



OPEN

Reservoir inventory for China in 2016 and 2021

DATA DESCRIPTOR

Haiying Zhang^{1,2} , Fang Chen^{1,2,3}, Lei Wang^{1,2,4} , Ning Wang⁵ & Bo Yu^{1,2}

Reservoir inventories are essential for investigating the impact of climate change and anthropogenic activities on water scape changes. They provide fundamental data sources to explore the sustainability and management efficiency of water resources. However, publicly released reservoir inventories are currently limited to a single temporal domain. As a result, the effectiveness of governmental policy implementation on water resources remains to be explored due to the lack of multi-time datasets. In this study, we generated a reservoir inventory for China for the years 2016 and 2021 with an overall accuracy of 99.71%. The reservoirs were visually interpreted from annually composited Landsat images, and each reservoir is represented by a polygon with attributes of reservoir name, area and storage capacity. About 10.32% of the reservoirs have increased storage capacity from 2016 to 2021, while 22.73% have decreased. Most provinces and river basins in China have expanded their accumulated storage capacity from 2016 to 2021.

Background & Summary

Reservoirs are recognized as one of the most effective means to deal with agricultural irrigation, hydropower generation, water supply, daily recreation, flood control and disaster relief^{1–3}. To make the best use of water resources and meet the continuous demands of water consumption and hydropower, numerous reservoirs have been constructed worldwide^{4–6}. One fifth of the global cultivated land is used for irrigation to provide 33% of the world's food⁷. Recent research suggests that by 2050, 70% more food will be needed to accommodate shifts in dietary patterns and the 40% increase in the world's population⁸. The additional food will be partly produced from irrigation, requiring 11% more water from reservoirs². Moreover, approximately 20% of the world's electricity is generated by hydropower, which amounts to 7% of worldwide energy². China is an arid country that seriously lacks water. The total freshwater resource of China accounts for only 6% of the world's total resources, but the average water resource capacity per capita is only one fourth of the global average level⁹. Moreover, China is the country that consumes the largest amount of water¹⁰. As a result, China has built numerous reservoirs over the past few decades to meet the increasing demands for effective water use, flood control, and power generation^{11–14}. In the 1950s, there were only 20 reservoirs with a storage capacity greater than 0.1 cubic kilometers³. By 2019, 98,112 reservoirs with a total capacity of 898.3 billion km³ had been constructed¹⁵.

To deal with the severe pressure from the increasing population and anthropogenic impacted droughts in the future, it is crucial to estimate water resource availability in terms of surface area, storage capacity, and spatial distribution for effective management strategy proposal¹⁶. The development of remote sensing technology made it possible to delineate georeferenced reservoirs with detailed area and storage capacity from remotely sensed images^{17,18}. Numerous research studies have delineated reservoirs from multi-source remotely sensed images to analyze the impact of anthropogenic activities and create global or regional datasets of reservoirs^{2,7,19,20}. Several dam and reservoir datasets have been proposed, such as the World Register of Dams (WRD)²¹, the Global Georeferenced Database of Dams (GOODD) V1²², and the Global Reservoir and Dam database (GRanD)². However, the datasets generally cover limited records, and the reservoir boundaries interpreted lack spatial resolution consistency. Some large reservoirs have coarse digitized boundaries while some small reservoirs have detailed boundaries due to the multiple spatial resolutions of images used for interpretation. Moreover, most

¹International Research Center of Big Data for Sustainable Development Goals, Beijing, 100094, China. ²Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, 100094, China. ³University of Chinese Academy of Sciences, Beijing, 100049, China. ⁴School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin, 541004, China. ⁵College of Urban and Environmental Sciences, Peking University, Beijing, 100871, China. ✉e-mail: wanglei@radi.ac.cn; yubo@radi.ac.cn

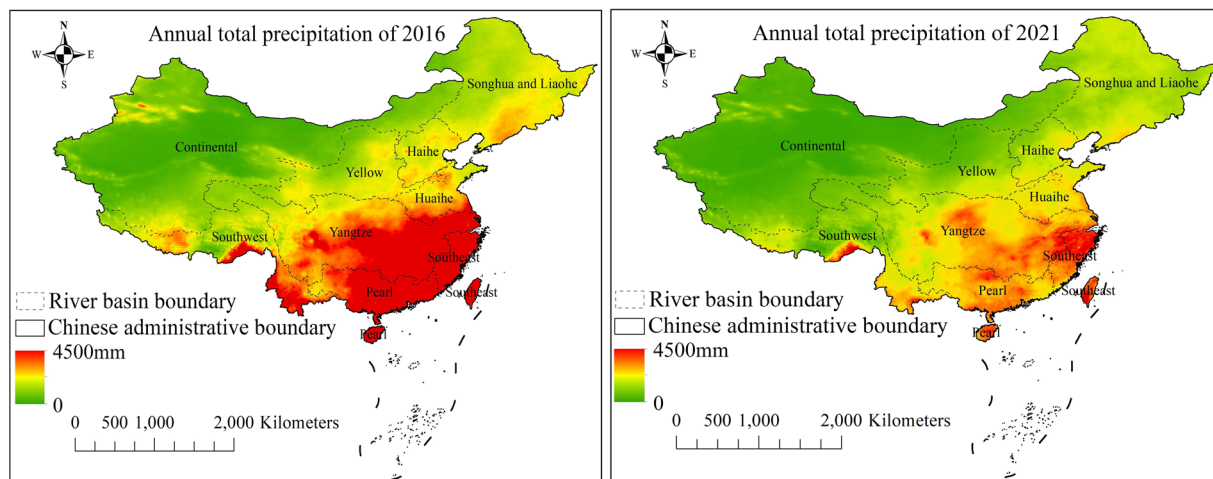


Fig. 1 Annual accumulated precipitation pattern across China⁶² in 2016 (a) and 2021 (b).

records in the datasets lack detailed reservoir information, including area and storage capacity. That hinders thorough and comprehensive analysis in the absence of detailed reservoir distribution patterns²³.

Landsat images have been widely used to delineate reservoirs in various locations, such as Zimbabwe²⁴, India²⁵, Ghana²⁶, the Yellow River of China²⁷, and mainland China⁷. There have been many reservoir datasets proposed based on Landsat images due to their global coverage and public availability, with a spatial resolution of 30 m. Application of the Google Earth Engine (GEE) platform has made it possible to map reservoirs for large-scale areas²⁸. Wang, *et al.* delineated large dams, reservoirs, and lakes of China based on spectral indices calculated from Landsat images obtained from GEE in 2019²⁹. Meanwhile, Song, *et al.* presented a comprehensive reservoir inventory dataset of China Reservoir Dataset (CRD) by compiling multiple public reservoir datasets constructed from GEE⁶. Although the published reservoir datasets cover detailed distribution patterns of a single temporal phase, they lack the variation patterns of reservoirs in different time domains, which make it difficult to analyze the sustainability improvement of water use in different time domains.

In 2016, the Chinese government released regulations on Farmland Water Conservancy to encourage the use of reservoirs for irrigation. Medium and large reservoirs are significant in agricultural irrigation. Although medium and large reservoirs in China account for less than 5% of total reservoirs, the corresponding storage capacity accounts for 92%. However, with the long-term use, some reservoirs have been dilapidated mainly because of lacking reinforcement³⁰. In 2021, the Chinese National Development and Reform Commission announced to double the budget to reinforce the hazardous reservoir to 4 billion. Before implementing the policy in strengthening the hazardous reservoirs with greater effort, the effectiveness of the policy in encouraging the use of reservoirs for irrigation remains to be explored through recent technical developments. However, analyzing the patterns of reservoir storage capacity change with irrigation cropland area between 2016 and 2021 from the China Water Resources Statistical Yearbook, released by the Chinese government, is challenging without detailed reservoir inventory in the two time-domains. Local investigation is the most commonly used method, but it is costly and laborious. Exploring the distribution pattern change of medium and large reservoirs through publicly available remotely sensed images in China can generally delineate the water resources utilization status change from year 2016 to 2021.

