Constructing dataset of classified drainage areas based on surface water-supply patterns in High Mountain Asia

Lu, Jieyu; Qiu, Yubao; Wang, Xingxing; Liang, Wenshan; Xie, Pengfei; Shi, Lijuan; Menenti, Massimo; Zhang, Dongshui

DOI
10.1080/20964471.2020.1766180

Publication date
2020

Document Version
Final published version

Published in
Big Earth Data

Citation (APA)
Lu, J., Qiu, Y., Wang, X., Liang, W., Xie, P., Shi, L., Menenti, M., & Zhang, D. (2020). Constructing dataset of classified drainage areas based on surface water-supply patterns in High Mountain Asia. *Big Earth Data*, *4*(3), 225-241. https://doi.org/10.1080/20964471.2020.1766180

Big Earth Data

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tbed20

Constructing dataset of classified drainage areas based on surface water-supply patterns in High Mountain Asia

Jieyu Lu , Yubao Qiu , Xingxing Wang , Wenshan Liang , Pengfei Xie , Lijuan Shi , Massimo Menenti & Dongshui Zhang

To cite this article: Jieyu Lu , Yubao Qiu , Xingxing Wang , Wenshan Liang , Pengfei Xie , Lijuan Shi , Massimo Menenti & Dongshui Zhang (2020) Constructing dataset of classified drainage areas based on surface water-supply patterns in High Mountain Asia, Big Earth Data, 4:3, 225-241, DOI: 10.1080/20964471.2020.1766180

To link to this article: https://doi.org/10.1080/20964471.2020.1766180

Published online: 13 Jul 2020.

BIG EARTH DATA
2020, VOL. 4, NO. 3, 225–241
https://doi.org/10.1080/20964471.2020.1766180

DATA ARTICLE

OPEN ACCESS Check for updates

Constructing dataset of classified drainage areas based on surface water-supply patterns in High Mountain Asia

Jieyu Lu[a,b], Yubao Qiu[b,c], Xingxing Wang[b,c], Wenshan Liang[b], Pengfei Xie[b], Lijuan Shi[b,c], Massimo Menenti[b,d] and Dongshui Zhang[a]

[a]School of Resource & Environment and Safety Engineering, Hunan University of Science and Technology, Xiangtan, Hunan, China; [b]Key Laboratory of Digital Earth Sciences, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China; [c]AIR-FMI Joint Research Center for Arctic Observations, Aerospace Information Research Institute, Chinese Academy of Sciences, and Finnish Meteorological Institute, Sodankylä, Finland; [d]Department of Geoscience and Remote Sensing, Civil Engineering, Delft University of Technology, Delft, The Netherlands

ABSTRACT

The High Mountain Asia (HMA) region, ranging from the Hindu Kush and Tien Shan in the west to the Himalaya in the south with an altitude between 2000 and 8844 m, holds the largest reservoir of glaciers and snow outside Earth Polar Regions. In the last decades, numerous glaciers and lake areas there have undergone tremendous changes with water redistribution. In order to increase understanding of the pattern of distribution of water resources, and their dynamic changes at the basin scale, a watershed classification based on the water replenishment patterns dataset was constructed. The input dataset are from the Randolph Glacier Inventory V.6.0 and the vector data of rivers and streams. Four datasets were thus obtained: Glacier-fed and Runoff-fed Drainage Area (GRDA), Glacier-fed and Runoff-free Drainage Area (GDA), Glacier-free and Runoff-fed Drainage Area (RDA), and the Glacier-free and Runoff-free Drainage Area (NGRDA), and the numbers of these four types of basins are 87, 107, 32, and 448 separately. The statistical results show GRDA has the largest surface area, accounting for 82.2% of the total basin area in HMA, mainly in the region of the basin with outflow rivers or streams. Dominated by small basins, the GDA area accounts for the smallest area, only 3.86% and the RDA accounts for 5.62%. For NGRDA, most are with small areas, accounting for 8.32%, and mainly distributes in the closed basin of the Qiangtang Plateau. This dataset provides a fundamental classified data source for research on water resources, climate, ecology, and environment in HMA. The published data are available at https://data.4tu.nl/download/uuid:d07d748f-d10b-4308-9626-199ef05cc9af/ and http://www.dx.doi.org/10.11922/sciencedb.923.

ARTICLE HISTORY
Received 4 February 2020
Accepted 22 April 2020

KEYWORDS
High Mountain Asia; water resources; glacial recharge; watershed classification; surface water system

1. Introduction

High Mountain Asia (HMA) is an Asia high-altitude mountain region centered with Qinghai-Tibet Plateau, which holds the largest reservoir of glaciers and snow outside Earth polar regions, ranging from the Hindu Kush and Tien Shan in the northwest to the

CONTACT Yubao Qiu qiuyb@aircas.ac.cn Key Laboratory of Digital Earth Sciences, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

eastern Himalaya in the southeast, with altitude between 2000 and 8844 m (Bolch et al., 2019; Guo et al., 2015; Loomis et al., 2019; Qiu et al., 2019; Shean et al., 2019). It is an important area of alpine water resources at low and middle latitudes, contains different water storage types, for example, glaciers, river and streams, lakes, permafrost, and feed to several large rivers and endorheic basins which are therefore known as "Asia's water tower" (Barnett, Adam, Lettenmaier, 2005; Immerzeel, Van Beek, & Bierkens, 2010; Pritchard, 2017; Qiu, 2018; Yao et al., 2019; Zhu et al., 2019). Dynamic changes of water resources have important impacts and feedback to the water and energy balance in HMA, influencing the regional and global climate.

Generally, the glaciers and lakes are main components of High Asian water resources, which are very sensitive to global warming and its induced environmental variation. The high concentration of lakes plays an important role in the water cycle of HMA. Recently, lakes over Tibetan Plateau have expanded rapidly (Zhang, Luo, Chen, & Zheng, 2019), and the glacier area decreased significantly in recent decades (Brun, Berthier, Wagnon, Kääb, & Treichler, 2017; Ding, Liu, Li, & Shangguan, 2006; Kääb, Berthier, Nuth, Gardelle, & Arnaud, 2012; Xiao et al., 2007; Yao et al., 2010; Zhang et al., 2015), for example, the area of the Gangdisê Glacier decreased by a total of 854.05 km^2 (−39.53%) from 1970 to 2016, with the rate of change being as high as −1.09%/year (Liu J, Yao, Liu, Guo, & Xu, 2019). The melt water from glaciers is one of the sources for several major rivers in the surrounding areas of HMA (Immerzeel et al., 2010; Phan, Lindenbergh, & Menenti, 2013; Qiu et al., 2017; Yang et al., 2017; Zhang et al., 2020). As for the second largest lake on top of Plateau, Nam Co, the water supply from glacial melt plays an increased role to its expansion, accounting for 8.55% contribution in 1971 – 1991 to 11.48% in 1992 – 2004, though the precipitation contributed more than 60% (Zhu, Xie, & Wu, 2010).

For most lakes, the supply is mainly from the supply pattern of surface recharge (Wang, Siegert, et al., 2013; Wang, Gong, et al., 2013; Ju et al., 2017; Song, Sheng, Ke, Nie, & Wang, 2016), including melting water and precipitation. Since the late 1990s, it was reported that the rapid expansion of lakes has been consistent with regional climate changes, especially including the increased precipitation being the dominant fraction of lake water storage gains, according to ~70%, of expansion of lakes in the basins over the Tibetan Plateau (Lei et al., 2014; Li & Sheng, 2013; Zhang et al., 2017a). A significant increasing in precipitation played a main role to the expansion of lakes (Dong, Xue, You, & Peng, 2014; Lei et al., 2014; Li et al., 2017). The area of lakes in the Shenza Basin is expanding – these changes in lake area are mainly linked to temperature and precipitation, which are in turn caused by the background climate warming and humidification (Yi, Deng, Li, & Zhang, 2015). Further detailed studies have shown that the main reasons of lake expansion with an area of more than 10 km^2 on the Tibetan Plateau after 2000 are changes in precipitation and evaporation (Song, Huang, Richards., Linghong, & Vu, 2014; Zhang et al., 2017b). While, for specific regions, a decrease in evaporation is the main reason for lake expansion in the Hoh Xil region, with an increase in glacial and frozen-soil melt water being a secondary factor (Yao et al., 2013). An increase in temperature, the precipitation, and glacier area induced by warming caused the lakes in the Nagqu and Hoh Xil regions resulting in general expansion between 1975 and 2016 (Wan et al., 2014).

In order to understand these changes of water distributions and the supply patterns in HMA, the drainage basins in this region were divided into two types: the endorheic or closed basins (here after, closed basins is used), where the water does not leave the basins

as surface runoff, and the outflow basins, where the water flow into region outside HMA, for example, ocean and inland area of Center Asia or Tarim basin, indicating the potential loss of water through the surface runoff (Gao, Wang, Yao, Lu, & Lu, 2018). The definition and vector data can be derived from China Lakes Record (Sumin & Hongshen, 1998) and the Atlas of Lakes Distribution in China (Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, 2015), and the existing lakes database for HMA. However, there are no fundamental classified dataset of basins with different water supply patterns, gain and loss of water that can be directly used to map the changes and distributions of water resource patterns in HMA.

In this paper, we introduced and constructed a classified dataset by different water supply patterns, which was produced based on the hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales, the HydroSHEDS (Lehner, Verdin, & Jarvis, 2006), the last updated Randolph Glacier Inventory V6.0 (RGI Consortium, 2017), and the detailed runoff vector dataset of rivers and streams for the whole HMA. The accuracy of the dataset was verified by visual interpretation based on the high resolution of Google Earth images. Finally, four vector dataset of drainage area (DA) were obtained, the Glacier-fed and Runoff-Fed Drainage Area (GRDA), Glacier-fed and Runoff-free Drainage Area (GDA), Runoff-fed and Glacier-free Drainage Area(RDA), and Glacier-free and Runoff-free Drainage Area (NGRDA, here N represents none). These publicly opened datasets provide the necessary support for further research about the distribution, management, and research on the changes of water resources in HMA.

2. Methodology and data processing

Except the Antarctic and the Arctic, the HMA known as the third largest volume of solid water on Earth covers the area between 25°N and 45°N, and 67°E and 107°E (Figure 1) (Barnett, Adam, Lettenmaier, 2005; Guo et al., 2015; Loomis et al., 2019). Through the personal communication with the author of boundary of Tibetan Plateau (Zhang, Li, & Zheng, 2002) and NASA High Mountain Asia Team (HiMAT) (Shean, et al., 2019; Joseph, 2019), the HMA boundary (Figure 1) was defined as the region with altitude above 2000 min High Asia, and some south part of it was adjusted to make it reasonable geographically by Professor Zhang Yili (Zhang et al., 2002). In Figure 1, it shows that the distribution of glaciers, river and streams, and lakes differs significantly from basin to basin. In terms of number and area, glaciers in closed basins are larger than that of basin. It can be seen that the spatial distribution of rivers and streams lies in region with densest outflow rivers and streams. Every basin contains one or more lakes, which are mainly located in the closed or endorheic basins of Tibetan Plateau, and there is the biggest number of highest altitude and largest lakes on Earth (Yao et al., 2013).

2.1. Input data

The fundamental dataset used in these data paper are described below.

2.1.1. The HydroSHEDS dataset

The Hydrological data and maps based on Shuttle Elevation Derivatives at multiple Scales (HydroSHEDS), which was developed by the Conservation Science Program of the World

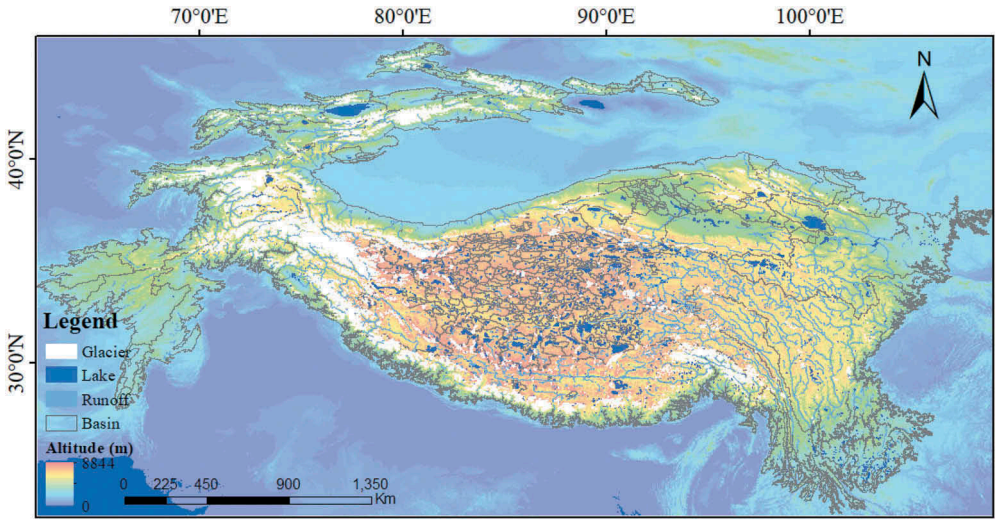

Figure 1. The boundary of High Mountain Asia is the outlier of Plateau region, where the HMA is with glacier, lakes, surface rivers and streams, and watershed boundaries inside. The boundary of HMA was one of the output of CAS-NASA collaboration on glacier and disaster by space observations (Guo & Freilich, 2015), defined by the private communication with author of boundary of Tibetan Plateau (Zhang et al., 2002), which was defined by the altitude above 2000 m, with some minor geographical adjusts. The runoff, rivers and streams, is from the surface water body separation from GlobeLand30-WTR2010 (J. Chen et al., 2014); The HydroSHEDS (Lehner et al., 2006) was used to delineate the basins. Lakes are from HydroLAKES (Messager, Lehner, Grill, Nedeva, & Schmitt, 2016), and glaciers are from RGI 6.0.

Wildlife Fund (WWF), is a global-scale watershed dataset obtained from http://hydrosheds.org/downloads (Lehner et al., 2006). Extracted from Shuttle Radar Topography Mission (SRTM) 90 m elevation data, the HydroSHEDS includes the surface runoff network, watershed boundaries, drainage direction, and other auxiliary data, such as river topology and runoff distance. According to this basis dataset, the basins in HMA were divided into 674 sub-basins through the intersection analysis of DA from HydroSHEDS and the boundary of HMA in Figure 1.

2.1.2. The Randolph glacier inventory v.6.0

The Randolph Glacier Inventory (newly updated v.6.0) is a globally complete inventory of glacier outlines (Pfeffer et al., 2014; RGI Consortium, 2017), which is supplemental to the database compiled by the Global Land Ice Measurements from Space initiative (GLIMS) (http://www.glims.org/RGI/), and also includes the updated glacier inventory dataset of China region. Considering that the SGI-China only covers China, and the Randolph Glacier Inventory V6.0 covers the entire HMA, the Randolph Glacier Inventory V6.0 is used as the glacier data in this paper.

The HMA part, including most of the glacier from Central Asia, South Asia West, and South Asia East from RGI 6.0 with an covering area of ~97,605 ± 7,935 km^2 (assuming ~8% uncertainty; Pfeffer et al., 2014), being taken from the Glacier Area Mapping for Discharge from the Asian Mountains (GAMDAM) (Nuimura et al., 2015), the Second Glacier Inventory

of China (SGI-China) at http://www.crensed.ac.cn/ (Liu et al., 2014, Guo et al., 2015) and as-yet unpublished work at the Technical University of Dresden and University of Zirich (RGI Consortium, 2017).

2.1.3. Vector data of rivers and streams

The vector dataset for rivers and streams in HMA was derived from the Global Land Surface Water Dataset at 30 m resolution obtained in 2010 (GlobeLand30-WTR2010) at http://dx.doi.org/10.3974/geodb.2014.02.01.v1 (Chen et al., 2014). The vector data of water surface for rivers, streams, and lakes are independently obtained through a series of steps of vectorization for surface water bodies, projection transformation, then all the vector dataset was separated into lakes (isolated water bodies), linear rivers, and streams based on the lakes from HydroLAKES as reference (Messager et al., 2016), with manual correction later on. The vector data of rivers or streams are cropped to construct the vector dataset for the whole HMA. The high-resolution images from recent Google Earth were used to evaluate; The result shows that the river and stream vector data have an accuracy of up to 98.897%, it achieved a more detailed surface water river and stream system, when compared with the existing river and stream database – Global River Widths from Landsat (GRWL) (Allen & Pavelsky, 2018; Messager et al., 2016), which shows a big discrepancy in Asia region, especially the HMA. All these datasets will be discussed in another paper.

2.2. Methodologies for data processing

The dataset based on the HydroSHEDS, RGI v.6.0 and HMA vector data for rivers and streams were used to create the map with different water resource supply patterns. The specific steps in this data process are shown in Figure 2, including data preparation with intersection processing and the accuracy of evaluation and compassion.

The details are described below:

Step 1: Determine whether there are glaciers in individual basins

Firstly, the boundary of HMA was used to clip the watershed data vector data from the HydroSHEDS. The basin boundary was then intersected with the vector data of glacier from RGI v.6.0. For each basin, it was then determined whether the basin had a glacier inside. The presence of glacier vector data for the basin means that the basin does have a melt water supply. Generally, one glacier normally links with one particular basin, while more could be also seen from RGI v.6.0 dataset that some glaciers fall into more than one basins. In such cases, we assumed that those basins do have a glacial melt water supply. The vector data for glacier-fed drainage areas and glacier-free drainage areas were separated as the grey basins and dark blue basins in Figure 3. Of the 674 basins in HMA, there were found to be 194 glacier-fed drainage areas and 480 glacier-free drainage areas for the whole HMA.

Step 2: Ascertain whether there is surface runoff rivers or streams in individual basins

The watershed boundary of HMA was then intersected with the vector data of the rivers and streams, represented to be runoff symbol. For each basin, it was determined whether there was any surface runoff, which are linearly rivers and streams. If there are rivers or streams in the basin, then it is considered to be recharged also by surface runoff through the rivers or streams, either from the precipitation or the glacier melt water. Generally, runoff, river, or stream belongs to the particular basins.

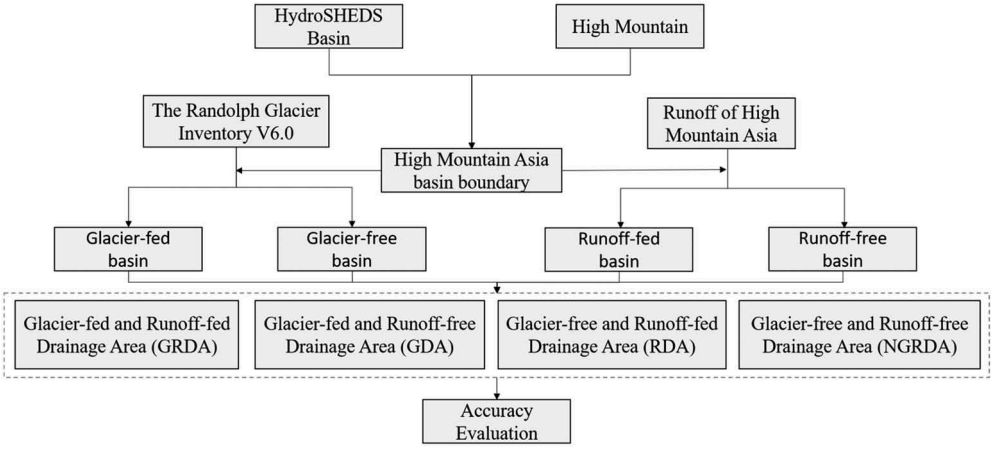

Figure 2. The data processing flow chart for the production of four types of basins in HMA. They are Glacier-fed and Runoff-fed Drainage Area (GRDA), Glacier-fed and Runoff-free Drainage Area (GDA), Glacier-free and Runoff-fed Drainage Area (RDA), and Glacier-free and Runoff-free Drainage Area (NGRDA).

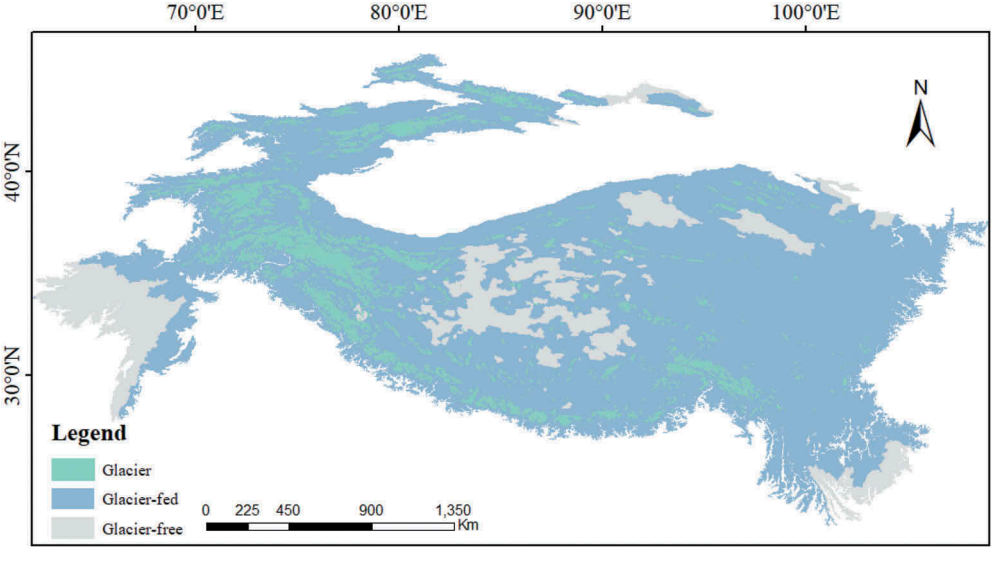

Figure 3. Distribution of basins with and without a potential glacier melt water supply, obtained from the intersection between the glacier map (RGI .0) and basins boundary data from HydroSHEDS in High Mountain Asia.

Data of the runoff-fed drainage area were then extracted through the geometric intersection between river and streams and the watershed dataset in HMA. This was followed by non-overlapping vector data to be recognized as runoff-free basin vector layer. Of the 674 basins in HMA, there were 119 runoff-fed drainage areas and 555 runoff-free drainage areas.

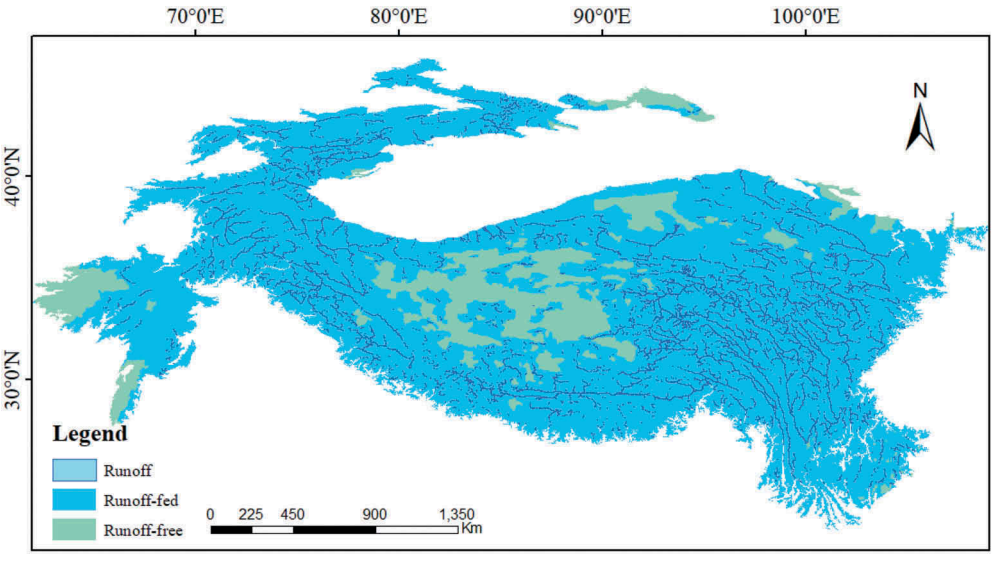

Figure 4. Distribution of basins with and without runoff inside, as determined using the remote-sensing surface water vector dataset, and its intersection with basins boundary data from HydroSHEDSin High Mountain Asia.

It can be seen from Figure 4, the distributions of the different types of basins exhibit significantly different spatial characteristics. The basins with dense rivers or streams are outflow area, mainly distributed in the east, southeast, and northwest of the HMA. The basins that are not obviously supplied by surface runoff water system, like rivers or streams, are mainly distributed in the closed basins in HMA and are located in the central, northern, and north eastern parts of the HMA region.

Step 3: Construction of dataset for classified drainage areas in HMA based on surface water-supply patterns

The overlapping parts of glacier, rivers or streams, and basin vector data were extracted to obtain the Glacier-fed or Glacier-free, and Runoff-fed or Runoff-free drainage areas. Similarly, the geometric intersection between the GRDA, GDA, RDA, and NGRDA was determined. This step produced four sets of vector data: GRDA, GDA, RDA, and NGRDA. Thus, the water resource supply patterns–based dataset for HMA was mapped and constructed. The accuracy evaluation was technically addressed in Section 5.

3. Result

The obtained dataset includes four sets of vector data, namely GRDA, GDA, RDA, and NGRDA. The distribution of these drainage areas are shown in Figure 5.

There are a total number of 87 basins that could be fed by glacier melting water with obvious runoff systems. Most of these are basins with outflow rivers or streams and distributed in the eastern, western, and southern regions of HMA. Dominated by large-scale basins, the GRDA has the largest surface area in four types of basins, accounting for 82.2% of the total basin area of HMA regions. Basins with the biggest number of glaciers and largest glacier area are distributed in the north western Tarim Basin and the southern Yarlung

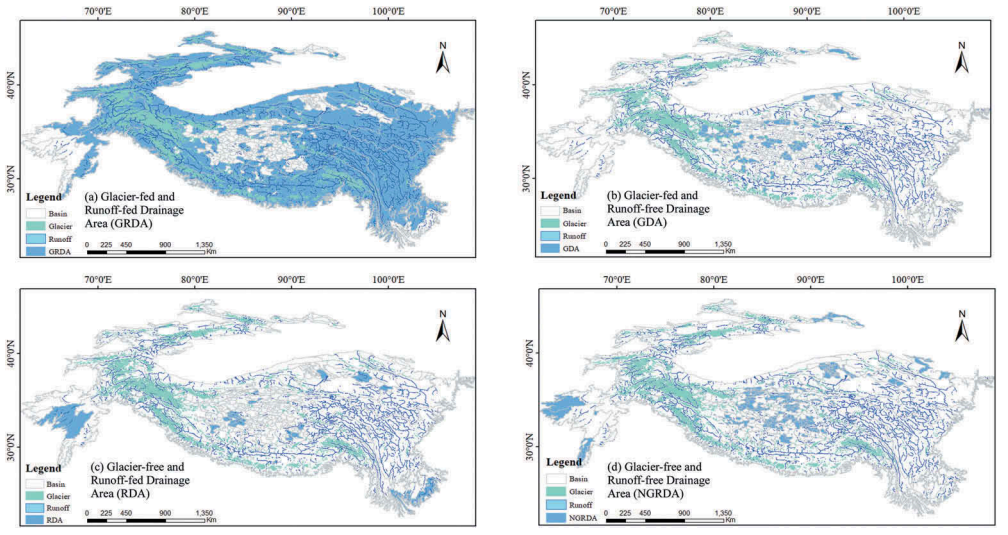

Figure 5. Dataset produced in this study, consisting of: (a) GRDA, basins with water supply from melting glaciers and an obvious runoff system, (b) GDA, basins with water supply from melting glaciers but without an obvious runoff system, (c) RDA, basins with water supply from an obvious runoff system but not melting glaciers, and (d) NGRDA, basins without an obvious water supply from glaciers or runoff water system.

Zangbo river basin. The basins with the densest rivers or streams system are found in the eastern Yangtze River Basin, Lancang River Basin, Salween River Basin, and the southern Brahmaputra River Basin; as for the GDA, 107 basins are potentially supplied by glaciers melting water but do not have obvious surface runoff systems, that is to say a recent supply from surface water runoff is weak; most of these are closed basins and mainly distributed in central and north eastern plateau of HMA. Dominated by small basins, the GDA area accounts for the smallest surface area of these four types of drainage area – only 3.86% of the total. There are 32 basins where the water supply is obviously witnessed by runoff systems, such as surface water of rivers or streams, but without glaciers melting water supply, namely RDA, which are mainly located in the western, south eastern, north eastern, and part of the central region of HMA. The RDA accounts for 5.62% of the total basin area of HMA region. A number of NGRDA basins are 448, which means that they are either surface runoff-fed nor the glacier melting water-fed. Most of these are small, closed basins, mainly distributed in the central, south western, and north eastern regions, they account for 8.32% of the total basin area of HMA regions.

4. Data records

The fundamental dataset described in the previous section was based on the HydroSHEDS hydrological data (basin range), combined with RGI v.6.0, and rivers and streams vector data produced by vectorization from GlobeLand30-WTR2010, and all were used to produce a complete classification dataset with different four water supply patterns for individual basins in HMA. The dataset includes four vector files at Shape file format, which have a total

volume size of 3.27 M, all these data files were stored in two databases for open sharing publicly, see section "Data Availability Statement." The data attributes include the area of each watershed. Users can use geographic information system (GIS) software, like ArcMap for data handling. This dataset can be used for research on water resource, hydrological, mass and energy balance, and climate change in HMA and is of importance for studies into the environment, disaster application in this region.

5. Technical validation

5.1. Reliability of the input dataset

The glacier dataset was compiled using high-resolution Landsat TM and ETM+ images, including that the boundaries of bare ice areas were extracted using the well-known band-ratio segmentation methods, in this way, the vector data for the boundary of each individual glacier for the main glacier regions in western China were obtained (Liu et al., 2014, 2015). The glacier attribute data were extracted based on the latest global digital elevation model, and a widely used algorithm was used to calculate the attribute for individual glaciers in the main glacier regions of western China. The RGI v.6.0 dataset in HMA is mainly from SGI-China, which has a high degree of positional accuracy, the absolute error, and relative error for the total glacier area were ±1411 km^2 and 3.2%, respectively (Liu et al., 2015). The dataset was recognized to be reliable as input to this study.

The HMA rivers and streams vector data were based on the published global 30-m resolution land surface water dataset compiled in 2010 (Chen et al., 2014). The vector data, which had been converted from raster data, were processed using projection transformation, and cropping to obtain the rivers, streams, and lakes with the aid of existing HydroLAKES data (Messager et al., 2016) as references. The rivers and streams vector dataset for the HMA region were examined and verified using the newly released GRWL dataset (Allen & Pavelsky, 2018), based on high-resolution Google Earth multi-temporal image data. The evaluation produced an accuracy of 98.897%, which was found to be a good complement to the GRWL dataset in HMA (Section 2.1).

5.2. Overall accuracy evaluation

5.2.1. Establishment of interpretation characteristics

Remote-sensing image interpretation is the process of determining image features to analyze the attributes of Earth surface, that is to identify surface features (Guan & Liu, 2007; Zhao, 2013). Therefore, it is necessary to interpret high-resolution satellite data as the reference data for these four datasets evaluation. As most of the Google Earth images are true color images, visual interpretation of imagery was used to establish interpretation characteristics for glaciers, rivers, and streams based on relevant literature and empirical knowledge.

The global imageries from Google Earth have an effective resolution of 100 m at least, and it also provides high-resolution images of resolutions about 0.5–1 m for large cities, which makes it easy to identify glaciers, rivers, streams, and lakes. The specific interpretation characteristics are described in Table 1, with detailed visual examples shown in Figure 6. Using this method, examples of basins for validation were built for the overall assessment to four types of datasets, described in Section 5.2.2.

Table 1. Interpretation characteristics for glacier and runoff in HMA.

Type	Shape	Tone	Texture
Runoff (rivers and streams)	Strips and branches	Brown or dark brown, even color tone	Single image structure
Glacier	Generally block-shaped with clearer boundaries	White with more uniform tone	Image structure is relatively rough

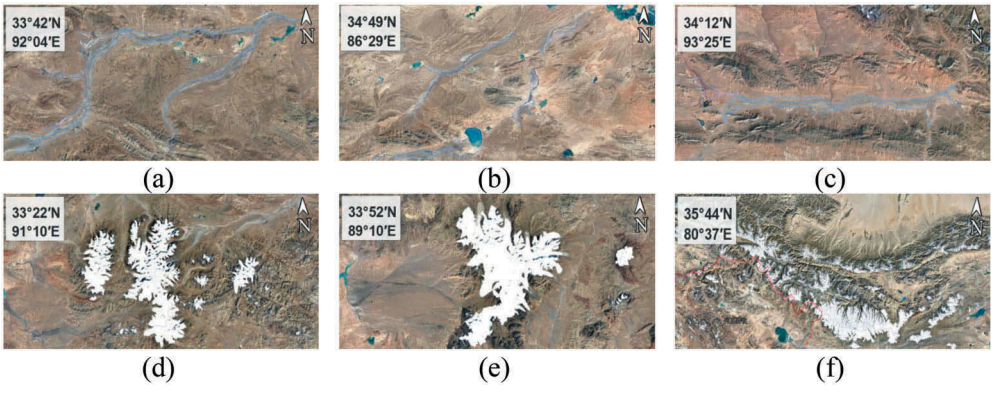

Figure 6. Examples for the surface interpretation characteristics: The figures (a), (b), and (c) are examples for runoff system, such as rivers and streams, where lake are easily interpreted with dark blue color, the rivers and streams are slightly in grey color. The figures (d), (e), and (f) are examples of glaciers in high-resolution images.

5.2.2. Result of evaluation

Stratified sampling method was used for the sample selection for verification to secure the representative, that is individual samples are randomly selected from different strata according to a prescribed sample ratio, this method provides better representation and a smaller sampling error (Cao, Wang, Li, & Jiang, 2008).

Samples were randomly selected from the four drainage areas: 44 GRDA, 16 GDA, 54 RDA, and 224 NGRDA were thus obtained. The samples were imported into Google Earth to calculate the accuracy for 11 years from 2006 to 2016. To reduce the impact from wintertime river-freezing and snow cover on the surface, 2006–2016 summer imagery from Google Earth was then used as verification samples. The accuracy formula can be expressed as a = b/(c + b), where a is the overall accuracy, b is the number of correct (true) random samples, and c is the number of incorrect (false) random samples.

From the error statistics in Table 2, it can be seen that the annual average accuracy is over 89.9% and the overall accuracy is high as 93.18%. The accuracy for 11 years is quite stable, and no significant variation was found yearly, this certificated the stable accuracy for almost 11 years. The highest overall accuracy (96.23%) is that for the NDRGA properly owing to the stability of the basins that are not affected by glaciers melting water, and runoff of river and streams. In contrast, the GRDA has a lower accuracy, which may be due to the glacier shrinkage resulting from global warming as well as changes in stream morphology (e.g., in straightness, position, and width) and length on a seasonal or annual basis. The GDA has the lowest average accuracy, which might be with the more frequency of glacier changes, as the accuracies are linearly decreasing from 2006 to 2016. With this

Table 2. Overall accuracies for four types of dataset in 11 years from 2006 to 2016.

Year	GRDA	GDA	RDA	NGRDA
2006	40/(4 + 40) = 90.91%	50/(4 + 50) = 92.59%	15/(1 + 15) = 93.75%	215/(9 + 215) = 95.98%
2007	40/(4 + 40) = 90.91%	49/(5 + 49) = 90.74%	15/(1 + 15) = 93.75%	216/(8 + 216) = 96.43%
2008	41/(3 + 41) = 93.18%	49/(5 + 49) = 90.74%	15/(1 + 15) = 93.75%	217/(7 + 217) = 96.88%
2009	40/(4 + 40) = 90.91%	51(3 + 51) = 94.44%	15/(1 + 15) = 93.75%	219/(5 + 219) = 97.77%
2010	42/(2 + 42) = 95.45%	49/(5 + 49) = 90.74%	16/(0 + 16) = 100%	214/(10 + 214) = 95.54%
2011	39/(5 + 39) = 88.64%	48/(6 + 48) = 88.89%	15/(1 + 15) = 93.75%	212/(12 + 212) = 94.64%
2012	40/(4 + 40) = 90.91%	45/(9 + 45) = 83.33%	14/(2 + 14) = 87.50%	215/(9 + 215) = 95.98%
2013	40/(4 + 40) = 90.91%	49/(5 + 49) = 90.74%	16/(0 + 16) = 100%	216/(8 + 216) = 96.43%
2014	40/(4 + 40) = 90.91%	48/(6 + 48) = 88.89%	15/(1 + 15) = 93.75%	216/(8 + 216) = 96.43%
2015	39/(5 + 39) = 88.64%	48/(6 + 48) = 88.89%	16/(0 + 16) = 100%	217/(7 + 217) = 96.88%
2016	40/(4 + 40) = 90.91%	48/(6 + 48) = 88.89%	16/(0 + 16) = 100%	214/(10 + 214) = 95.54%
Average	91.12%	89.90%	95.45%	96.23%

overall accuracy testing, the dataset is suggested to be used as a basic data source for research, especially the water management, as well as provides support to the global big Earth data-based Earth three poles comparison study at a relevant high accuracy.

6. Statistics and analysis

The lake dataset from HydroLAKES was also used as basic dataset for understanding the lake changes in water resource analysis to these four datasets. Statistical analysis revealed that the GRDA has the largest number of lakes, accounting for 3,853 lakes, with a surface area of 38,356.96 km^2, while the GDA has the smallest number of lakes, accounting for 340 lakes, covering an area of 5,656.58 km^2. The RDA contains 398 lakes with a total surface area of 1,889.16 km^2, and the NGRDA has 653 lakes with a total surface area of 4,523.95 km^2, the results are all listed in Table 3.

There are significant differences in the distribution of lakes between four types of drainage area. From Table 3, the lakes in the GRDA account for 73.47% of all lakes and 75.99% of the total area of lakes; the lakes in the GDA are smallest in number, accounting for only 6.48% of the total number of lakes and 11.34% of the total lake area; the lakes in the RDA account for 7.59% of the lakes and 3.73% of the total lake area; and the lakes in the NGRDA account for 12.46% of the number of lakes and 8.94% of the total lake area. The spatial distributions are shown in Figure 7. The map shows the coverage of individual lakes as a percentage of the drainage areas with gradient blue color legend. Lakes covering the largest percentages of their respective drainage areas are found in the center to north western part of the HMA region; generally speaking, the closed basins contains lakes that cover a high percentage of their respective drainage areas, that is to say the surface water are quite dominated and active in the center of HMA, most of which are GRDAs.

Table 3. Statistical analysis of number and area for four different water supply drainage areas in HMA.

Types	Basin area (km^2)	Number of lakes	Lake area (km^2)	Proportion in total number of lakes and total area of lakes
GRDA	3,700,581.54	3,853	38,356.96	73.47%/75.99%
GDA	173,499.31	340	5,656.58	6.48%/11.34%
RDA	252,740.26	398	1,889.16	7.59%/3.73%
NGRDA	374,526.66	653	4,523.95	12.46%/8.94%

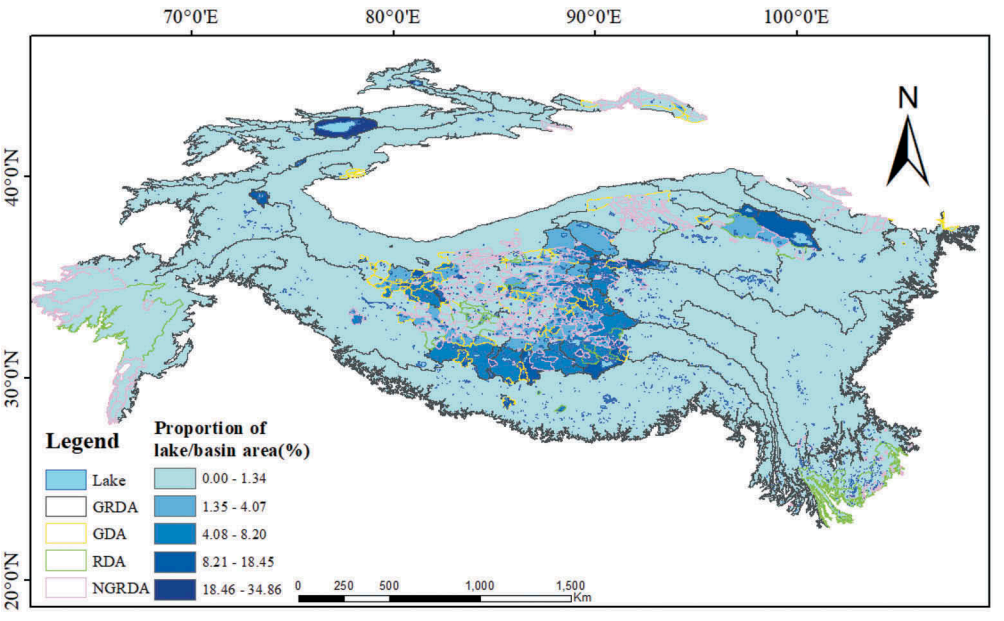

Figure 7. This map shows the fractional area (%) covered by lakes potentially supplied by both glacier-fed and runoff-fed (gray), glacier-fed and runoff-free (yellow), glacier-free and runoff-fed (green), and glacier-free and runoff-free (pink), the runoff represents the existing rivers or streams, with the lake base map from HydroLAKE (Messager et al., 2016).

A dataset that provides a long-time series of change on lake areas over Tibetan Plateau (Zhang et al., 2017b) was used to further analyze the changes of the lakes in each basin. These data contain dataset of area for 364 lakes with area of more than 10 km^2 from the 1970s to 2013. An analysis was carried out using the lake area data from 2001 to 2013, the statistics for the changes of lake area in each drainage area revealed that 272 lakes experienced an expansion whereas 92 lakes experienced shrinkage during this period. The total area of the first group of lakes increased to 3,966.748 km^2; the second group decreased to 1,312.83 km^2, it says that more than 75% of surface water of lakes are expanding.

From the statistic map in Figure 8, we can see the area of lake expansion was mainly concentrated in the closed basins of HMA, with few of them experienced a shrinking process. Of these lakes, 98 expanding lakes and 44 shrinking lakes were in the GRDA, the lakes that experienced expansion grew by a total of 2,943.31 km^2, whereas the area of the lakes that shrank fell by 556.68 km^2. There were 63 expanding lakes and 16 shrinking lakes in the GDA; in this area, the expanding lakes grew by a total of 754.25 km^2 and the total area of the shrinking lakes reduced by a total of 71.19 km^2, 10% of the expanding lakes in area. In the RDA, there were 11 expanding lakes and seven shrinking lakes. Here, the total areas of expansion and shrinkage were 153.39 km^2 and 22.81 km^2, respectively. In the NGRDA, 100 lakes experienced expansion, 25 experienced shrinkage and the total areas of expansion and shrinkage were 859.79 km^2 and 215.51 km^2, respectively. The NGRDA areas with the biggest number of expanding lakes are mainly distributed in the closed basins. Most of the lakes in the closed basins are expanding now, while some lakes in the basins with outflow rivers or streams are shrinking. These datasets show a clear understanding of the water changes with the supply

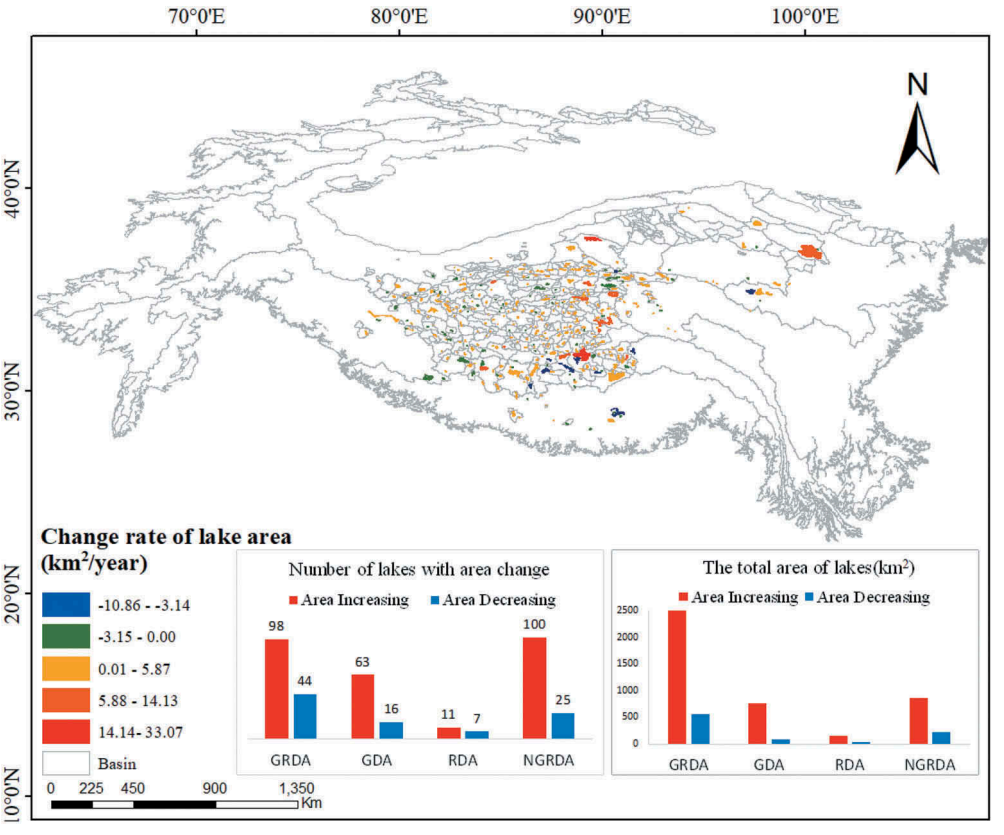

Figure 8. Change rate of lake area from 2001 to 2013 by G. Zhang (2017b),and its distribution of number and total area of all lakesfor four categories of drainage area, GRDA, GDA, RDA and NGRDA in High Mountain Asia. The glacier dominated drainage areas, like GRDA and GDA, experienced an obvious increasing trend in both number and area.

from the glacier. For example, GRDA, GDA is experiencing mainly increase in the number and area, while the lakes in NGRDA are also experiencing an obvious increase in number and area. That indicates directly other forms of water transportation is still very active for those without obvious glacier, rivers, and streams.

7. Conclusion

In this data paper, the research explains clearly the work to construct the dataset for four types of drainage areas, GRDA, GDA, RDA, and NGRDA. The aim for this dataset is to provide a simple vector dataset that can be easily used by the researcher for understanding the water resource and its changes distribution through the water supply patterns, like glacier-fed or glacier -free, surface runoff of rivers or streams runoff-fed or runoff-free. The dataset was evaluated by the recent high-resolution images from Google Earth, resulting in an apparent stable and acceptable over accuracy.

The analysis was conducted for these four datasets with the lake's distribution and the lake area changes from 2003 to 2013; several points were found, firstly, the water change

is very active in the center and north western of HMA, especially for the lake in the GRDA. Another important finding is that the glacier dominant drainage areas, like GRDA and GDA, experienced a very clear increasing in number and area of lakes, roughly accounting for three times than that of the decreasing lakes. While the lake changes in the NGRDA give a clear information that the water transportation need to be understood at a more detailed level. This calls for an potential improvement at the identification of the permafrost changes information need to be included in this dataset.

The dataset are open for public and free for the researchers working in the field of the water management, environment, and climate changes.

Acknowledgments

We thank the providers of fundamental datasets, including the Randolph Glacier Inventory V6.0 by the Global Land Ice Measurements from Space (http://www.glims.org/RGI/), the Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales (HydroSHEDS) (http://hydrosheds.org/downloads) provided by the World Wildlife Fund (WWF) and its Conservation Science Program, and the long-time series of dataset of lake areas (http://poles.tpdc.ac.cn/). The vector data of rivers and streams in HMA were provided by the "Big Earth Data Science Engineering" (No. XDA19070201) of Chinese Academy of Sciences, and thanks for the support from the Working Group on High Mountain and Cold Regions (HiMAC) under Digital Belt and Road Program (DBAR).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was implemented in the Key Laboratory of Digital Earth Sciences, Chinese Academy of Sciences, and supported by the Strategic Priority Research Program of the Chinese Academy of Sciences [XDA19070201], the National Key Research and Development Program of China, MARIS Project [2017YFE0111700], and the International Cooperation Program of the Chinese Academy of Sciences, [131211KYSB20150035].

ORCID

Yubao Qiu http://orcid.org/0000-0003-1313-6313

Data Availability Statement

The datasets produced in this study are openly available in the 4TU. Research Data at https://data.4tu.nl/download/uuid:d07d748f-d10b-4308-9626-199ef05cc9af/and in the Science Data Bank at http://www.dx.doi.org/10.11922/sciencedb.923.

References

Allen, G. H., & Pavelsky, T. M. (2018). Global extent of rivers and streams. *Science*, *361*(6402), 585–588.

Barnett, T., Adam, J., & Lettenmaier, D. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, *438*, 303–309.

Bolch, T., Shea, J., Liu, S., Azam, F., Gao, Y., Gruber, S., … Zhang, Y. (2019). Status and change of the cryosphere in the extended Hindu Kush Himalaya region. *The Hindu Kush Himalaya Assessment*, 209–255. doi:10.1007/978-3-319-92288-1_7

Brun, F., Berthier, E., Wagnon, P., Kääb, A., & Treichler, D. (2017). A spatially resolved estimate of High Mountain Asia glacier mass balances from 2000 to 2016. *Nature Geoscience*, *10*(9), 668–673.

Cao, Z., Wang, J., Li, L., & Jiang, C. (2008). Strata efficiency and optimization strategy of stratified sampling on spatial population. *Progress in Geography*, *27*(3), 152–160.

Chen, J., Liao, A., Chen, L., Zhang, H., He, C., Peng, S., … Sang, H. (2014). *Global 30m resolution land surface water dataset*. Retrieved from: http://dx.doi.org/10.3974/geodb.2014.02.01.v1(2014)

Consortium, R. G. I. (2017). Randolph glacier inventory(RGI)-A dataset of global glacier outlines: Version 6.0, Technical Report, *Global Land Ice Measurements from Space*, Boulder, Colorado, USA. Digital Media. doi: 10.7265/N5-RGI-60.

Ding, Y., Liu, S., Li, J., & Shangguan, D. (2006). The retreat of glaciers in response to recent climate warming in western china. *Annals of Glaciology*, *43*(1), 97–105.

Dong, S., Xue, X., You, Q., & Peng, F. (2014). Remote sensing monitoring of the lake area changes in the Tibetan Plateau in recent 40 years. *Journal of Lake Sciences*, *26*(4), 535–544.

Gao, Y., Wang, W., Yao, T., Lu, N., & Lu, A. (2018). Hydrological network and classification of lakes on the third pole. *Journal of Hydrology*, *560*, 582–594.

Guan, Z., & Liu, J. (2007). *Remote sensing image interpretation* (pp. 188–189). Wuhan, China: Wuhan University Press.

Guo, H., & Freilich, M. (2015). Report to the 1st CAS-NASA collaboration conference on glacier and disaster in High Mountain Asia by space observations, Kathmandu, Nepal.

Guo, W., Liu, S., Xu, J., Wu, L., Shangguan, D., Yao, X., … Jiang, Z. (2015). The second Chinese glacier inventory: Data, methods and results. *Journal of Glaciology*, *61*(226), 357–372.

Immerzeel, W., Van Beek, L., & Bierkens, M. (2010). Climate change will affect the asian water towers. *Science*, *328*(5984), 1382–1385.

Joseph, M., (2019)*High Mountain Asia – cryospheric change*. Retrieved from https://josephmichaelshea.wordpress.com/2019/02/04/high-mountain-asia-cryospheric-change/

Ju, J., Zhu, L., Wang, J., Cui, Y., Huang, L., Yang, R., … Wang, Y. (2017). Estimating the contribution of glacial meltwater to Ranwu Lake, a proglacial lake in SE Tibet, using observation data and stable isotopic analyses. *Environmental Earth Sciences*, *76*(5), 229.

Kääb, A., Berthier, E., Nuth, C., Gardelle, J., & Arnaud, Y. (2012). Contrasting patterns of early twenty-first-century glacier mass change in the Himalayas. *Nature*, *488*(7412), 495–498.

Lehner, B., Verdin, K., & Jarvis, A. (2006). *HydroSHEDS technical documentation, version 1.0.[R]* (pp. 1–27). Washingtou, DC: World Wildlife Fuud US.

Lei, Y., Yang, K., Wang, B., Sheng, Y., Bird, B. W., Zhang, G., & Tian, L. (2014). Response of inland lake dynamics over the Tibetan Plateau to climate change. *Climatic Change*, *125*(2), 281–290.

Li, B., Zhang, J., Yu, Z., Liang, Z., Chen, L., & Acharya, K. (2017). Climate change driven water budget dynamics of a Tibetan inland lake. *Global and Planetary Change*, *150*, 70–80.

Li, J. L., & Sheng, Y. (2013). Spatiotemporal pattern and process of Inland Lake change in the Qinghai-Tibetan Plateau during the period of 1976–2009. *Arid Zone Research*, *30*(4), 571–581.

Liu J., Yao, X., Liu, S., Guo, W., & Xu, J. (2019). Glacier changes in the Gangdisê mountains from 1970 to 2016. *Acta Geographica Sinica*, *74*(7), 333–1344.

Liu, S., Guo, W., Xu, J., ShangGuan, D., Wu, L., Yao, X., … Wang, Y. (2014). The second glacier inventory dataset of China (Version 1.0). Cold and arid regions science data center at Lanzhou. doi:10.3972/glacier.001.2013.db

Liu, S., Yao, X., Guo, W., Xu, J., ShangGuan, D., Wei, J., … Wu, L. (2015). The contemporary glaciers in China based on the second Chinese glacier inventory. *Acta Geographica Sinica*, *70*(1), 3–16.

Loomis, B. D., Richey, A. S., Arendt, A. A., Appana, R., Deweese, Y.-J. C., Forman, B. A., … Shean, D. E. (2019). Water storage trends in High Mountain Asia. *Frontiers in Earth Science*, *7*. doi:10.3389/feart.2019.00235

Messager, M., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, *7*, 13603.

Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences ed. Distribution of lakes Atlas in China. [EB/OL]. (2015).Retrieved from http://lake.geodata.cn/lake_museum/

Nuimura, T., Sakai, A., Taniguchi, K., Nagai, H., Lamsal, D., Tsutaki, S., … Fujita, K. (2015). The GAMDAM glacier inventory: A quality-controlled inventory of Asian glaciers. *The Cryosphere*, *9* (3), 849–864.

Pfeffer, W., Arendt, A., Bliss, A., Bolch, T., Cogley, J., Gardner, A., … Sharp, M.; The Randolph Consortium. (2014). The Randolph glacier inventory: A globally complete inventory of glaciers. *Journal of Glaciology*, *60*(221), 522–537.

Phan, V., Lindenbergh, R., & Menenti, M. (2013). Geometric dependency of Tibetan lakes on glacial runoff. *Hydrology and Earth System Sciences*, *17*(10), 4061.

Pritchard, H. D. (2017). Asia's glaciers are a regionally important buffer against drought. *Nature*, *545* (7653), 169.

Qiu, J. (2018). World meteorological organization: Scaling the peaks for social benefits, *National Science Review 5*, 947–952.

Qiu, Y., Lu, J., Shi, L., Xie, P., Liang, W., & Wang, X. (2019, March 29). Passive microwave remote sensing data of snow water equivalent in High Asia. *China Scientific Data*, *4*(1), 1–16.

Qiu, Y., Massimo, M., Li, X., Birendra, B., Joni, K., Narantuya, D., … Zhao, T. (2017). Observing and understanding High Mountain and cold regions using Big Earth Data. *Bulletin of Chinese Academy of Sciences*, *32*(Z1), 82–94.

Shean, D., Bhushan, S., Montesano, P., Rounce, D., Arendt, A., & Osmanoglu, B. (2019). A systematic, regional assessment of High Mountain Asia glacier mass balance. *Frontiers in Earth Science*, 7. doi:10.3389/feart.2019.00363

Song, C., Huang, B., Richards., K., Linghong, K., & Vu, H. (2014). Accelerated lake expansion on the Tibetan Plateau in the 2000s: Induced by glacial melting or other processes. *Water Resources Research*, *50*(4), 3170–3186.

Song, C., Sheng, Y., Ke, L., Nie, Y., & Wang, J. (2016). Glacial lake evolution in the southeastern Tibetan Plateau and the cause of rapid expansion of proglacial lakes linked to glacial-hydrogeomorphic processes. *Journal of Hydrology*, *540*, 504–514.

Sumin, W., & Hongshen, D. (1998). *China lakes record*. Beijing: Science Press Ltd. (in Chinese).

Wan, W., Xiao, P., Feng, X., Hui, L., Rong, H., Hong, T., & Li, M. (2014). Monitoring lake changes of Qinghai Tibetan Plateau over the past 30 years using satellite remote sensing data. *Science Bulletin*, *59*(10), 1021–1035.

Wang, X., Gong, P., Zhao, Y., Xu, Y., Cheng, X., Niu, Z., … Li, X. (2013). Water-level changes in China's large lakes determined from ICESat/GLAS data. *Remote Sensing of Environment*, *132*, 131–144.

Wang, X., Siegert, F., Zhou, A., & Franke, J. (2013). Glacier and glacial lake changes and their relationship in the context of climate change, central Tibetan Plateau 1972–2010. *Global and Planetary Change*, *111*, 246–257.

Xiao, C., Liu, S., Zhao, L., Wu, Q., Peiji, L., Chunzhen, L., … Pu, J. (2007). Observed changes of cryosphere in china over the second half of the 20th century: An overview. *Annals of Glaciology*, *46*(1), 382–390.

Yang, K., Lu, H., Yue, S., Zhang, G., Lei, Y., La, Z., & Wang, W. (2017). Quantifying recent precipitation change and predicting lake expansion in the inner Tibetan Plateau. *Climatic Change*, *147*(1–2), 149–163.

Yao, T., Li, Z., Yang, W., Guo, X., Zhu, L., Kang, S., … Yu, W. (2010). Glacial distribution and mass balance in the Yarlung Zangbo river and its influence on lakes. *Chinese Science Bulletin*, *55*(20), 2072–2078.

Yao, T., Yu, W., Wu, G., Xu, B., Yang, W., Zhao, H., … You, C. (2019). Glacier anomalies and relevant disaster risks on the Tibetan Plateau and surroundings. *Chinese Science Bulletin*, *64*(27), 2770–2782.

Yao, X., Liu, S., Li, L., Sun, M., Luo, J., & Feng, Y. (2013). Spatial-temporal variations of lake area in Hoh Xil region in the past 40 years. *Acta Geographica Sinica*. *68*(7), 886–896.

Yi, G., Deng, W., Li, A., & Zhang, T. (2015). Response of lakes to climate change in Xainza Basin Tibetan Plateau using multi-mission satellite data from 1976 to 2008. *Journal of Mountain Science*, *12*(3), 604–613.

Zhang, G., Chen, W., Li, G., Yang, W., Yi, S., & Luo, W. (2020). Lake water and glacier mass gains in the northwestern Tibetan Plateau observed from multi-sensor remote sensing data: Implication of an enhanced hydrological cycle. *Remote Sensing of Environment*, *237*, 111554.

Zhang, G., Luo, W., Chen, W., & Zheng, G. (2019). A robust but variable lake expansion on the Tibetan Plateau. *Science Bulletin*, *64*(18), 1306–1309.

Zhang, G., Yao, T., Piao, S., Bolch, T., Zhang, H., Chen, D., … Zhang, H. (2017b). Extensive and drastically different alpine lake changes on Asia's high plateaus during the past four decades. *Geophysical Research Letters*, *44*(1), 252–260.

Zhang, G., Yao, T., Shum, C., Yi, S., Yang, K., Xie, H., … Yu, J. (2017a). Lake volume and groundwater storage variations in Tibetan Plateau's endorheic basin. *Geophysical Research Letters*, *44*, 5550–5560.

Zhang, G., Yao, T., Xie, H., Wang, W., & Yang, W. (2015). An inventory of glacial lakes in the Third Pole region and their changes in response to global warming. *Global and Planetary Change*, *131*, 148–157.

Zhang, Y., Li, B., & Zheng, D. (2002). A discussion on the boundary and area of the Tibetan Plateau in China. *Geographical Research*, *21*(1), 2–9.

Zhao, Y. S. (2013). *Principles and Methods of Remote Sensing Application Analysis* (2nd ed., pp. 155–158). Beijing, China: Science Press.

Zhu, L., Ju, J., Qiao, B., Yang, R., Liu, C., & Han, B. (2019). Recent lake changes of the Asia Water Tower and their climate response: Progress, problems and prospects. *Chinese Science Bulletin*, *64*, 2796–2806.

Zhu, L., Xie, M., & Wu, Y. (2010). Quantitative analysis of lake area variations and the influence factors from 1971 to 2004 in the Nam Co Basin of the Tibetan Plateau. *Chinese Science Bulletin*, *55*, 1294–1303.