100 years of lake evolution over the Qinghai-Tibet Plateau

Guoqing Zhang[1], Youhua Ran[2], Wei Wan[3], Wei Luo[1,4], Wenfeng Chen[1,5], Fenglin Xu[1,5]

[1] State Key Laboratory of Tibetan Plateau Earth System Science, Institute of Tibetan Plateau Research, Chinese

5   Academy of Sciences, Beijing, China

[2] Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou, China

[3] Institute of Remote Sensing and GIS, School of Earth and Space Sciences, Peking University, Beijing, China

[4] Natural Resources and Planning Bureau, Qujing, Yunnan, China

[5] University of Chinese Academy of Sciences, Beijing, China

Correspondences: Guoqing Zhang (guoqing.zhang@itpcas.ac.cn) and Youhua Ran (ranyh@lzb.ac.cn)

15 **Abstract**: Lakes can be effective indicators of climate change, and this is especially so for the lakes over the Qinghai-Tibet Plateau (QTP), the highest plateau in the world, which undergo little direct human influence. The QTP has warmed at twice as the mean global rate, and the lakes there respond rapidly to climate and cryosphere changes. The QTP has ~1200 lakes larger than 1 km^2 with a total area of ~46000 km^2, accounting for approximately half the number and area of lakes in China. The lakes over the QTP have been selected as an

20 essential example for global lakes or water bodies studies. However, concerning lake data over the QTP are limited to the Landsat era and/or available at sparse intervals. Here, we extend the record to provide the comprehensive lake evolution data sets covering the past 100 years (from 1920 to 2020). Lake mapping in 1920 was derived from an early map of the Republic of China, and in 1960 from the topographic map of China. The densest lake inventories produced so far between 1970 and 2020 (covering all lakes larger than 1 km^2 in 14

25 epochs) are mapped from Landsat MSS, TM, ETM+ and OLI images. The lake evolution shows remarkable transitions between four phases: significant shrinkage in 1920−1995, rapid linear increase in 1995−2010, relative stability in 2010−2015, and further increase in 2015−2020. The spatial pattern indicates that the majority of lakes shrank in 1920−1995, and expanded in 1995−2020, with a dominant enlargement for central-north lakes in contrast to contraction for southern lakes in 1976−2020. The time series of precipitation between

30 1920 and 2017 indirectly supports the evolution trends of lakes identified in this study. The lake data set is freely available at http://doi.org/10.5281/zenodo.4678104 (Zhang et al., 2021).

Keywords: lake, map of Republic of China, topographic map, Landsat, remote sensing, Qinghai-Tibet Plateau

1 Introduction

35 The Qinghai-Tibet Plateau (QTP) is the highest plateau in the world, and has a large number of lakes

widely distributed across it. The QTP is sensitive to climate change: between 1970 and 2018 it warmed faster

than other continental areas, with a warming rate of ~0.36 °C/decade compared to the global mean of

~0.19 °C/decade (GISTEMP-Team, 2019). The warmer climate in the region during the last half century

(Kuang and Jiao, 2016) has induced dramatic changes in both the hydrosphere and the cryosphere (Chen et al.,

40 2015). Cryospheric melting reflects in accelerated glacier retreat and mass loss (Shean et al., 2020; Yao et al.,

2012), a lower snowline (Shu et al., 2021), and degradation of permafrost (Ran et al., 2018), which feeds

crucial water to alpine lakes. Human effects on these alpine lakes' evolution can be considered to be negligible,

due to their remoteness and the harsh weather conditions. These lakes respond rapidly to climate and

cryosphere changes, as most of them are located in closed watersheds (basins), and they have been selected as

45 typical examples in studies of global water bodies (Pekel et al., 2016) and lakes (Woolway et al., 2020).

Changes in lake area can have important influences on terrestrial ecosystems and climate change due to

the change between water and land. Knowledge of lake changes over the QTP has been greatly improved by

the application of remote sensing techniques which allow changes in lake area, level and volume to be derived

from satellite data. Lake mapping and the determination of changes in lake number and area (Sun et al., 2018;

50 Zhang et al., 2020b) are the most extensively researched applications of these techniques on the QTP. Lake

mapping from satellite data can acquire data for lakes with an extensive range of areas (greater than 4 pixels in

size) and for long monitoring periods (Ma et al., 2010; Wan et al., 2014).

Several studies have evaluated the potential and accuracy of maps of the Republic of China (from 1912 to

1949). For example, Han et al. (2016) compared the changes of rocky desertification in Guangxi province from

55 a 1930s topographic map (with a scale of 1:100 000) and Landsat TM in ~2000. Kong (2011) evaluated the

accuracy (offset) of the early Republic of China map of Henan Province (1:100 000). Su et al. (2018) provided

a lake and wetland data set for Xinjiang of late Qing and Republican China (~1909 and 1935). Yu et al. (2020)

described spatial-temporal evolution of Dongting lake in Hunan Province, southeastern China since the late

Qing Dynasty using a topographical map from the Republican period. All of these studies confirm the value of
60   the early maps in geoscience science, as a historical and rare archive. However, no studies have yet reported

lake mapping for the remote QTP in the Republic of China.

Changes in lake number and area have been considered in several studies: 1) Lake inventories from the

1960s topographic mapping (Wan et al., 2014), Landsat images in 2010 (Zhang et al., 2019), and Chinese

Gaofen-1 satellite data in 2014/2015 (Wan et al., 2016; Zhang et al., 2017b); 2) changes in lake number and

65   area between two or multi-phases (Zhang et al., 2019); 3) changes in lake area for some selected large lakes

with continuous satellite observations (Lei et al., 2013; Wu et al., 2017; Zhang et al., 2020b); 4) changes in area

for dominant lake distribution regions, such as the Inner Plateau (Qiangtang Basin), between periods such as

1970s−2011 (Song et al., 2013) and 2009−2014 (Yang et al., 2017). Global surface water bodies have been

mapped by the Google Earth Engine (GEE) platform (Pekel et al., 2016). However, the lakes in that study are

70   defined as permanent water surfaces (i.e. those which persist throughout the year) (Pekel et al., 2016). A

detailed classification of water bodies (including lakes, reservoirs and rivers) and precise mapping of their

number and area are important for discovering surface water characteristics and their changes. We emphasize

that lakes should be differentiated carefully from other water body types, such as rivers and reservoirs, when

examining their changes (Zhang et al., 2020a), and they should be mapped in a relatively stable season (Sheng

75   et al., 2016). Therefore, lakes cannot be extracted directly from this global water body data set.

Here, we provide the most comprehensive lake mapping yet produced over the QTP covering the past 100

years (from 1920 to 2020). The new features of this data set are: 1) its temporal length - it provides the longest

period of lake observations from maps; 2) it provides a state-of-the-art lake inventory for the Landsat era (from

the 1970s to 2020); 3) it provides the densest lake observations for lakes with areas larger than 1 km^2.

80

2 Study area

The QTP, with an area of ~200×10^4 km^2, consists of Tibet Autonomous Region and Qinghai province, and

has the border of China as its southern boundary (Figure 1). The region is sometimes referred to as the Tibetan

Plateau (TP), but this is defined here slightly differently as the area with an altitude above 2500 m a.s.l., which

85   also includes the glaciers in the Himalaya, Karakoram, and Hindu Kush (Figure S1) and has an area of

~300×10^4 km^2. For the creation of the lake inventory from the historical maps, we use the QTP, rather than the

TP, as the map of the early Republic of China and the topographic map of China do not include areas outside
China. For the post-1970s Landsat era, we provide lake mapping for both the QTP and the TP as the Landsat
images provide global coverage.

90   There are 70 China Meteorological Administration (CMA) weather stations in the QTP. The data from
these stations reveal that the annual warming rate between 1980 and 2018 was 0.05 ± 0.01 ºC/yr ($P<0.0001$)
(Figure S2). The annual precipitation also shows an increasing trend, especially after 1998 (an increase of ~7%
in 1999−2018 relative to 1980−1998). The QTP includes 87% of the lakes of the TP and 92% of their area. The
number of lakes larger than 1 km^2 is ~1200, and they have a total area of ~46000 km^2 (Figure 1). The glacier

95   area in this region, obtained from the second Chinese glacier inventory (Guo et al., 2015), is ~3×10^4 km^2. The
area of clear glacier shrank by 0.08%/yr in Inner Tibet between 1990 and 2018 (Huang et al., 2021).

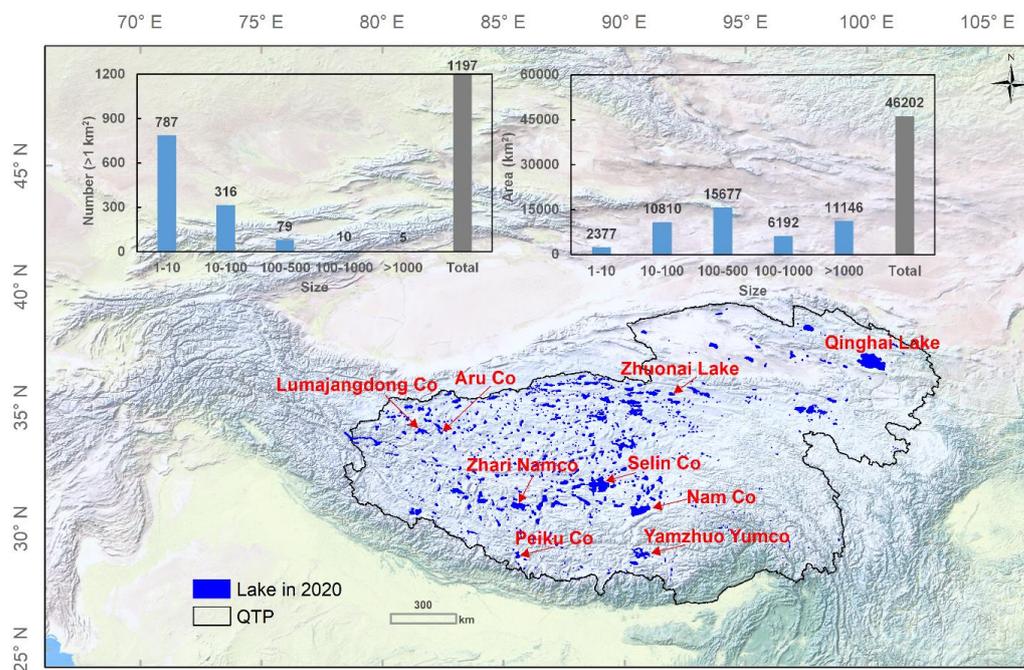



Figure 1. Distribution of lakes over the QTP. Insets show, for lakes larger than 1 km^2, the number and area of lakes
in different size ranges and the total number and area (status: 2020). The names refer to the lakes presented as
100   examples in Figure 6. The background of this figure is from Natural Earth at https://www.naturalearthdata.com.

3 Data and methods

Three different data sources were used: the early map of the Republic of China in ~1920 (hereafter 1920 used), a digitalized topographic map of China from the ~1960s (hereafter 1960), and Landsat

105 MSS/TM/ETM+/OLI images from 1970−2020. The lake mapping process for the period 1920 to 2020, shown in Figure 2, consists of three main steps: data collection, preprocessing, and water body classification.

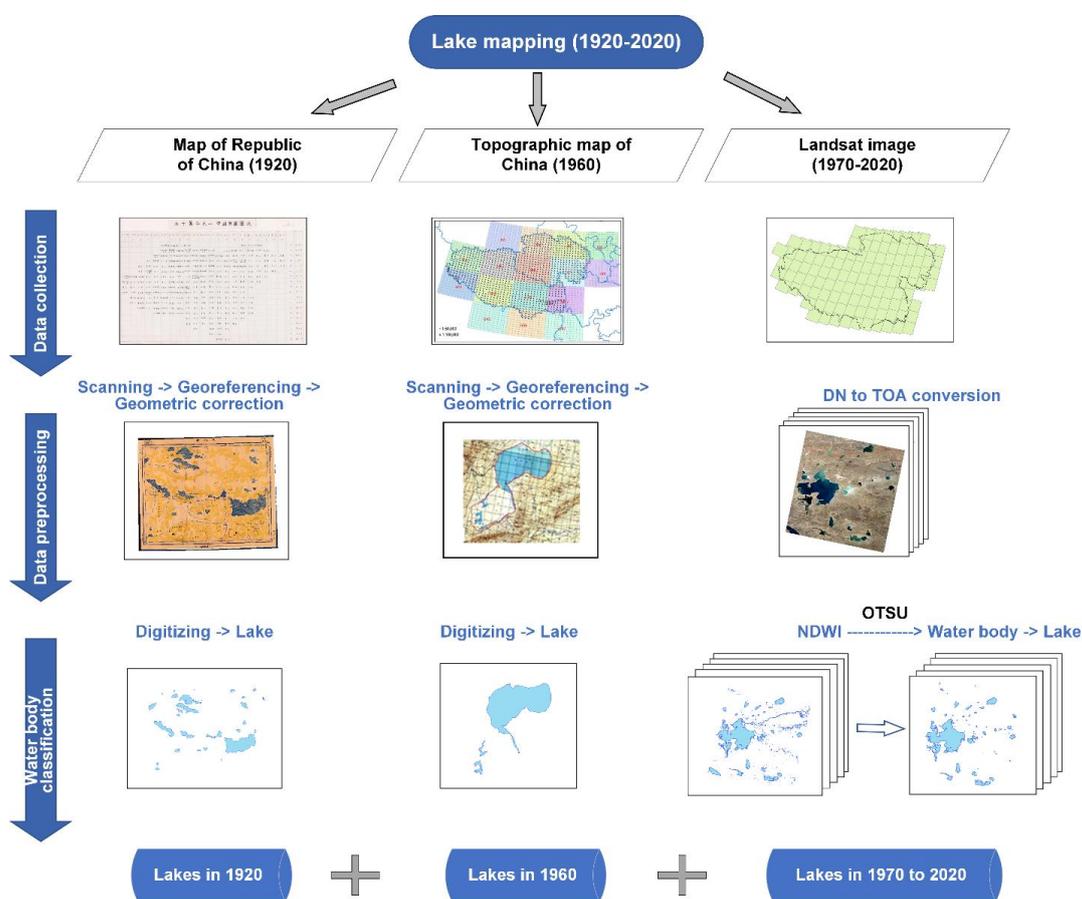



Figure 2. Flowchart showing the process for lake mapping of the QTP from 1920 to 2020. Three key steps: data collection, data preprocessing, and water body classification for maps of the early Republic of China in 1920,

110 topographic map of China in 1960, and Landsat images in 1970−2020 are shown. The distribution of topographic maps over the QTP is from Wan et al. (2014).

3.1 Lake mapping in 1920

Maps of the Republic of China surveyed and drawn by Beijing Survey Bureau, Staff Cartographic Bureau,

115 and Guangdong Survey Bureau in about 1914−1916 with a scale of 1:500 000

(http://www.ccartoa.org.tw/news/2018/180901.html) were used (Table 1). This is the most accurate

topographic map with the most complete coverage available from the early 20th century. About 50 maps,

covering the entire QTP, were obtained from Academia Sinica (https://www.sinica.edu.tw/ch). Each map was

scanned and georeferenced. A geometric correction was also applied to remove distortion introduced by the

120 scanning process. Manual digitizing processes were used to vectorize water bodies from the raster maps. 760

lake water body units were identified, and annotated with names (old names from the early map and current

Chinese names).

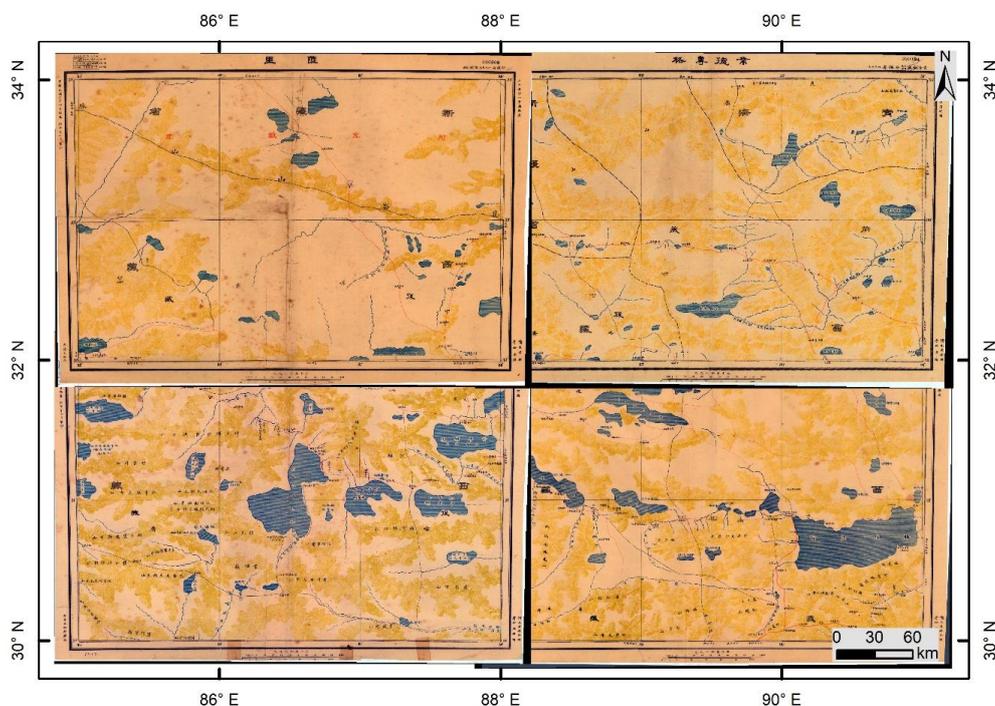



Figure 3. An example showing lake distribution in 1920 on maps of the Republic of China with georeferencing.

125

3.2 Lake mapping in 1960

The first Chinese lake inventory by field investigation and topographic maps created from aerial

photographs surveyed mainly in the 1960s (with a scale of 1:250 000) were obtained (Table 1) (Wan et al.,

2016; Wang and Dou, 1998). All maps were georeferenced to an Albers Equivalent Conical Projection, with an

130   Root Mean Square (RMS) error of <1.5 pixel (Ma et al., 2010). Visual interpretation and manual digitization

were chosen to vectorize lake boundaries. Salt crusts, as found around the salt lakes in the northeastern QTP,

were not considered to be a component of the lakes. All lakes were coded along with other attributes (names

and area).

135   **Table 1.** Maps (images) and methods used for lake mapping from 1920 to 2020.

Map	Sense	Period	Resolution	Number	Method for lake mapping
Early Republic of China	-	1920s	1:500,000 or ~130 m	~50	Manual digitization
Topographic map of China	-	1960s	1:250,000 or ~75 m	-	Manual digitization
Landsat	MSS	1970s	~60 m	93	Manual digitization
	TM, ETM+, OLI	1990−2020	30 m	1384	NDWI + Otsu

3.3 Lake mapping in 1972−2020

Remotely sensed images at a high spatial-temporal resolution from the Landsat series of satellites provide

the longest data set (~50 years) of global lake observations (Wulder et al., 2019). The initial Landsat-1

140   Multispectral Scanner System (MSS) launched in 1972 provides land surface observations with a resampled

resolution of ~60 m. The Thematic Mapper (TM) on Landsat-4 (launched in 1982) and -5 (launched in 1984)

has a spatial resolution of 30 m for the six reflective bands, which offers a significant advance in spatial,

spectral, and geometric performance (Chander et al., 2009). The Enhanced Thematic Mapper Plus (ETM+)

onboard Landsat-7 launched in 1999 has the same spatial resolution as the TM sensor. Unfortunately, the scan-

145   line corrector (SLC) of the Landsat 7 ETM+ sensor failed in 2003, resulting in ~22% pixels missing from each

scene (Chen et al., 2011). The Operational Land Imager (OLI) on Landsat-8 which launched in 2013 included a

significant improvement of image radiometric quality providing better surface water mapping (Masek et al., 2020; Roy et al., 2014; Wulder et al., 2019). A policy of open and free access to archived and new Landsat images operated by the United States Geological Survey (USGS) since 2008 has led to a substantial increase in

150 Landsat applications (Zhu et al., 2019). However, the generation of temporally-continuous lake mapping over the QTP can still be a challenge due to the contamination of these optical images by high amounts of cloud cover and snow (Yu et al., 2016).

Landsat Level 1 terrain-corrected (L1T) products from the MSS/TM/ETM+/OLI sensors between 1972 and 2020 were downloaded from https://glovis.usgs.gov. This product is defined as the Universal Transverse

155 Mercator (UTM) projection with World Geodetic System 84 (WGS84) datum. Radiometric calibration and a systematic geometric correction have been applied to Landsat L1T products before their release.

It is not possible to map lakes at an annual interval over the QTP due to the high levels of cloud contamination of the Landsat images and because of seasonal fluctuations. Here, our purpose is to cover all lakes (with area larger than 1 km^2) in each epoch. Lakes of area less than 1 km^2 are also selected for water body

160 classification so that no lakes are missed. The images in seasons with the largest and relatively stable lake area are preferentially chosen (Sheng et al., 2016). An evaluation of the seasonal lake area cycle, and the presence of relatively low levels of cloud coverage, identifies October as the optimal season over the QTP (Zhang et al., 2020c). If there are no images available for October, we widen the search to September and November, and so on. This process leads to the selection of some images from the winter months, but the number is small and

165 none contain lake ice. This time-window extension is required particularly in the early Landsat period (before 1990) for which there are few available images and data quality is low. For lake mapping in a specific target year, the data in the optimal season for this year are first selected, and then, if not available, have a slight extension in a preceding year. If there is still not enough data, images from the subsequent year are also used. With this procedure, a total of 1477 images between 1972 and 2020 were used to provide 14 epochs of lake

170 mapping (for lakes >1 km^2) (Table 1). We use 1970 for the 1970s, 1990 for the 1990s, 1995 for the period 1994−1996, 2000 for the 2000s, 2005 for the period 2003−2006, and 2010 for the 2010s to better present the spatial-temporal changes in the lakes and to match with the time series of climate data.

A summary of the images used for each epoch was presented in Figure 4, which shows the year and month
in which the images were acquired. In the 1970s, 93 images from Landsat MSS between 1972 and 1977 were
175   selected, with most of the images from 1976. In the 1980s, there are few available Landsat data over the QTP
and no alternative data sources, so there is a gap in the data at this time. For lake mapping in the 1990s, data
from 1990 were the first choice. Data from the preceding period, 1987−1989, and the subsequent year, 1991,
were also selected to reduce seasonal anomalies. This choice resulted in 99 images spanning five years. For
lake mapping in 1995, data from 1995 were the first choice for selection, but to meet the requirements the data
180   selection period had to be extended to 1994 and 1996, finally allowing 97 images to be chosen. For the 2000s,
91 images, predominantly from 2000 but with a small extension to the preceding years 1998 and 1999, were
selected. For the 2005 lake mapping, 93 images were collected including 64 (69%) in 2005, 10 in 2003, 18 in
2004, and an additional scene in 2006. For the 2010 lake mapping, most of the data (75%) was from 2010, with
an additional 22 images (25%) from 2008 and 2009. After 2013, Landsat 8 provides more available data
185   enabling annual lake mapping to be achieved between 2013 and 2020. For this period about 120 Landsat
images in each year are required to cover the whole QTP. The small variations in this number from year to year
are due to cloud coverage over some lake surfaces which leads to more data having to be employed.

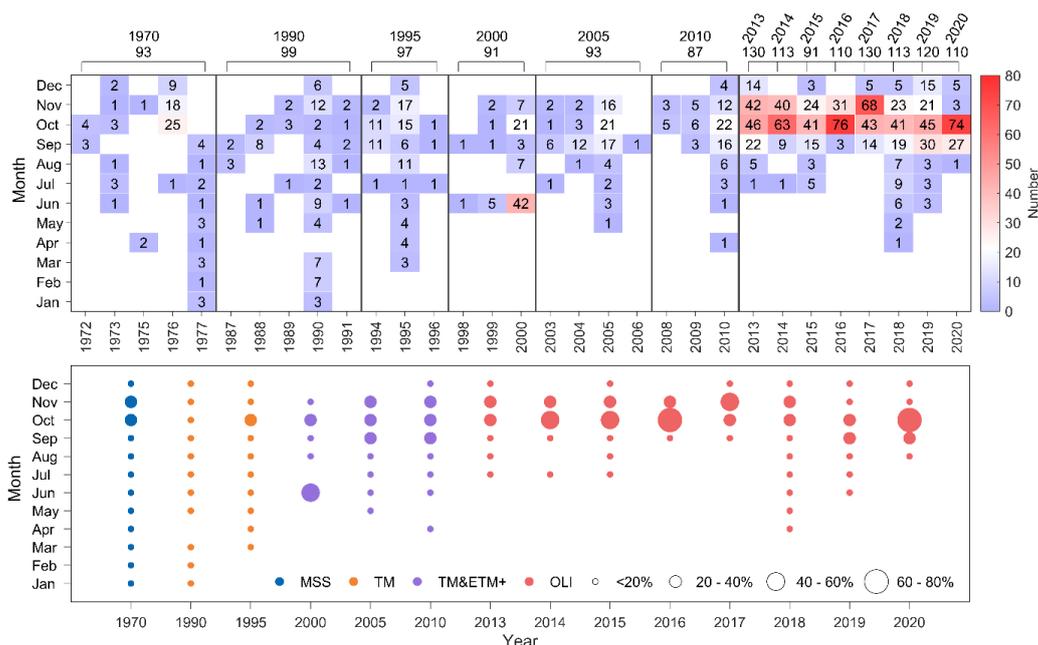



Figure 4. Landsat images used for lake mapping from 1970 to 2020. a) Data span in different years for each epoch,

190   and number of months of data acquired. b) Images used from different sensors, and percentage acquired in each

month.

Visual interpretation was used for lake boundary delineation in 1970 due to the low quality of the MSS

data. From 1990 onwards, a semi-automated water-body classification method was used to distinguish water

195   from non-water features. First, the Digital Numbers (DNs) of raw Landsat TM/ETM+/OLI images are

converted to Top-Of-Atmosphere (TOA) reflectance (Chander et al., 2009), which is often used to generate the

normalized difference water index (NDWI), a binary water / non-water classification (Li and Sheng, 2012;

McFeeters, 2013). Second, an automated water extraction process is executed (Zhang et al., 2017b). Initially, a

small threshold, such as "0", was used to separate water from non-water features. An outward buffer zone is

200   then created for each water body, and an optimal threshold was determined by the Otsu method (Otsu, 1979).

The water / non-water classification is performed again using the optimized threshold to generate the water

body products. Finally, the lake boundaries are extracted, visually compared with the original Landsat images, and manually edited if any mismatches are found.

205 **3.4 Validation of lake mapping**

A combination of direct and indirect validation was used to assess the accuracy of lake boundaries determined from the three different data sources. The lakes determined from the 1920 map of the Republic of China and the 1960 topographic map of China cannot be validated directly as there are no records of field measurements of lake boundaries for the relevant periods. However, for these, we can indirectly compare with

210 time series of precipitation changes (see Section 5.2) as the plateau lakes are dominantly driven by variations in precipitation (>70% contribution to lake water balance) (Biskop et al., 2016; Zhang et al., 2020b). Moreover, the veracity of the 1960 topographic maps has been demonstrated by previous studies (Ma et al., 2010; Wan et al., 2016; Wang and Dou, 1998).

Landsat images have been widely used for water body classification and lake water-extent mapping

215 globally (Pekel et al., 2016; Pickens et al., 2020; Verpoorter et al., 2014; Yamazaki et al., 2015). Here, we further confirm the accuracy of the water classification algorithm both directly and indirectly. High-precision GPS measurements of the boundaries of Aru Co in the northwestern plateau were made on July 28, 2017 (Figure 5). The Landsat-8 image for July 29, 2017 was acquired for the purpose of comparison with these measurements, and Figure 5b shows that the lake boundary delineated by the method of this study matches well

220 with the data from the GPS survey. In addition, we compared the area extracted for Qinghai Lake with the lake level recorded by a gauging station between 1976 and 2019. A well-established relationship (r^2=0.95) between lake area and water level was exhibited (Figure 5d), indirectly suggesting that our algorithm for lake boundary determination provides a true reflection of lake state.

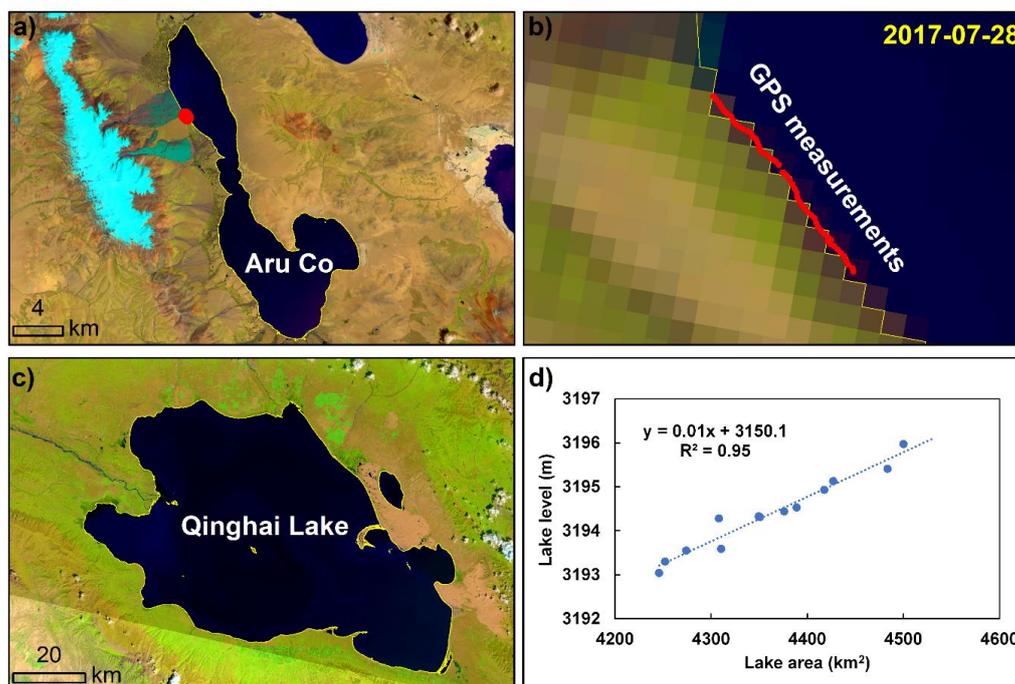



Figure 5. Validation of delineated lake boundaries from GPS measurements and comparison with lake level, with background of Landsat images. a) Location of GPS measurements points for Aru Co. b) Distribution of GPS measurements points along the water boundary of Aru Co. c) Boundary of Qinghai Lake from Landsat image in 2020. d) Comparison between lake area and water level from gauging station for Qinghai Lake between 1976 and 2019. Locations of Aru Co and Qinghai Lake on the QTP are shown in Figure 1.

3.5 Meteorological data

Annual temperature and precipitation data (~1 km resolution) between 1920 and 2017 over the QTP were derived from Peng et al. (2019). This data set is spatially downscaled from the Climatic Research Unit (CRU) data, and shows good agreement with CMA weather stations across China between 1951 and 2016. In addition, changes in annual precipitation between 1980 and 2018, determined from 70 CMA weather stations, were also utilized to indirectly compare with lake area variations (Figure S2).

4 Results

4.1 Past and present lake states

240   Lakes larger than 1 km^2 in 1920, 1960, and 1972−2020 were extracted. For 1920, a total of 710 lakes were

identified, with areas ranging from 1.06 to 4303.40 km^2, and a total area of 44232.12 km^2. The number of lakes

larger than 1 km^2 (from 1.0 to 4284.13 km^2) increased to 1107 in 1960, but their total area decreased to

38314.51 km^2, a 13% drop relative to the area in 1920. By 2020, the number of lakes larger than 1 km^2

increased to 1197 with a total area of 46201.62 km^2 (21% higher than the area in 1960). Out of the ~1200 lakes

245   in 2020, 787 lakes (2377 km^2 in total area, 5%) are within the size range of 1−10 km^2 (inset of Figure 1). Out of

410 lakes with areas greater than 10 km^2, 5 are larger than 1000 km^2 (11146 km^2 in total area, 24%). In total,

the number of lakes increased by ~69% and the area by ~4% between 1920 and 2020. The great increase in

number may be attributed to the appearance of new lakes such as those detected by multi-period satellite

images (Ma et al., 2010; Zhang et al., 2020b).

250   Individual lakes exhibit different patterns of evolution. Eight large lakes distributed in different climatic

regions (Figure 1) have been selected to display their distinguishing characteristics (Figure 6a). Qinghai Lake,

the largest lake on the QTP, had a relatively stable area before 2010 (except for a fluctuation in 1976), but a

continuous rapid expansion in the most recent decade. Selin Co, the largest lake in Tibet, exhibits a robust

expansion during the past 100 years, especially after 2000. In the Landsat era, the area expanded by ~40%

255   (Figure 6b). This study extended the record to early 1920 and found a triple increase in area has occurred. Nam

Co, the second largest lake in Tibet, shrank in area before 1960, and then remained in a relatively stable state

until 2020. Lumajangdong Co, in the northwestern plateau, expanded monotonically between 1920 and 2020

with an overall ~55% increase in area. The graph for Zhari Namco in the central plateau reveals a rapid

expansion before 1960, followed by a stable state, or slight increase (~6%), between 1960 and 2020. Zhuonai

260   Lake in the northeastern plateau expanded until 1960, and then remained in a stable state between 1960 and

2010. However, an outburst flood event occurred on September 15, 2011 (Liu et al., 2016), resulting in a rapid

lake water retreat as indicated by the area shrinkage between 2013 and 2020 (Figure 6b). Two large lakes in the

southern plateau, Paiku Co and Yamzhuo Yumco, have undergone a slight lake level decline according to

altimetry data or observations of area shrinkage from Landsat images (Li et al., 2019; Phan et al., 2012; Zhang

265   et al., 2020b). Here, it is shown that the shrinkage in area for these two lakes can be traced back to 1920, with

decreases in area of ~22% for Paiku Co and ~49% for Yamzhuo Yumco, found between 1920 and 2020.

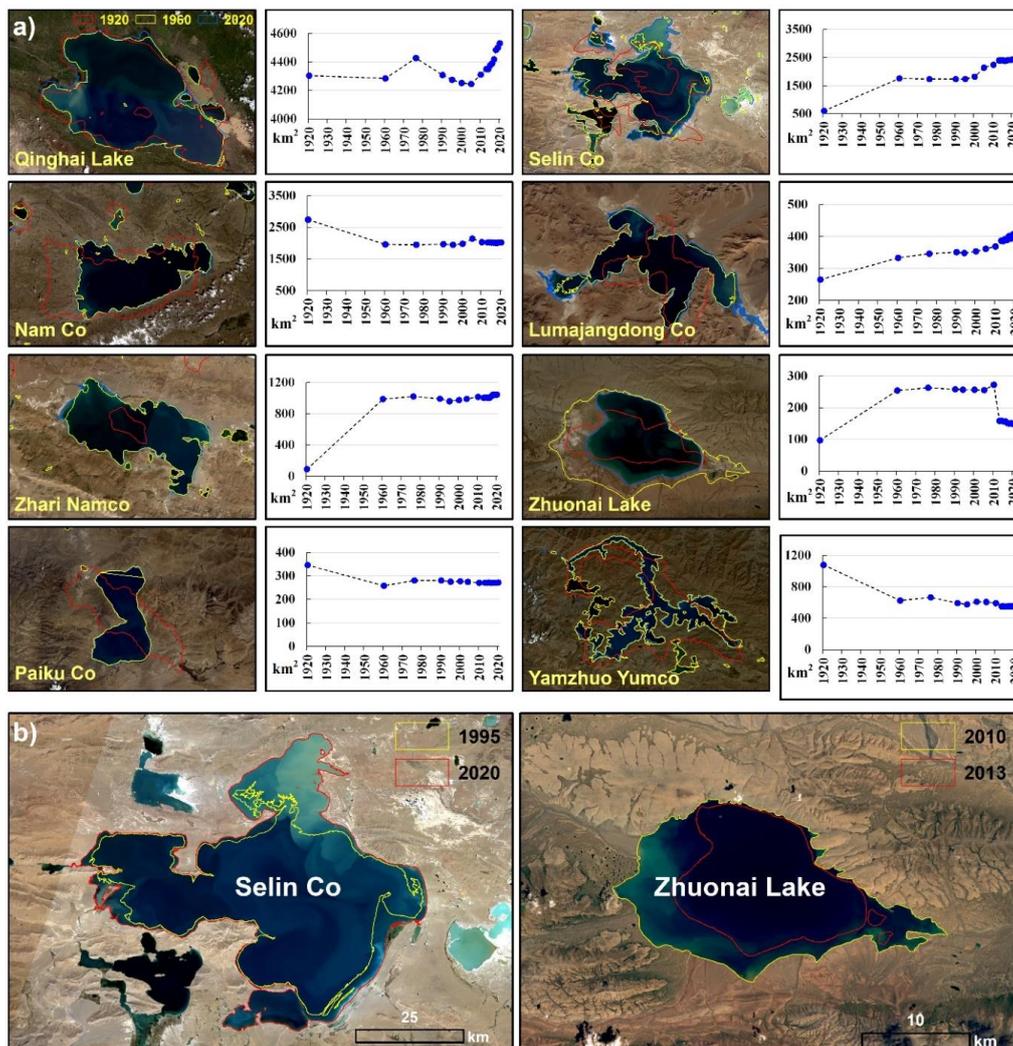



Figure 6. Examples showing the evolution of eight selected lakes from 1920 to 2020, with background of Landsat

images (Bands 3, 2, 1 composite). a) Outlines of eight selected lakes in 1920, 1960, and 2020, and time series of lake

270   area from 1920 to 2020. b) Images for Selin Co and Zhuonai Lake as examples of rapid lake expansion and rapid

shrinkage, respectively. The locations of these lakes are shown in Figure 1.

4.2 Time series of lake changes between 1920 and 2020

Here, the long-term evolution of lakes with areas larger than 1 km^2 between 1920 and 2020 over the QTP

275 are described. Over the period 1920−1995, total lake area showed a remarkable 13% reduction, from 40779.33

in 1920 to 35308.31 km^2 in 1995 (Figure 7a). Between 1995 and 2010, lake area increased rapidly in a linear

fashion, almost recovering to its initial 1920 value by 2005 and reaching a value of 43194.82 km^2 in 2010

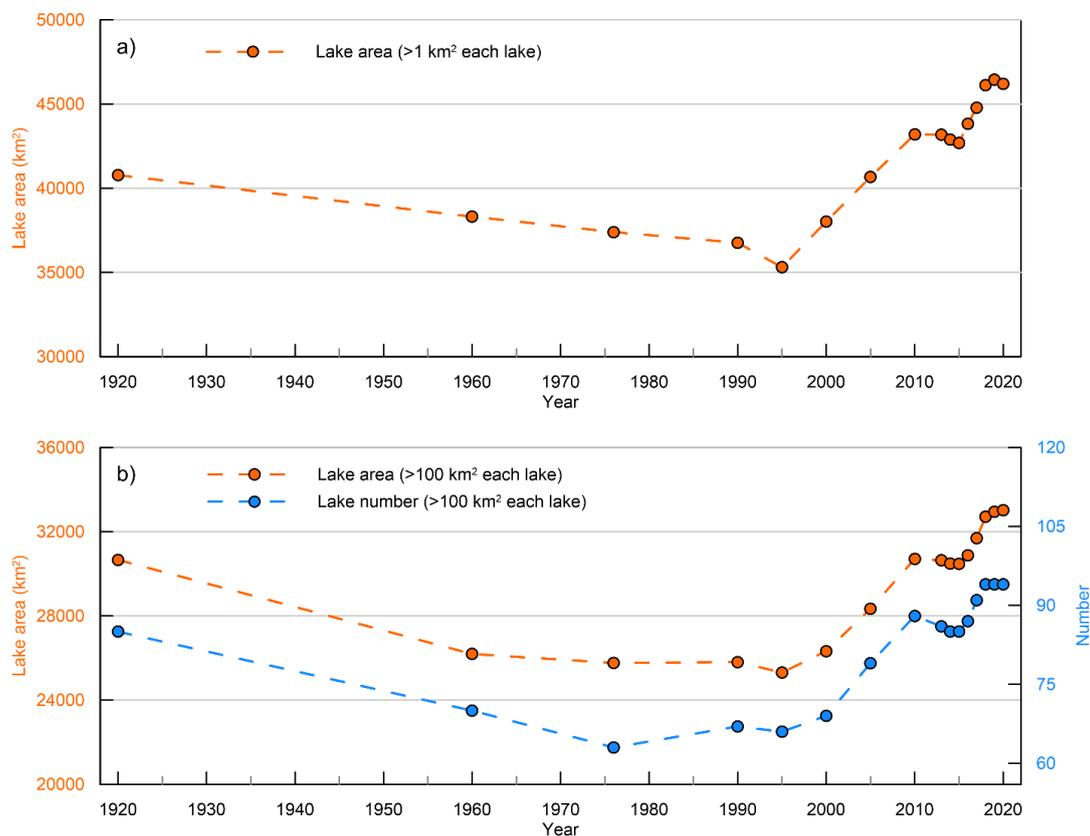
(+22%). However, a more stable period with a slight shrinkage occurred between 2010 and 2015. The most

recent five years (2015−2020) were another period of rapid expansion in lake area.

280 Since the number and area of lakes of small size are variable, we also examined the trends in number and

area of large lakes of more than 100 km^2 in area. The patterns of evolution of number and area for these large

lakes (Figure 7b) are consistent with the total area of all lakes greater than 1 km^2, supporting the several

different lake evolution phases during the past 100 years identified above.

285 **Figure 7.** Changes in number and area of lakes from 1920 to 2020. a) Total area of all lakes greater than 1 km². b) Total area and number of lakes greater than 100 km² to exclude the effects of variability in the number and area of small lakes (1−100 km²).

4.3 Spatial patterns of lake changes between 1920 and 2020

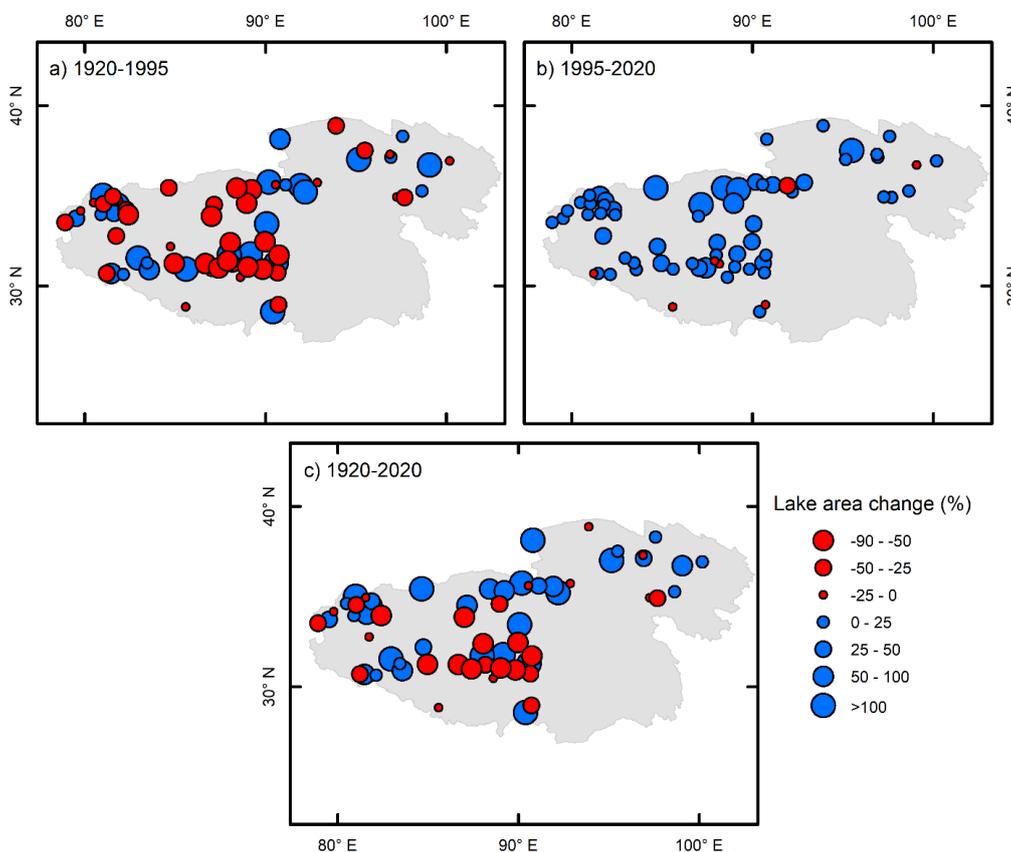
290   The spatial patterns of lake changes for three different periods are presented in Figure 8. Between 1920 and 1995, most of the lakes throughout the plateau shrank in area. However, almost all of the lakes increased in area between 1995 and 2020. Over the entire study period, 1920−2020, the spatial pattern is heterogeneous, with the majority of shrinking lakes in the central-south plateau, but most of the enlarging lakes in the northern plateau.

295

Figure 8. Spatial patterns of lake area changes during different evolution stages. a) 1920−1995. b) 1995−2020. c) 1920−2020.

The spatial patterns of lake area changes during the Landsat era are further examined (Figure 9) as

300   Landsat images are used popularly, and the past five decades cover the period with the most significant lake changes. Between 1970 and 1995, the majority of lakes decreased in area. However, this pattern was reversed during 1995−2010 when the vast majority of lakes grew in area. In the most recent stage, 2010−2020, most of lakes were still enlarging, although there were some contracting lakes in the central plateau. Over the entire period of 1970−2020, the predominant lake behavior was expansion, but some, mainly southern, lakes shrank.

305   One such exception is Zhuonai Lake (Figure 9c−d) which underwent a noticeable reduction in area, which is in agreement with evolutionary process observations (Figure 6a).

In general, the shrinkage of lake area in 1920−1995 is reflected in the spatial patterns. The comprehensive lake expansion in 1995−2010 is also in agreement with the evolution revealed by the time series. The overall increase, albeit with variability in the 2010−2020 period, is visible in the spatial differences, with a clearer

310   contrast through Landsat era (1970−2020).

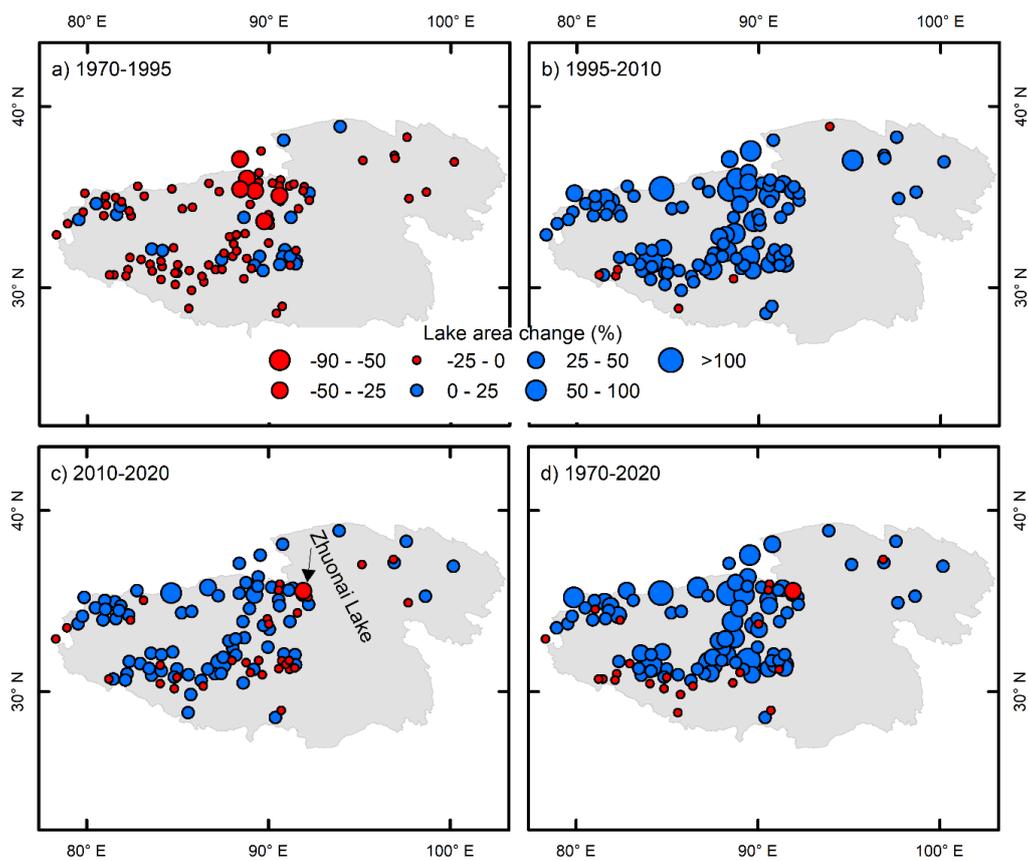



Figure 9. Spatial patterns of lake area changes during the Landsat era. a) 1970−1995. b) 1995−2010. c) 2010−2020. d) 1970−2020.

5 Discussion

5.1 Uncertainties in lake mapping

The uncertainties in lake mapping from topographic maps and Landsat images have usually been estimated by half a pixel inside and outside the delineated lake boundary (Fujita et al., 2009; Salerno et al., 2012). However, for the 1920 lake mapping, such an uncertainty estimate is inappropriate because the process and principle of the map creation are not known. In addition, there are offsets of the maps for some lakes, as shown in Figure 6a, although a geometric correction has been conducted before delineation. In any case, such offsets should not have an impact on the calculation of lake area.

Having three diverse data sources with different spatial resolutions (Table 1) could lead to differences in lake area estimates. However, a long-term (100-year) lake evolution is examined, and lakes over the QTP have

325   experienced a remarkable change. The inconsistency in spatial resolution between the data sets seems to have had little effect on altering the trend of lake evolution. The methods used for delineating lake boundaries are also different for the different data sources. Visual interpretation and manual digitization were used for the 1920 and 1960 maps, and for the 1970 Landsat MSS data. The results were cross checked to ensure a high-quality output of lake boundaries. For Landsat TM/ETM+/OLI, an automated water classification with the

330   optimal threshold for NDWI images is used, with the purpose of providing highly efficient and consistent water extraction. We have examined the output from this process by comparing it with false color composites of the original Landsat images to ensure lake boundaries are correct despite cloud and snow contamination of the water body surface. Inevitably, some Landsat images outside the optimal season have had to be selected, although at a several-year interval for each epoch (Figure 4). However, few data from these seasons are used,

335   and those images that are used contain few lakes. Similar conditions apply to the selection of ETM+ images when no suitable Landsat TM data are available. Again, the proportion of lakes covered by these ETM+ images is small compared to those covered by the Landsat TM data.

5.2 Causes of lake changes

340   Lake water balances for the Inner plateau as a whole (covering most of the lakes) (Zhang et al., 2017a) or for typical lake basins such as Selin Co (Zhou et al., 2015), Nam Co (Li et al., 2017), and Qinghai Lake (Zhao et al., 2017) have been quantitatively assessed. These studies reveal that increased precipitation, rather than glacier meltwater supply, has made the dominant contribution (>70%) to lake growth in recent decades. In addition, the geodetic mass balance also clearly shows that glacier mass loss contributes only a small fraction

345   (an average of ~10% for lake basins) to lake water volume gain between 2000 and ~2018 (Brun et al., 2020; Zhang et al., 2021). Based on this assumption, a comparison of time series of precipitation with lake area can reveal the dominant driver of lake evolution.

Comprehensive warming has occurred over the QTP during the past 100 years (Figure 10). The time series of temperature shows an increase followed by a decrease before ~1965 and then a warming trend until

350   the present day. The spatial pattern of precipitation change indicates a wetting trend everywhere except for the

southwestern and southeastern peripheries of the QTP. The time series of precipitation shows a drying trend

between 1920 and 1995, followed by a wetting between 1995 and 2010 but with variability after 2010. The

spatial-temporal patterns of precipitation changes, especially the time-series, match well with the evolution of

total lake area, suggesting that precipitation has mainly controlled lake evolution for the past 100 years.

355   Additionally, these data also give further weight to the argument that the lake development patterns before

1960 are reliable, although no field measurements or satellite observations are available to validate them.

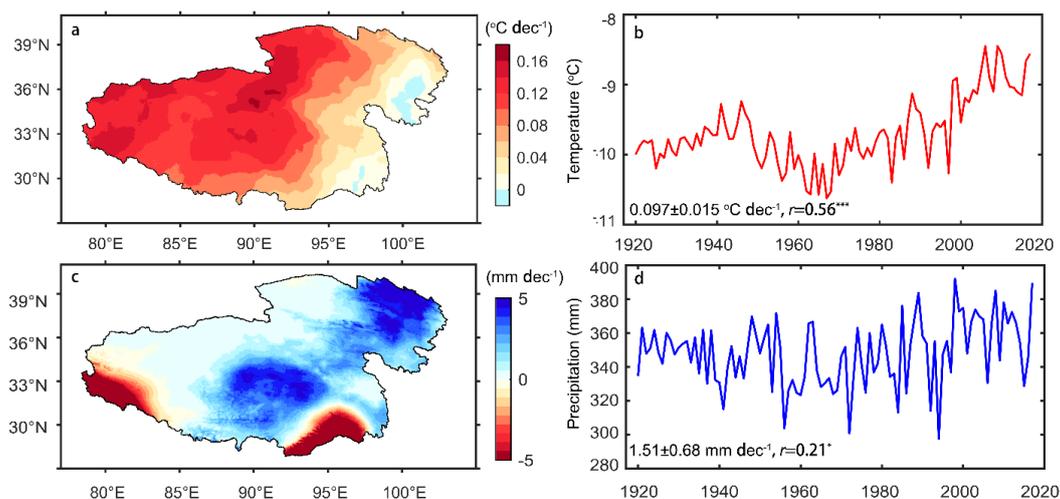



Figure 10. Spatial-temporal variations of temperature and precipitation from 1920 to 2017. a) Spatial pattern of

temperature change. b) Time series of temperature change. c) Spatial pattern of precipitation variation. d) Time series

360   of precipitation change.

6   Data availability

This lake data set includes the lake evolution during the past 100 years (from 1920 to 2020) over the QTP,

which is defined by the boundaries of Tibet Autonomous Region and Qinghai province, with an area of about

365   200×10^4 km^2. The 1920 and 1960 lake mapping are derived from manual digitizing of an early map of

Republic of China, and the topographic map of China, respectively. The continuous lake mappings from 1970

to 2020 are derived from Landsat MSS, TM, ETM+ and OLI images by a semi-automatic water classification.

All lake boundaries have been visually checked against original maps/images. All the analysis presented in this study is based on this lake data set.

370   We also provide the lake data set for the period 1970 to 2020 for the TP, the boundary of which is defined by the altitude of 2500 m, and which has an area of about 300×10^4 km^2. The boundary of the TP was first defined in this way by Zhang et al. (2013), and is popularly used. This lake data set, which includes more of the western plateau, is useful for more extensive analyses of climate and cryosphere variations.

Both lake data sets can be freely downloaded at http://doi.org/10.5281/zenodo.4678104 (Zhang et al.,

375   2021). The lake data sets are provided as ArcGIS Shapefiles, which are easy to read or reanalyze in GIS environments.

7   Conclusions

In this study, we have compiled a lake data set for the QTP spanning the early 20th century to the early

380   21th century with a relatively high spatial resolution from map observations. This is the longest, and the most comprehensive, lake mapping available for the QTP. The lake products generated, including lakes in 1920 from maps of the Republic of China, lakes in 1960 from the topographic map of China, and lakes for the period from 1970 through 2020 from 1477 Landsat images.

In 1920 there were 710 lakes larger than 1 km^2, with a total area of 44232.12 km^2. One hundred years

385   later, the number has increased to 1197 (~69%) with a total area of 46201.62 km^2 (~4%). The increase in numbers can probably be mainly attributed to the fact that lakes smaller than 1 km^2 have exceeded 1 km^2 over time and thus have been added to the count. Detailed lake evolution is also examined. The lakes underwent a decrease in area over the period 1920−1995, followed by a rapid increase in 1995−1995. They tended to stability in 2010−2015, and increased in area again in 2015−2020. The spatial patterns of overall shrinkage or

390   expansion are consistent with time series, but a pattern of enlargement for central-north lakes against contraction for southern lakes is revealed, which is especially clear between 1970 and 2020. The time series of precipitation is coincident with the trends of lake evolution.

This high-quality lake data set is a great asset for interdisciplinary climate, cryosphere, and hydrosphere studies, and can even form a bridge linking together earth system science over the world's highest plateau. We

395 will update the lake data set in the future when further data become available, for example with annual updates

with new Landsat images.

Supplement

The supplement related to this article is available online at: XXX

400

Author contributions: GZ designed the study and wrote the manuscript. YR provided the lake data set for the

Republic of China in 1920, and WW provided the lake data set from the topographic map of China in 1960. All

authors contributed to the writing and editing of this paper.

405 **Competing interests:** The authors declare that they have no conflict of interest.

Acknowledgements: The first author would like to thank former students or team members, Kexiang Zhang,

Fujing Zhu, Guoxiong Zheng, for data/codes accumulation. We also thank Hsiung-Ming Liao at Academia

Sinica for providing maps of the Republic of China.

410

Financial support: This study was supported by grants from the Natural Science Foundation of China

(41871056, 41831177), the Second Tibetan Plateau Scientific Expedition and Research (STEP) program

(2019QZKK0201), and the Strategic Priority Research Program (A) of the Chinese Academy of Sciences

(XDA20060201).

415

Review statement: This paper was edited by XXX and reviewed by XXX anonymous referees.

References

Biskop, S., Maussion, F., Krause, P., Fink, M., 2016. Differences in the water-balance components of four lakes in
420   the southern-central Tibetan Plateau. Hydrol. Earth Syst. Sci., 20(1): 209−225. doi:10.5194/hess-20-209-2016
Brun, F., Treichler, D., Shean, D., Immerzeel, W.W., 2020. Limited Contribution of Glacier Mass Loss to the Recent
        Increase in Tibetan Plateau Lake Volume. Frontiers in Earth Science, 8. doi:10.3389/feart.2020.582060

Chander, G., Markham, B.L., Helder, D.L., 2009. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. Remote Sens Environ, 113(5): 893−903. doi:10.1016/j.rse.2009.01.007

Chen, D. et al., 2015. Assessment of past, present and future environmental changes on the Tibetan Plateau. Chinese Sci Bull, 60(32): 3025−3035. doi:10.1360/N972014-01370

Chen, J., Zhu, X., Vogelmann, J.E., Gao, F., Jin, S., 2011. A simple and effective method for filling gaps in Landsat ETM+ SLC-off images. Remote Sens Environ, 115(4): 1053−1064. doi:10.1016/j.rse.2010.12.010

Fujita, K., Sakai, A., Nuimura, T., Yamaguchi, S., Sharma, R.R., 2009. Recent changes in Imja Glacial Lake and its damming moraine in the Nepal Himalaya revealed by in situ surveys and multi-temporal ASTER imagery. Environmental Research Letters, 4(4): 045205. doi:10.1088/1748-9326/4/4/045205

GISTEMP-Team, 2019. GISS Surface Temperature Analysis (GISTEMP), version 4. NASA Goddard Institute for Space Studies. Dataset accessed 2019-10-06 at https://data.giss.nasa.gov/gistemp/.

Guo, W. et al., 2015. The second Chinese glacier inventory: data, methods and results. J Glaciol, 61(226): 357−372. doi:10.3189/2015JoG14J209

Han, Z., Ran, Y., Liu, J., Li, J., 2016. The changing distribution of rocky desertification in the Guangxi Region, 1930s to 2000. Acta Geographica Sinica, 71(3): 390-399.

Huang, L., Li, Z., Zhou, J.M., Zhang, P., 2021. An automatic method for clean glacier and nonseasonal snow area change estimation in High Mountain Asia from 1990 to 2018. Remote Sens Environ, 258. doi:10.1016/j.rse.2021.112376

Kong, Y., 2011. Accuracy analysis of the 1:100 000 topographic maps of Henan province in early period of the Republic of China. Journal of Geo-Information Science, 13(2): 226-233.

Kuang, X., Jiao, J.J., 2016. Review on climate change on the Tibetan Plateau during the last half century. Journal of Geophysical Research: Atmospheres, 121(8): 3979-4007. doi:10.1002/2015JD024728

Lei, Y. et al., 2013. Coherent lake growth on the central Tibetan Plateau since the 1970s: characterization and attribution. J Hydrol, 483: 61−67. doi:10.1016/j.jhydrol.2013.01.003

Li, B. et al., 2017. Climate change driven water budget dynamics of a Tibetan inland lake. Global Planet Change, 150: 70-80. doi:10.1016/j.gloplacha.2017.02.003

Li, J., Sheng, Y., 2012. An automated scheme for glacial lake dynamics mapping using Landsat imagery and digital elevation models: a case study in the Himalayas. Int J Remote Sens, 33(16): 5194−5213. doi:10.1080/01431161.2012.657370

Li, X. et al., 2019. High-temporal-resolution water level and storage change data sets for lakes on the Tibetan Plateau during 2000–2017 using multiple altimetric missions and Landsat-derived lake shoreline positions. Earth Syst. Sci. Data, 11(4): 1603-1627. doi:10.5194/essd-11-1603-2019

Liu, B. et al., 2016. Outburst Flooding of the Moraine-Dammed Zhuonai Lake on Tibetan Plateau: Causes and Impacts. IEEE Geoscience and Remote Sensing Letters, 13(4): 570-574. doi:10.1109/lgrs.2016.2525778

Ma, R. et al., 2010. A half-century of changes in China's lakes: Global warming or human influence? Geophys Res Lett, 37(24): L24106. doi:10.1029/2010gl045514

Masek, J.G. et al., 2020. Landsat 9: Empowering open science and applications through continuity. Remote Sens Environ, 248. doi:10.1016/j.rse.2020.111968

McFeeters, S., 2013. Using the Normalized Difference Water Index (NDWI) within a Geographic Information System to Detect Swimming Pools for Mosquito Abatement: A Practical Approach. Remote Sensing, 5(7): 3544-3561. doi:10.3390/rs5073544

Otsu, N., 1979. A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man and Cybernetics, 9(1): 62−66. doi:10.1109/TSMC.1979.4310076

Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature, 540: 418−422. doi:10.1038/nature20584

Peng, S., Ding, Y., Liu, W., Li, Z., 2019. 1 km monthly temperature and precipitation dataset for China from 1901 to 2017. Earth System Science Data, 11(4): 1931-1946. doi:10.5194/essd-11-1931-2019

Phan, V.H., Lindenbergh, R., Menenti, M., 2012. ICESat derived elevation changes of Tibetan lakes between 2003 and 2009. International Journal of Applied Earth Observation and Geoinformation, 17: 12−22. doi:10.1016/j.jag.2011.09.015

Pickens, A.H. et al., 2020. Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full Landsat time-series. Remote Sens Environ, 243. doi:10.1016/j.rse.2020.111792

Ran, Y., Li, X., Cheng, G., 2018. Climate warming over the past half century has led to thermal degradation of permafrost on the Qinghai–Tibet Plateau. The Cryosphere, 12(2): 595-608. doi:10.5194/tc-12-595-2018

Roy, D.P. et al., 2014. Landsat-8: Science and product vision for terrestrial global change research. Remote Sens Environ, 145: 154−172. doi:10.1016/j.rse.2014.02.001

480 Salerno, F. et al., 2012. Glacial lake distribution in the Mount Everest region: Uncertainty of measurement and conditions of formation. Global Planet Change, 92–93: 30-39. doi:10.1016/j.gloplacha.2012.04.001

Shean, D.E. et al., 2020. A Systematic, Regional Assessment of High Mountain Asia Glacier Mass Balance. Frontiers in Earth Science, 7: 363. doi:10.3389/feart.2019.00363

Sheng, Y. et al., 2016. Representative lake water extent mapping at continental scales using multi-temporal Landsat-8 imagery. Remote Sens Environ, 185: 129-141. doi:10.1016/j.rse.2015.12.041

485 Shu, S. et al., 2021. Estimation of snow accumulation over frozen Arctic lakes using repeat ICESat laser altimetry observations – A case study in northern Alaska. Remote Sens Environ, 216: 529-543. doi:10.1016/j.rse.2018.07.018

Song, C., Huang, B., Ke, L., 2013. Modeling and analysis of lake water storage changes on the Tibetan Plateau using multi-mission satellite data. Remote Sens Environ, 135: 25−35. doi:10.1016/j.rse.2013.03.013

490 Su, R., Wang, F., Pan, W., 2018. A dataset of Xinjiang region lakes and wetlands in late Qing and Republican China. Science Data Bank, 3(3). doi:10.11922/sciencedb.618

Sun, J. et al., 2018. Linkages of the dynamics of glaciers and lakes with the climate elements over the Tibetan Plateau. Earth-sci Rev, 185: 308-324. doi:10.1016/j.earscirev.2018.06.012

495 Verpoorter, C., Kutser, T., Seekell, D.A., Tranvik, L.J., 2014. A Global Inventory of Lakes Based on High-Resolution Satellite Imagery. Geophys Res Lett, 41: 6396−6402. doi:10.1002/2014GL060641

Wan, W. et al., 2014. Monitoring lake changes of Qinghai-Tibetan Plateau over the past 30 years using satellite remote sensing data. Chinese Sci Bull, 59(10): 1021−1035. doi:10.1007/s11434-014-0128-6

Wan, W. et al., 2016. A lake data set for the Tibetan Plateau from the 1960s, 2005, and 2014. Scientific Data, 3:
500 160039. doi:10.1038/sdata.2016.39

Wang, S., Dou, H., 1998. Chinese Lakes Inventory. Science Press, Beijing, China.

Woolway, R.I. et al., 2020. Global lake responses to climate change. Nature Reviews Earth & Environment, 1: 388–403. doi:10.1038/s43017-020-0067-5

Wu, Y., Zhang, X., Zheng, H., Li, J., Wang, Z., 2017. Investigating changes in lake systems in the south-central
505 Tibetan Plateau with multi-source remote sensing. Journal of Geographical Sciences, 27(3): 337-347. doi:10.1007/s11442-017-1380-x

Wulder, M.A. et al., 2019. Current status of Landsat program, science, and applications. Remote Sens Environ, 225: 127-147. doi:10.1016/j.rse.2019.02.015

Yamazaki, D., Trigg, M.A., Ikeshima, D., 2015. Development of a global ~ 90 m water body map using multi-
510 temporal Landsat images. Remote Sens Environ, 171: 337−351. doi:10.1016/j.rse.2015.10.014

Yang, K. et al., 2017. Recent dynamics of alpine lakes on the endorheic Changtang Plateau from multi-mission satellite data. J Hydrol, 552: 633-645. doi:10.1016/j.jhydrol.2017.07.024

Yao, T. et al., 2012. Different glacier status with atmospheric circulations in Tibetan Plateau and surroundings. Nature Clim. Change, 2(9): 663−667. doi:10.1038/nclimate1580

515 Yu, J. et al., 2016. Developing daily cloud-free snow composite products from MODIS Terra-Aqua and IMS for the Tibetan Plateau. Ieee T Geosci Remote, 54(4): 2171−2180. doi:10.1109/TGRS.2015.2496950

Yu, S. et al., 2020. Spatio-temporal evolution of riparian lakes in Dongting Lake area since the late Qing Dynasty. Acta Geographica Sinica, 75(11): 2346-2361.

Zhang, G., Yao, T., Xie, H., Kang, S., Lei, Y., 2013. Increased mass over the Tibetan Plateau: From lakes or
520 glaciers? Geophys Res Lett, 40(10): 2125−2130. doi:10.1002/grl.50462

Zhang, G. et al., 2017a. Lake volume and groundwater storage variations in Tibetan Plateau's endorheic basin. Geophys Res Lett, 44(11): 5550-5560. doi:10.1002/2017GL073773

Zhang, G. et al., 2017b. Automated water classification in the Tibetan Plateau using Chinese GF-1 WFV data. Photogrammetric Engineering & Remote Sensing, 83(3): 33-43. doi:10.14358/PERS.83.7.415

525 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. doi:10.1016/j.scib.2019.07.018

Zhang, G., Chen, W., Zheng, G., Xie, H., Shum, C.K., 2020a. Are China's water bodies (lakes) underestimated? Proceedings of the National Academy of Sciences, 117(112): 6308-6309. doi:10.1073/pnas.1922250117

Zhang, G. et al., 2020b. Response of Tibetan Plateau lakes to climate change: Trends, patterns, and mechanisms.
530 Earth-sci Rev, 208: 103269. doi:10.1016/j.earscirev.2020.103269

Zhang, G., Bolch, T., Chen, W., Crétaux, J.F., 2021. Comprehensive estimation of lake volume changes on the Tibetan Plateau during 1976–2019 and basin-wide glacier contribution. Science of the Total Environment 772: 145463. doi:10.1016/j.scitotenv.2021.145463

Zhang, Y., Zhang, G., Zhu, T., 2020c. Seasonal cycles of lakes on the Tibetan Plateau detected by Sentinel-1 SAR
    data. Sci Total Environ, 703: 135563. doi:10.1016/j.scitotenv.2019.135563

Zhao, L., Simon Wang, S.Y., Meyer, J., 2017. Interdecadal Climate Variations Controlling the Water Level of Lake
    Qinghai over the Tibetan Plateau. J Hydrometeorol, 18(11): 3013-3025. doi:10.1175/jhm-d-17-0071.1

Zhou, J. et al., 2015. Exploring the water storage changes in the largest lake (Selin Co) over the Tibetan Plateau
    during 2003–2012 from a basin-wide hydrological modeling. Water Resour Res, 51(10): 8060−8086.
    doi:10.1002/2014WR015846

Zhu, Z. et al., 2019. Benefits of the free and open Landsat data policy. Remote Sens Environ, 224: 382-385.
    doi:10.1016/j.rse.2019.02.016