Aiming to propose a reservoir inventory of China for 2016 and 2021 to explore the reservoir storage capacity change of medium and large reservoirs and evaluate the change in response to the regulation of Farmland Water Conservancy, we generated a reservoir inventory dataset (MLRC) over China for 2016 and 2021 by classification based on annually composited Landsat images from GEE, aided by geological datasets of the National 1:250,000 Public Basic Geographic Database of 2019, and reservoir locations acquired from Baidu Maps in 2016 and 2021. We composited the Landsat images annually as base maps for potential reservoir extraction and validation through visual interpretation. The locations of reservoirs in Geological datasets and Baidu Maps are used to initialize the reservoirs for the sake of visual interpretation. The reservoir inventories were evaluated by stratified random sampling³¹, referring to higher spatial resolution images from Google Earth. Our shared dataset consists of a detailed reservoir inventory with area and storage capacity. Based on the reservoir inventories for 2 years, we explored the storage capacity change of medium and large reservoirs, together with the irrigation cropland area change, aided by publicly available 30 m spatial resolution land cover land use products³². This study not only evaluates the reservoir capacity change for water conservancy but also advances research in other water resource utilization-related research by providing reservoir inventories in two different time domains.

Methods

Study area. Our study covers the entire administrative area of China. Due to its location on the west coast of the Pacific Ocean, China experiences a significant monsoon climate³³, resulting in uneven spatial distribution of freshwater resources. The very uneven distribution of accumulated precipitation, as shown in Fig. 1, leads to severe imbalances in water resource distribution across different regions. The cultivated area in the Yangtze

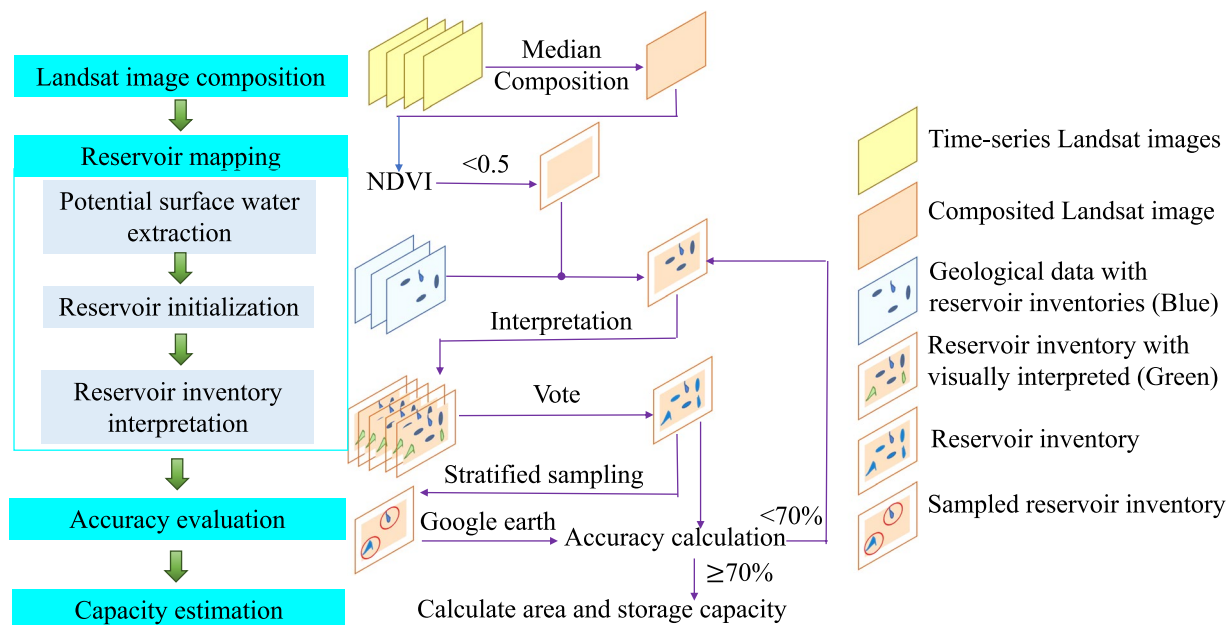


Fig. 2 Flowchart for creating reservoir inventory.

River Basin and south of the Yangtze River accounts for 36% of China's total cultivated land, while its freshwater resources make up 80% of the total capacity³⁴. In contrast, the water resources of the Yellow River Basin, Huaihe River Basin, and Haihe River Basin only account for 8% of the total capacity, but their cultivated land accounts for 40%³⁵. In addition to the spatially imbalanced water resource distribution, interannual variation in precipitation in Northern and Southern China is another factor leading to frequent floods and droughts, resulting in greater instability for agricultural production³⁶. To balance the water resources and address the aforementioned issues, the storage capacity of numerous reservoirs has been expanded to support water conservancy and irrigation for farmland, promoted by the regulation on Farmland Water Conservancy from the Chinese government. However, the efficiency of the regulation implementation remains to be explored.

Methodological framework. Reservoir inventories are mainly achieved in two phases: water body detection and classification. In terms of water body detection, a variety of spectral band ratios have been used to adjust the threshold to distinguish water from background objects, such as snow, bare land, and vegetation, through visual interpretation³⁷. The spectral band ratios include normalized difference water index (NDWI)³⁸, normalized difference vegetation index (NDVI)³⁹, etc. Regarding reservoir classification, visual interpretation is usually used, with the aid of Google Earth images and ancillary data⁴⁰, to distinguish reservoir polygons from other water bodies, such as lakes.

The flow chart for creating the reservoir inventory in our study consists of four main steps: annual Landsat image composition, reservoir mapping, accuracy evaluation, and capacity estimation, as illustrated in Fig. 2.

Step 1: Annual landsat image composition. The annual Landsat image is composed by assigning the medium top of atmosphere (TOA) reflectance calculated from the collected annual all available Landsat7 and Landsat8 images captured during the growing season in 2016 and 2021 over China with cloud coverage of <math><50\%</math>. The composition generally comprises of five steps: (1) conversion from digital number of raw Landsat images to TOA⁴¹; (2) cloud and shadow removal by the proposed bagging model⁴²; (3) radiometric correction to resist the reflectance difference of images taken in different dates triggered by different atmospheric conditions⁴³; (4) calculate NDWI for each corrected Landsat image and rank the values in terms of each pixel location; (5) assign the TOA in each channel of pixel whose NDWI ranks the medium as that of corresponding composited pixel location.

Step 2: Reservoir mapping. Reservoir mapping involves potential surface water extraction based on spectral indices, reservoir location initialization based on geological data, and reservoir interpretation by five experts with rich experiences in dam related research. Using annually composited Landsat images of China, we first calculated the NDVI spectral index according to Eq. (1), enhancing the surface water information. The notations of I_{Red} and I_{NIR} indicate DN values of the Red and NIR bands, respectively. We set an experienced threshold of 0.5 to extract potential water bodies from background objects.

$$\text{NDVI} = (I_{\text{NIR}} - I_{\text{RED}}) / (I_{\text{NIR}} + I_{\text{RED}}) \quad (1)$$

Second, we initiated the reservoir locations based on geological data and supplemented the reservoir inventories through visual interpretation by five different experts. In cases where experts disagreed, higher spatial

resolution images from Google Earth were used for further judgment. The reservoir datasets we generated had a 5-year gap, so different geological datasets were applied to initialize the reservoir locations. For the inventories of 2016, we used the Wetland Map of China and downloaded records from Baidu Maps. For the inventories of 2021, we used the National Public Basic Geographic Database and downloaded records from Baidu Maps. Additionally, we adopted the reservoir inventories generated in 2016 to initialize the 2021 inventories due to the stable attributes of reservoirs constructed over a long period of time.

With the limited spatial resolution of Landsat images, we extracted reservoirs with an area larger than 0.405 km² (450 pixels × 30 m × 30 m/pixel) in the generated inventories.

Step 3: Mapping accuracy evaluation. The accuracy of the generated reservoir inventories in both years is evaluated using stratified sampling^{44,45} with reference to the annually composited Landsat images and higher spatial resolution images from Google Earth⁴⁶. Five experienced experts are involved into the validation interpretation.

Step 4: Reservoir capacity estimation. For each mapped reservoir polygon, several attributes are added through calculations on the ArcGIS platform. These attributes include area, storage capacity, and reservoir name. Area is calculated in units of square kilometers. The storage capacity of each reservoir (R_c) is calculated using an experienced equation derived from work done by Yang *et al.*⁷. As illustrated in Eq. (2), the storage capacity is calculated based on area of each reservoir, indicated by notation of *Area*. The experienced equation has been validated to calculate storage capacity of 2185 reservoirs over China, referring to the statistics released by the Chinese government with R^2 of 0.9097⁷. The strong approximate to official statistics demonstrates the strong applicability of the experienced equation in calculating storage capacity for Chinese reservoirs. Therefore, it has been adopted in our study to calculate the storage capacity for each mapped reservoir in China. Based on the storage capacity, the reservoirs were grouped into two categories: medium and large reservoirs. Medium reservoirs are defined as having a storage capacity of less than 100 million m³, but no less than 10 million m³, while large reservoirs have a capacity of no less than 100 million m³.

$$R_c = 25.841 \times Area^{1.05} \quad (2)$$

Data source used for reservoirs delineation. *Landsat images.* The Landsat images used to interpret reservoirs are median composited following the methods outlined by Potapov, *et al.*⁴³. The median spectral reflectance of the four spectral bands—Red, Near-Infrared (NIR), Shortwave Infrared 1 (SWIR1), and Shortwave Infrared 2 (SWIR2)—from all Landsat images throughout each year are used to determine the spectral reflectance of the annually composited image. The composited spectral reflectance is then converted to an 8-bit digital number (DN) using Eq. (3) for computation purposes. The values of the notation g vary in the different bands of Red, NIR, SWIR1, and SWIR2, with values of 508, 254, 363, and 423, respectively, as established by Potapov, *et al.*⁴³.

$$DN = \text{top} \times g + 1 \quad (3)$$

Geographical data. The geographic data used in this study include the Wetland Map of China in 2008, the National 1:250,000 Public Basic Geographic Database of 2019, and reservoir locations obtained from Baidu Maps for 2016 and 2021. Due to the coarse spatial resolution or biased local reports, most of these data sources cover only a limited number of reservoirs in China. The Wetland Map of China contains 68,521 reservoir and pond inventories, which were visually interpreted based on Landsat images from 2008 and used to initialize reservoir boundaries in China for 2016. In addition to the Wetland Map of China, the names and locations (longitude and latitude) of about 25,897 reservoirs downloaded from Baidu Maps in 2016 were used to aid in the visual interpretation of reservoir inventories.

The National Public Basic Geographic Database for 2019 is publicly available and was used to initialize reservoir inventories for 2021. It covers 197,664 inventories of surface water systems, such as lakes, reservoirs, and rivers, with reservoirs accounting for 13,020 inventories. This database is a significant public data source for initializing reservoir inventories for 2021. Additionally, the reservoir locations and names for 2021 were downloaded from Baidu Maps to aid in the visual interpretation of reservoir inventories in this study.

Land cover land use product. The impact of changes in reservoir capacity on irrigation cropland is analyzed using the Global Land Cover Product with Fine Classification System in 30 m (GLC_FCS30), which was proposed by Zhang *et al.*³². This publicly available global land use product has an overall accuracy of 82.5%⁴⁷ and is generated from time-series Landsat images taken every 5 years from 1985 to 2020. It is classified into 29 land use categories, covering different levels of cropland, forest, shrubland, grassland, wetlands, impervious surfaces, bare areas, water bodies, and permanent ice and snow. Within the fine classification system of cropland⁴⁷, irrigated cropland is extracted for analysis. As our study focuses on the impact of irrigation policy on reservoir storage capacity in China from 2016 to 2021, we adopt the GLC_FCS30 products for 2015 and 2020 for analysis.

Spatial distribution pattern of medium and large and reservoirs in China. In this study, aided by high-resolution images on the Google Earth platform, we utilized manual interpretation method based upon the preprocessed Landsat images on the ArcGIS 10.8.2 platform, to generate a 30 m spatial resolution reservoir inventory of China for 2016 and 2021, and calculated the corresponding attributes of reservoirs' area and capacity, which are demonstrated in Fig. 3. The inventory was visually interpreted by five experienced experts using annually composited Landsat images after initializing them with publicly available geological datasets. The water surface area and storage capacity of reservoirs have undergone significant changes due to artificial

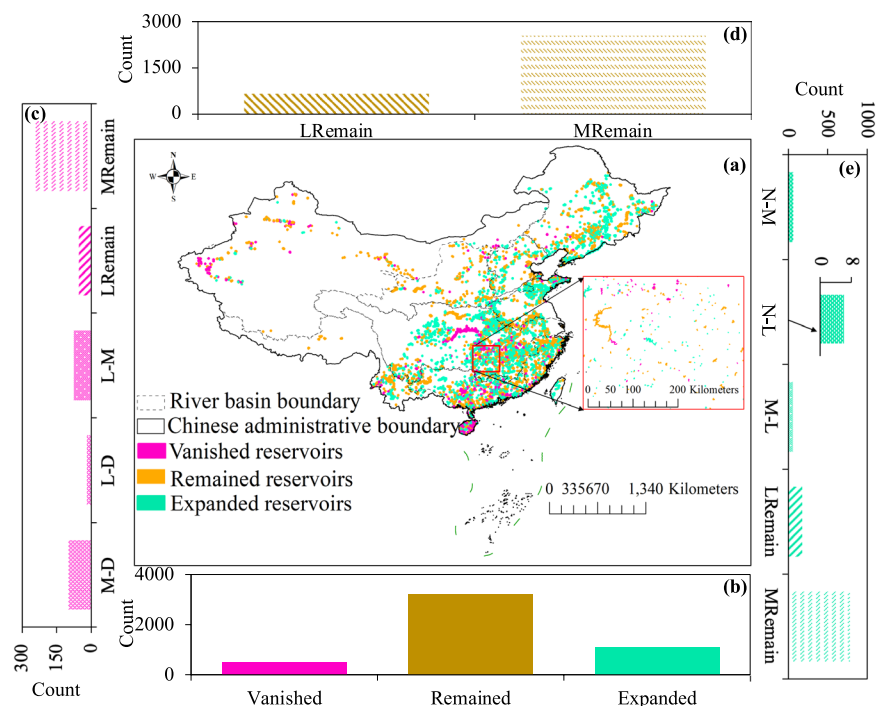


Fig. 3 Comparison of reservoirs across China between the years 2016 and 2021. **(a)** Change in reservoir capacity from 2016 to 2021; **(b)** Number of reservoirs in different change categories (vanished, remained, expanded) from 2016 to 2021; **(c)** Vanished reservoirs with different capacity sizes from 2016 to 2021; **(d)** Remained reservoirs with different capacity sizes from 2016 to 2021; **(e)** Expanded reservoirs with different capacity change categories from 2016 to 2021. M-D and L-D represent the disappeared medium and large reservoirs, while N-M and N-L indicate newly constructed medium and large reservoirs. L-M and M-L demonstrate the category change of reservoirs from 2016 to 2021. LRemain and MRemain stand for reservoirs without category change.

regulation or watershed planning, to meet the continually increasing demand for hydropower energy, irrigation, flood control, drought relief, and aquaculture⁴⁸. Since reservoir storage capacity is more widely used to indicate reservoir water supply capability⁴⁹, storage capacity is used to distinguish the reservoirs into three categories based on the storage capacity change during the study period: expanded reservoirs, remained reservoirs, and vanished reservoirs. Expanded reservoirs are the ones with a capacity increase of more than 20% compared to that in 2016. The reservoirs whose capacity decreased by more than 20% are labeled as vanished reservoirs. The rest of the reservoirs, whose capacity changes were within 20% of that in 2016, are grouped as remained reservoirs. The 20% capacity change threshold is adopted to avoid seasonal water surface changes and boundary interpretation errors.

There are a total of 4,777 medium and large reservoirs in China from 2016 to 2021, which approximates the 4,872 reservoirs in the China water statistical yearbook. The medium and large reservoirs in 2016 had a capacity of 100,214.59 and 617,578.41 million m³, respectively, which is in accordance with the statistics released by the China Water Statistical Yearbook. Compared with the storage capacity in 2016, the accumulated medium and large reservoir capacity increased by 8,019.59 and 43,435.22 million m³ in 2021, where the storage capacity of newly constructed medium and large reservoirs contributed to 18.60% and 6.71%, respectively. With the enhanced storage capacity, the capability of water supply, flood control, and irrigation has significantly improved and been restored⁵⁰. In 2016, the total storage capacity of medium and large reservoirs was 717,793.00 million m³, which was increased by 51,454.81 million m³ in 2021. The accumulated reservoir storage change varied significantly in different provinces and river basins of China. Most reservoirs, accounting for 66.95%, have remained the same in storage capacity, as shown in Fig. 3b. The proportion of reservoirs with expanded storage capacity, including newly extracted reservoirs, is 10.32%. The remaining reservoirs, representing about 22.73% in 2016, have vanished or disappeared until 2021. Within the vanished reservoirs (Fig. 3c), 75 large reservoirs in 2016 vanished to medium reservoirs in 2021, 120 medium and large reservoirs turned into reservoirs with a capacity of less than 10 million m³ or disappeared, and 298 large or medium reservoirs remained in the same category despite their vanishing storage capacity. The expanded reservoirs stem from different extents of capacity change (Fig. 3e), including 71 medium reservoirs and large reservoirs newly constructed in 2021, 57 medium reservoirs transformed to large reservoirs from 2016 to 2021, and 958 medium and large reservoirs with a remained category change.

Due to varying water use strategies and uneven precipitation patterns, the accumulated storage capacity of medium and large reservoirs across different provinces in China is significantly imbalanced. Between 2016 and 2021, most provinces have expanded their reservoir storage capacity. As shown in Fig. 4a, Heilongjiang ranks first

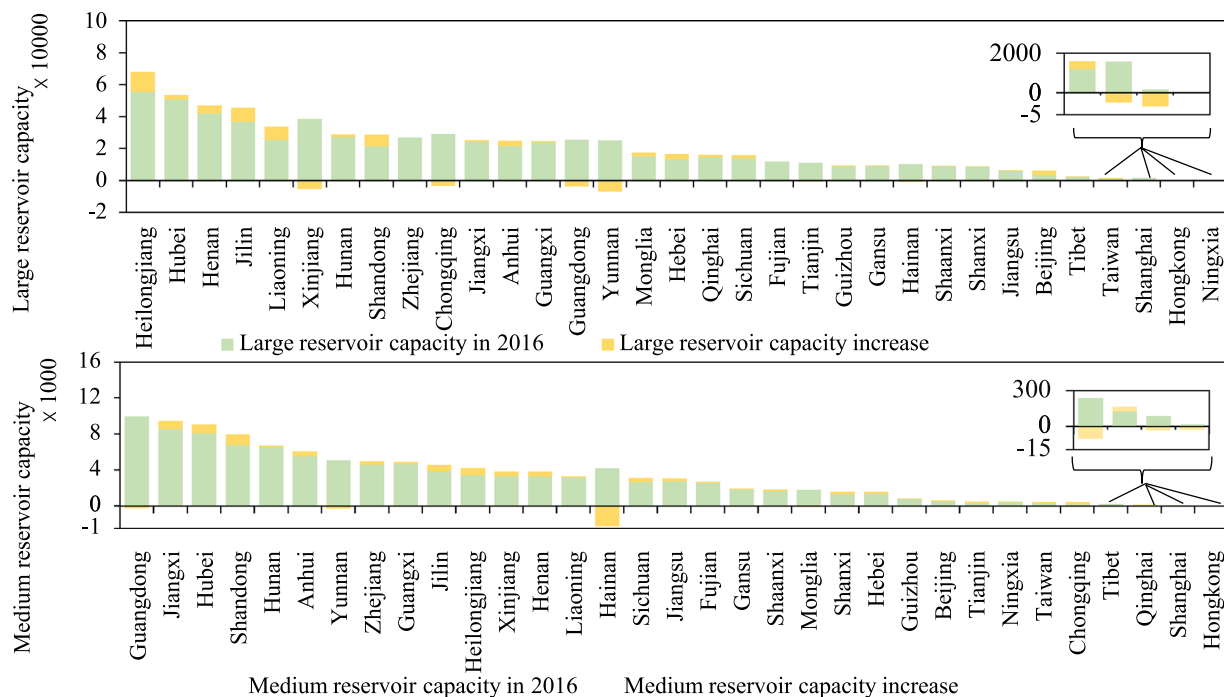


Fig. 4 Changes in accumulated reservoir storage capacity (million m^3) by province in China from 2016 to 2021, separated by large and medium reservoirs. The provinces are ranked from highest to lowest in terms of accumulated storage capacity.

in large reservoir capacity in 2021 with a total of 68,033.33 million m^3 , reflecting a large increase of 12,464.07 million m^3 from 2016. In addition to Heilongjiang, three other provinces—Hubei, Henan, and Jilin—also have capacities above 40,000 million m^3 in 2021. Notably, the storage capacity of large reservoirs in Liaoning province surpassed that of Xinjiang province, with a remarkable increase between 2016 and 2021.

The large precipitation in Liaoning province and the large water leakage resulting from the permeability of plain reservoirs in Xinjiang province⁵¹ are the main reasons for the change in water storage. Apart from Xinjiang province, Guangdong, Yunnan, and Chongqing provinces have lost more than 2,000 million m^3 of water during the study period, mainly due to drought and lack of rainfall in 2021. The provinces with decreased water storage capacity require more specific and effective strategies to recover and improve their water use efficiency. In our inventory, there is no obvious large reservoir in Ningxia province that can be extracted from Landsat images.

As shown in Fig. 4b, Guangdong, Jiangxi, and Hubei provinces have the largest medium reservoir storage capacities, exceeding 9,000 million m^3 in 2021. Guangdong province ranks first with the highest accumulated medium reservoir storage capacity throughout the study period. Shandong province witnessed the largest increase in medium reservoir storage capacity from 2016 to 2021, with 1,200.31 million m^3 , due to new reservoir construction and increased reservoir capacity implementation⁶. This corresponds with the statistics in the Jinan Water Resources Bulletin. The accumulated medium reservoir storage capacity of Hainan province decreased the most by 883.30 million m^3 , mainly due to fluctuations in annual precipitation in wet and dry years. This corresponds with the statistics in the Water Resources Bulletin of Hainan Province from 2016 to 2021. Moreover, it shows that the water conservation constructions in Hainan province require more advanced measures to improve the current relatively backward facilities to deal with fluctuating precipitation⁵².

Referring to Fig. 5, most river basins in China have increased their storage capacity in both medium and large reservoirs during the study period from 2016 to 2021. The Yangtze River Basin has the highest accumulated storage capacity in both medium and large reservoirs, with proportions of 35.18% and 30.38%, respectively, in 2021. The Songhua and Liaohe River Basin expanded the largest storage capacity with 31,051.85 million m^3 in large reservoirs of all the river basins (Fig. 5a). This is mainly due to the large increase in the accumulated capacity of Heilongjiang, Jilin, and Liaoning provinces in 2021, as shown in Fig. 4a. Due to the water leakage issue mainly in plain reservoirs in Xinjiang province, as demonstrated in Fig. 4, the accumulated storage capacity of medium reservoirs in the Continental River Basin decreased from 2016 to 2021 in Fig. 5a. The Pearl River Basin and Southwest River Basin witnessed an obvious decrease in both medium and large reservoir storage capacity during the study period, partly due to drought in Yunnan, Hainan, and Guangdong provinces.

Use case results: reservoir storage capacity changes. We synthesized irrigated cropland products to explore the impact of changes in medium and large reservoir storage capacity on the irrigated cropland area between 2016 and 2021 in nine agricultural regions of China. As shown in Fig. 6, the accumulated reservoir storage capacity in 2016 increased with the irrigated cropland area in most agricultural regions. Notably, the accumulated large reservoir storage capacity of the Southern China and Yunnan-Guizhou Plateau agricultural

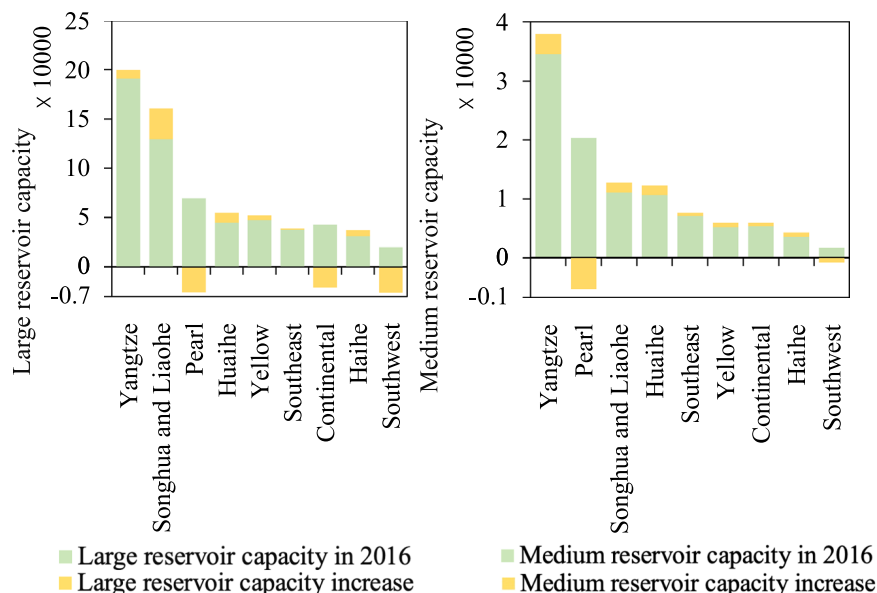


Fig. 5 Change in accumulated reservoir storage capacity (million m³) in each river basin across China from 2016 to 2021. (a) Change in capacity of large reservoirs; (b) Change in capacity of medium reservoirs.

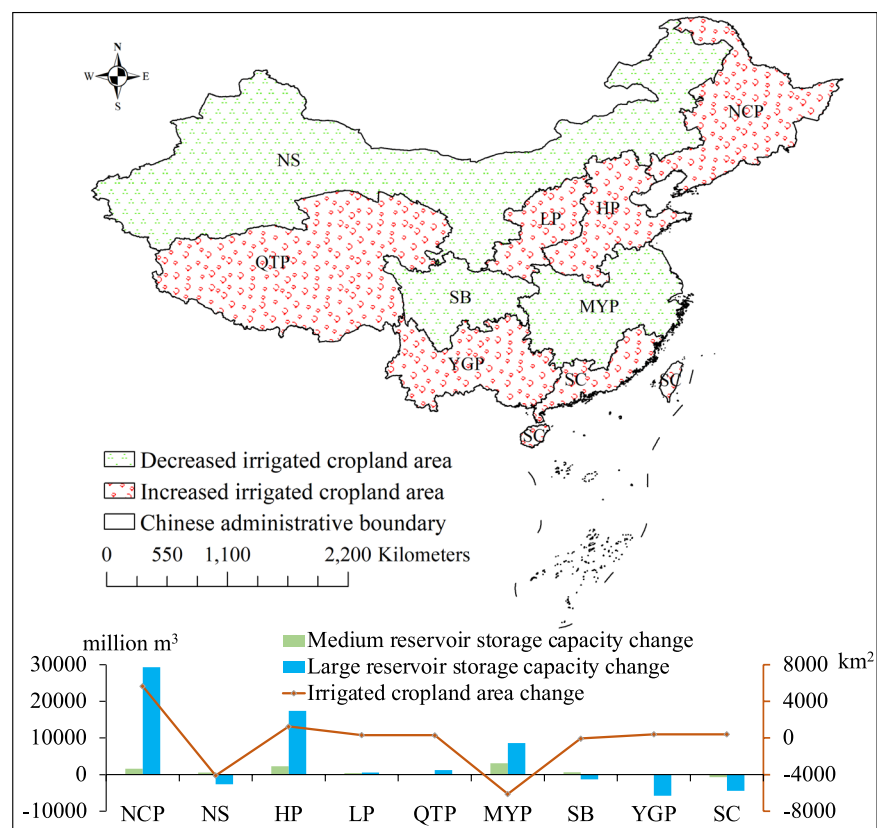


Fig. 6 Changes in large and medium reservoir storage capacity (million m³) and irrigated cropland area (km²) for nine agricultural regions in China during the study period. (NCP: Northeast China Plain; NS: Northern arid and semiarid region; HP: Huang-Huai-Hai Plain; LP: Loess Plateau; QTP: Qinghai Tibet Plateau; MYP: Middle-lower Yangtze Plain; SB: Sichuan Basin and surrounding regions; YGP: Yunnan-Guizhou Plateau; and SC: Southern China).

regions decreased significantly, while the irrigated cropland area increased during the study period. The reason for the Southern China case mostly attributes to the droughts in 2021 and the significant improvement of water use efficiency achieved in Guangdong province through continuous optimization of water use structure⁵³.

It demonstrates the importance of proposing advanced strategies to improve water use efficiency and resist the impact of climatic factors such as precipitation. As for the Yunnan-Guizhou Plateau agricultural region, Yunnan province suffered from serious droughts in 2021, resulting in a remarkable decrease in storage capacity. However, the Provincial Water Resources Department in Yunnan attaches great importance to the work of saving water for farmland irrigation and places water conservation in a priority position with a series of measures, such as improving water-saving irrigation technology⁵⁴, so that the irrigated cropland area has remained and slightly increased.

Apart from the agricultural regions with increased irrigated cropland area, the Middle-lower Yangtze Plain and Northern arid and semiarid region have significantly decreased the irrigated cropland area due to urban expansion and an increase in population for economic development⁵⁵. Compared with the Northern arid and semiarid region, the accumulated reservoir storage capacity of the Middle-lower Yangtze Plain has increased remarkably. This is mainly owing to the extreme precipitation in Northern China in 2021⁵⁶. With the Chinese government encouraging reservoir application for irrigation in 2016, most agricultural regions of China have witnessed an increased irrigated cropland area with enhanced water conservation capability of increased accumulated reservoir storage capacity. However, the issue of decreasing irrigated cropland area remains to be resolved with more practical solutions to avoid food crises, especially in the Middle-lower Yangtze Plain and Northern arid and semiarid region.

Data Records

We present an inventory of medium and large reservoirs in China for the years 2016 and 2021, named (1) MLRC_2016 and (2) MLRC_2021, based mainly on Landsat 8 images and other ancillary data sources. The inventory dataset is developed by Environmental Systems Research Institute, stored in the form of a shapefile (.shp) with each record represented as a polygon and including attributes such as reservoir area, storage capacity, and reservoir name. In the attributes tables of the.shp files, the field *Shape Area* stands for the area of a reservoir polygon in square kilometers; the field *StorCap* represents the estimated storage capacity of a reservoir in million cubic meters; and the field *Name_En* refers to the name of a reservoir. The dataset gives a national medium and large reservoir extent of 23314.00 and 24888.95 km² in 2016 and 2021 respectively, with the spatial resolution of 30 m. The dataset can be accessed at repository under data doi (<https://doi.org/10.5281/zenodo.8278702>)⁵⁷.

The projected coordinate system of the dataset is as follows:

Projected Coordinate System: Krasovsky_1940_Albers

Projection: Albers;
False Easting: 0°;
False Northing: 0°;
Central Meridian: 105°E;
Standard Parallel 1: 25°N;
Standard Parallel 2: 47°N;
Latitude of Origin: 0°;
Linear Unit: Meter.

Geographic Coordinate System: GCS_Krasovsky_1940

Datum: D Krasovsky 1940;
Prime Meridian: Greenwich;
Angular Unit: Degree.

Technical Validation

In order to evaluate the comprehensiveness and accuracy, we compared the proposed reservoir inventory MLRC_2016 and MLRC_2021 with the publicly accessible reservoir datasets covering China in terms of large and medium reservoir number, area, storage capacity, the number of reservoirs with attribute of names, and acquisition year. The publicly accessible reservoir datasets comprise of China-LDRL²⁹, CRD database⁶, GOODD²², GeoDAR v1.1²³, GRanD v1.3², National 1:250000 Public Basic Geographic Database of 2019 (ND2019) (<https://www.webmap.cn/main.do?method=index>), and the recently published Global Reservoir Storage (GRS) dataset⁵⁸, as listed in Table 1. It is challenging to conduct a fair comparison between our two-phase inventory of 2016 and 2021 and the published reservoir datasets, because the publicly released datasets are generally collected from images and statistics of different years than our proposed MLRC dataset (Table 1). Since actual reservoir mapping over China is limited accessible, the publicly released census of China water statistical yearbook (CWY) of year 2016 and 2021 are referred to as ground truth in this study with actual statistics of reservoir number and storage capacity. Detailed comparisons are listed in Table 1. The reservoirs of the inventories for comparison are subtracted to Chinese administrative area with the storage capacity of no less than 10 million m³, in accordance with the focus of our reservoir inventory on medium and large reservoirs.

It is apparent that our proposed reservoir inventory draws closest to the actual number of reservoirs and storage capacity in both years of 2016 and 2021, as listed in the census of China water statistical yearbook. The dataset of GOODD only comprises of locations of dams, rather than detailed inventories, lacking detailed reservoir spatial characteristics. ND2019 is a composite dataset with of 6728 reservoir inventories, which is far larger than the actual reservoir number, indicating many falsely extracted reservoirs in the ND2019 dataset. The datasets of China-LDRL and GeoDAR in China are captured in year of 2019, which should cover more reservoirs than that of year 2016, but the number of their recorded reservoirs is far smaller than that of China water statistical yearbook in year of 2016. In terms of CRD and GRanD datasets, the numbers of reservoirs are both 885, indicating

Reservoir Dataset	Acquisition year	Number of reservoirs	Coverage Area	Storage Capacity	Number of reservoirs with names
China-LDRL	2019	2194	16351.95		
ND2019	2015–2017	6728	21432.57		6728
CRD	1984–2020	885	28211.08	79569.41	856
GOODD	2020	9215 dams			
GRanD	2019	885	16844.41		
GeoDAR	2019	1713	18162.84		
GRS	1999–2018	885			
MLRC	2016	4631	23314.00	717793.00	3862
	2021	4602	24888.95	769247.81	4071
CWY	2016	4610		826200	
	2020	4872		858000	

Table 1. Statistical comparison between our proposed reservoir inventory with publicly available datasets, in terms of reservoir number, coverage area(km²), storage capacity (million m³), number of reservoirs with names, and acquisition year.

Category	Reservoir	Background	UA (%)	PA (%)	OA (%)
Reference reservoir	395	2	99.5014	98.7514	99.71
Reference background	5	1998	99.75	99.90	

Table 2. Reservoir inventory proposal evaluation statistics.

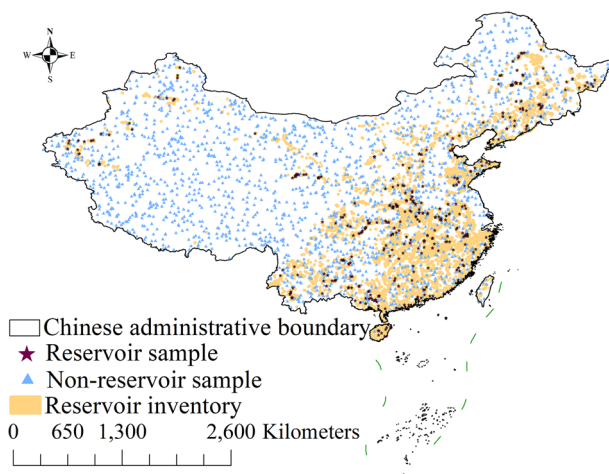


Fig. 7 Evaluation sample distributions of reservoir and non-reservoir ground objects.

large quantities of large and medium reservoirs are missed in the datasets. It needs to point the dataset GRS provides monthly reservoir statistics, but the detailed inventory is inaccessible publicly and the inventory shown on the platform of Google Earth Engine for each reservoir does not change over time as well. With the limited number of reservoirs recorded in China in GRS, the application of GRS dataset is largely restricted. Referring to the census of China water statistical yearbook, our proposed MLRC dataset is recognized as most reliable with the most approximate number of reservoirs and storage capacity.

To present an objective evaluation, our inventory is evaluated by visual interpretation with the strategy of stratified sampling^{44,45}. Stratified sampling is an evaluation strategy that has been widely used for large-scale object mapping evaluation⁵⁹. The reservoir inventories in MLRC_2016 and MLRC_2021 are evaluated using selected samples, as shown in Fig. 7. We randomly selected 400 reservoir sample points within the inventory and 2,000 non-reservoir sample points outside the inventory. Referring to annually composited Landsat images and high-resolution images from Google Earth, the selected reservoir and non-reservoir samples are validated through visual interpretation by five experienced experts and voted in case of disagreement. We calculated the indicators of user accuracy (UA), producer accuracy (PA), and overall accuracy (OA) to objectively evaluate the accuracy of our proposed inventory. UA is used to measure the percentage of correctly extracted reservoir samples. PA is used to evaluate the percentage of correctly extracted reservoir samples over all the actual reservoir samples. OA is a comprehensive indicator of both accuracies. The inventories of reservoirs are mostly accurate, with an OA of 99.71%, as listed in Table 2. Most extracted reservoir pixels are correct, with a 99.50% UA, and

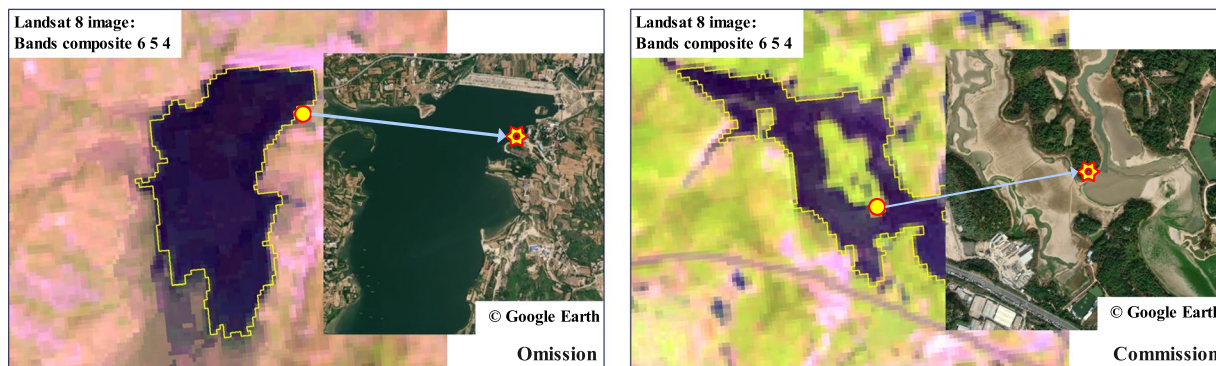


Fig. 8 Demonstration of omitted and committed reservoir samples with reference to high spatial resolution images from Google Earth. **(a)** Example of omitted actual reservoir pixel; **(b)** Example of committed actual background pixel but mis-extracted as a reservoir.

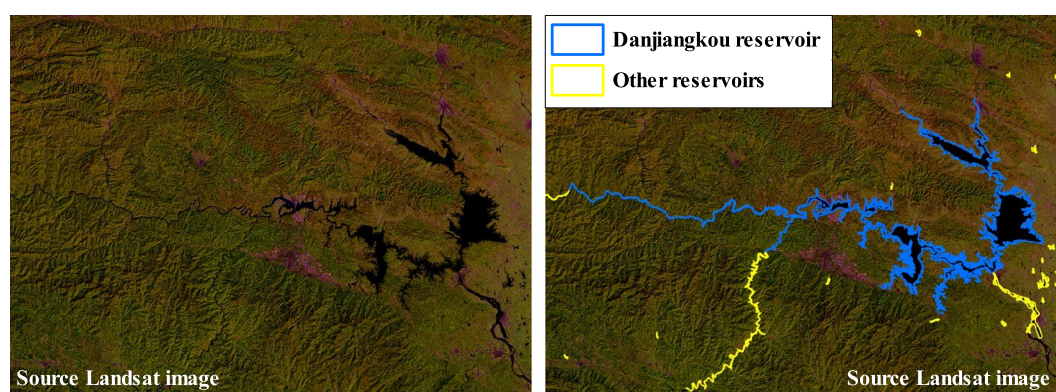


Fig. 9 Demonstration of Danjiangkou reservoir.

most actual reservoir pixels have been correctly extracted with a 98.75% PA in our inventory as well. Due to the limit of Landsat image resolution and pixel mixture issue⁶⁰, some reservoir points were omitted and committed as background objects on the boundary of the reservoirs, as referred to high-resolution Google Earth images in Fig. 8. This issue can be further addressed by adopting a mixed pixel decomposition strategy⁶¹ in future studies.

Usage Notes

By distinguishing the reservoirs based on uniform data sources and methodology, the proposed reservoir inventory dataset is intended for further analysis to explore the impact of anthropogenic activities and climate change on changes in China's water landscape. Additionally, it can be used to investigate the effects of changes in reservoir storage capacity on geological and hydrological hazards. This detailed reservoir inventory provides a fundamental data source to develop a more effective plan for allocating limited water resources in China. And it is believed to be useful for promoting the research of basic disciplines such as hydrology and environmental science, and assisting watershed-scale environmental planning/monitoring and the national macro-decision-making in terms of integrated management of surface water resources.

In addition, unlike current studies that present single temporal inventories, our study proposes a consistent inventory for two different years, allowing the observation of variations in reservoir construction. In response to the regulations on Farmland Water Conservancy proposed by the Chinese government in 2016, which encourage the application of reservoirs for irrigation, our two-phase reservoir inventory has shown the remarkable positive impact of increased accumulated reservoir storage capacity on the expansion of irrigation cropland area. In the face of uncontrollable climatic factors, particularly extreme precipitation or droughts, proposing advanced strategies for higher water use efficiency is more significant in guaranteeing water use. Compared with a single temporal reservoir inventory, the proposed two-phase reservoir inventory can contribute to providing more insight analysis of a wide range of water conservation management-related aspects for better sustainable development.

The inventories in this study cover medium and large reservoirs, which can be easily confused with large rivers or lakes, especially those located closely. Danjiangkou reservoir, spanning across Hubei and Henan provinces, is a typical reservoir constructed beside the Yangtze River (as shown in Fig. 9). The starting location is easy to recognize in the annually composited Landsat image, but the ending boundary is difficult to extract compared to closed reservoirs. In this study, the ending boundary of such reservoirs without exact closed boundaries is

extracted by referring to the official census released from the yearbook through calculating reservoir area and storage capacity and the relative location to nearby reservoirs. Moreover, due to the issue of mixed pixels in the Landsat image, as demonstrated in Fig. 8, the extracted boundaries of reservoirs are difficult to delineate the detailed shapes. This issue can be addressed by improving the inventory based on higher spatial resolution images in future studies.

The names provided in the dataset for each reservoir are collected from public geological datasets and local media, but there are still several names of reservoirs missing. These can be filled or updated by local field survey, which can be time-consuming and expensive to implement. To the best of our knowledge, the names provided in our proposed inventory cover the state-of-the-art largest coverage. This can be recognized as a comprehensive and reliable reservoir dataset for hydrological modeling and improving water management efficiency.

Code availability

The code for extracting names, longitudes and latitudes of reservoirs, ponds, etc. from Baidu web map (<http://map.baidu.com>) can be obtained from <https://github.com/licd54/webMap>. The annual Landsat image composition is implemented on the platform of Google Earth Engine. One of the main tools to visualize the dataset is to use the software ArcGIS Desktop see <https://desktop.arcgis.com/en/arcmap/latest/get-started/installation-guide/installing-on-your-computer.htm>. By using ArcGIS 10.8.2 platform, we distinguished and vectorized the reservoirs' boundaries based on manual interpretation on Landsat 8 images.

Received: 15 May 2023; Accepted: 30 August 2023;

Published online: 09 September 2023

References

- Liu, J. *et al.* Water conservancy projects in China: Achievements, challenges and way forward. *Global Environ. Chang.* **23**(3), 633–643, <https://doi.org/10.1016/j.gloenvcha.2013.02.002> (2013).
- Lehner, B. *et al.* High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Front. Ecol. Environ.* **9**(9), 494–502 (2011).
- Yang, X. & Lu, X. X. Delineation of lakes and reservoirs in large river basins: An example of the Yangtze River Basin, China. *Geomorphology* **190**, 92–102, <https://doi.org/10.1016/j.geomorph.2013.02.018> (2013).
- Chao, B. F., Wu, Y.-H. & Li, Y. Impact of artificial reservoir water impoundment on global sea level. *Science* **320**(5873), 212–214 (2008).
- Wada, Y. *et al.* Recent changes in land water storage and its contribution to sea level variations. *Surv. Geophys.* **38**(1), 131–152 (2017).
- Song, C. *et al.* A comprehensive geospatial database of nearly 100 000 reservoirs in China. *Earth Syst. Sci. Data* **14**(9), 4017–4034 (2022).
- Yang, X. & Lu, X. Drastic change in China's lakes and reservoirs over the past decades. *Sci. Rep.* **4**, 6041 (2014).
- Bruinsma, J. The resource outlook to 2050 in *Expert meeting on how to feed the world*. 1–33.
- He, C., Harden, C. P. & Liu, Y. Comparison of water resources management between China and the United States. *Geogr. Sustain.* **1**(2), 98–108 (2020).
- Udimal, T. B., Jincai, Z., Ayamba, E. C. & Owusu, S. M. China's water situation; the supply of water and the pattern of its usage. *Int. J. Sustain. Built Environ.* 491–500 (2017).
- Tao, S. *et al.* Changes in China's water resources in the early 21st century. *Front. Ecol. Environ.* **18**(4), 188–193 (2020).
- Wu, J. *et al.* The three gorges dam: an ecological perspective. *Front. Ecol. Environ.* **2**(5), 241–248 (2004).
- Zhang, Q. *et al.* Has the Three-Gorges Dam made the Poyang Lake wetlands wetter and drier? *Geophys. Res. Lett.* **39**, 20 (2012).
- Wang, J., Sheng, Y., Gleason, C. J. & Wada, Y. Downstream Yangtze River levels impacted by three Gorges dam. *Environ. Res. Lett.* **8**(4), 044012 (2013).
- Ministry of Water Resources of the People's Republic of China. *China Water Resources Statistical Yearbook*, Vol. 3-4 (China WaterPower Press, Beijing, 2019).
- Jackson, R. B. *et al.* Water in a changing world. *Ecol. Appl.* **11**(4), 1027–1045 (2001).
- Yu, B., Yang, A., Chen, F., Wang, N. & Wang, L. SNNFD, spiking neural segmentation network in frequency domain using high spatial resolution images for building extraction. *Int. J. Appl. Earth Obs. Geoinf.* **112**, 102930, <https://doi.org/10.1016/j.jag.2022.102930> (2022).
- Chen, F., Wang, N., Yu, B. & Wang, L. Res2-Unet, a new deep architecture for building detection from high spatial resolution images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **15**, 1494–1501 (2022).
- Lehner, B. & Doll, P. Development and validation of a global database of lakes, reservoirs and wetlands. *J. Hydrol.* **296**(1-4), 1–22 (2004).
- Vörösmarty, C. *et al.* The aging and storage of reservoir systems large continental runoff of the world. (1997).
- International Commission on Large Dams (ICOLD). *World Register of Dams: General Synthesis*, <http://www.icold-cigb.org> (2011).
- Mulligan, M., van Soesbergen, A. & Sáenz, L. GOODD, a global dataset of more than 38,000 georeferenced dams. *Sci. Data* **7**(1), 31, <https://doi.org/10.1038/s41597-020-0362-5> (2020).
- Wang, J. *et al.* GeoDAR: georeferenced global dams and reservoirs dataset for bridging attributes and geolocations. *Earth Syst. Sci. Data* **14**(4), 1869–1899 (2022).
- Sawunyama, T., Senzanje, A. & Mhizha, A. Estimation of small reservoir storage capacities in Limpopo River Basin using geographical information systems (GIS) and remotely sensed surface areas: Case of Mzingwane catchment. *Phys. Chem. Earth* **31**(15-16), 935–943 (2005).
- Mialhe, F., Gunnell, Y. & Mering, C. Synoptic assessment of water resource variability in reservoirs by remote sensing: General approach and application to the runoff harvesting systems of south India. *Water Resour. Res.* **44**, (2008).
- Annor, F. O., Giesen, N., Liebe, J., Zaag, P. & Odai, S. N. Delineation of small reservoirs using radar imagery in a semi-arid environment: A case study in the upper east region of Ghana. *Phys. Chem. Earth* **34**(4–5), 309–315 (2007).
- Ran, L. & Lu, X. X. Delineation of reservoirs using remote sensing and their storage estimate: an example of the Yellow River basin, China. *Hydrol. Process.* **26**(8), 1215–1229, <https://doi.org/10.1002/hyp.8224> (2012).
- Yu, B., Xu, C., Chen, F., Wang, N. & Wang, L. HADeenNet: A hierarchical-attention multi-scale deconvolution network for landslide detection. *Int. J. Appl. Earth Obs. Geoinf.* **111**, 102853, <https://doi.org/10.1016/j.jag.2022.102853> (2022).
- Wang, X. *et al.* Improved maps of surface water bodies, large dams, reservoirs, and lakes in China. *Earth Syst. Sci. Data* **14**(8), 3757–3771, <https://doi.org/10.5194/essd-14-3757-2022> (2022).
- Chai, X., Shen, Z. & Nie, B. Study on the anti-seepage reinforcement scheme of earth-rock dam for Changli Reservoir. *Adv. Mater.* **904**, 389–392 (2014).
- Aoyama, H. A study of stratified random sampling. *Ann. Inst. Stat. Math.* **6**(1), 1–36 (1954).

32. Zhang, X. *et al.* GLC_FCS30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery. *Earth Syst. Sci. Data* **13**(6), 2753–2776 (2021).
33. Zhang, R., Wu, B., Han, J. & Zuo, Z. in *Climate Change - Realities, Impacts Over Ice Cap, Sea Level and Risks* (ed Singh, B. R.) (IntechOpen, 2013).
34. Zhu, M. *et al.* Population and economic projections in the Yangtze River Basin based on shared socioeconomic pathways. *Sustainability-Basel* **12**, (2020).
35. Liu, L. L. & Du, J. J. Documented changes in annual runoff and attribution since the 1950s within selected rivers in China. *Adv. Clim. Chang. Res.* (2017).
36. Yang, Z. *et al.* Interannual variation and regime shift of the evaporative moisture sources for wintertime precipitation over Southern China. *J. Geophys. Res. Atmos.* (2018).
37. Chen, F., Wang, J., Li, B., Yang, A. & Zhang, M. Spatial variability in melting on Himalayan debris-covered glaciers from 2000 to 2013. *Remote Sens. Environ.* **291**, 113560 (2023).
38. Gao, B. NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **58**, (1996).
39. Tucker, C. Red and photographic infrared linear combination for monitoring vegetation. *Remote Sens. Environ.* **8**, (1979).
40. Yu, B. *et al.* Temporal expansion of the nighttime light images of SDGSAT-1 satellite in illuminating ground object extraction by joint observation of NPP-VIIRS and sentinel-2A images. *Remote Sens. Environ.* **295**, 113691 (2023).
41. Chander, G., Markham, B. L. & Helder, D. L. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sens. Environ.* **113**(5), 893–903 (2009).
42. Potapov, P., Turubanova, S. & Hansen, M. C. Regional-scale boreal forest cover and change mapping using Landsat data composites for European Russia. *Remote Sens. Environ.* **115**(2), 548–561 (2011).
43. Potapov, P. V. *et al.* Quantifying forest cover loss in Democratic Republic of the Congo, 2000–2010, with Landsat ETM+ data. *Remote Sens. Environ.* **122**, 106–116 (2012).
44. Lin, Y. *et al.* Leveraging optical and SAR data with a UU-Net for large-scale road extraction. *Int. J. Appl. Earth Obs. Geoinf.* **103**, 102498 (2021).
45. Mahdianpari, M. *et al.* A large-scale change monitoring of wetlands using time series Landsat imagery on Google Earth Engine: a case study in Newfoundland. *GLSci. Remote Sens.* **57**(8), 1102–1124 (2020).
46. Yu, B., Chen, F., Wang, N., Wang, L. & Guo, H. Assessing changes in nighttime lighting in the aftermath of the Turkey-Syria earthquake using SDGSAT-1 satellite data. *The Innovation* **4**(3), 100419 (2023).
47. Zhang, X., Liu, L., Chen, X., Xie, S. & Gao, Y. Fine land-cover mapping in China using Landsat datacube and an operational SPECLib-based approach. *Remote Sens.* **11**(9), 1056 (2019).
48. Li, C., Chen, F., Wang, N., Yu, B. & Wang, L. SDGSAT-1 nighttime light data improve village-scale built-up delineation. *Remote Sens. Environ.* **297**, 113764 (2023).
49. Turner, S. W. & Galelli, S. Water supply sensitivity to climate change: An R package for implementing reservoir storage analysis in global and regional impact studies. *Environ. Modell. Softw.* **76**, 13–19 (2016).
50. Liu, B. *et al.* Spatiotemporal assessment of water security in China: An integrated supply-demand coupling model. *J. Clean. Prod.* **321**, 128955 (2021).
51. Han, K.-W., Shi, K.-B., Yan, X.-J. & Cheng, Y.-Y. Water savings efficiency of counterweighted spheres covering a plain reservoir in an arid area. *Water Resour. Manag.* **33**, 1867–1880 (2019).
52. Feng, H. *et al.* Study on the integrated protection strategy of water environment protection: The case of Hainan Province of China. *Environ. Technol. Innov.* **24**, 101990 (2021).
53. Wang, S., Zhou, L., Wang, H. & Li, X. Water use efficiency and its influencing factors in China: based on the data envelopment analysis (DEA)—Tobit model. *Water-Sui.* **10**(7), 832 (2018).
54. Zhang, M.-G. *et al.* Using species distribution modeling to improve conservation and land use planning of Yunnan, China. *Biol. Conserv.* **153**, 257–264 (2012).
55. Liu, L., Xu, X. & Chen, X. Assessing the impact of urban expansion on potential crop yield in China during 1990–2010. *Food Secur.* **7**, 33–43 (2015).
56. Sun, Y., Li, J., Wang, H., Li, R. & Tang, X. Extreme rainfall in Northern China in September 2021 tied to air–sea multi-factors. *Clim. Dyn.* 1–15 (2022).
57. Zhang, H., Chen, F., Wang, L., Wang, N. & Yu, B. *Reservoir inventory for China in 2016 and 2021* <https://doi.org/10.5281/zenodo.8278702> (2023).
58. Li, Y., Zhao, G., Allen, G. H. & Gao, H. