Himalaya, and Inner Tibet. The rate of expansion in Central Himalaya and Southern and Eastern Tibet was found to have decreased with increasing elevation (Fig. 7).

7 Data availability
The data set developed in this study comprised two .shp file documents containing the glacial lake inventory of the HMA region in 1990 and 2018. The data now can be accessed on the National Special Environment and Function of Observation and Research Stations Shared Service Platform (China) at http://dx.doi.org/10.12072/casnw.064.2019.db (Wang et al., 2019a).

8 Conclusions
A glacial lake inventory of the HMA region was realized based on satellite remote sensing data and GIS techniques. Eight attribute items were recorded in the glacial lake inventory data set of the HMA region. Lake area error was assessed carefully with respect to theoretical analysis of lake boundary pixels and actual boundaries derived by GPS field-based surveys. On average, the deviations between the delineation of lake boundaries derived using the two methods were within ±0.5 pixels (±15 m). The relative area errors of each lake in 2018 varied from 2–85 %, and the average relative area errors of ±13.3 % in the entire HMA region were characterized by increase in the relative area error with decreasing lake size.

Overall, 28,953 glacial lakes with a total area of 1955.93 ± 259.68 km² were catalogued in 2018 in the HMA region. Glacial lakes survived in all 13 sub-regions of HMA from the elevation of 1600 to 6300 m. Glacial lakes were found concentrated in the sub-regions of Eastern Himalaya and Inner Tibet and at elevation bands of 4500–5400 m. The HMA glacial lakes have experienced widespread expansion in area with an average rate of increase in area of 15.5 %. Lake area expanded most in the higher elevation bands during 1990–2018. The data set has been developed as basic data for cryosphere hydrology research; however, it is expected that it could support practical utilization and management of water resources and assessment of glacier-related hazards in the HMA region.

9 Acknowledgements
The study was funded by the National Natural Science Foundation of China (No. 41771075, No. 41571061, and No. 41271091). The authors are grateful to the lake boundary vectorization operators who were not written as authors, there are Yao Chao, Chen Shiyin, Zhu Xiaoxi, Li Ruijia, Huang Rong,
Peng Xin, Xiang Lili, Yi Ying, Liu Yanlin, Fu Yongqiao, Ran Weijie, Gu Ju. We thank James Buxton MSc from Liwen Bianji, Edanz Group China (www.liwenbianji.cn/ac), for editing the English text of this manuscript.

10 References


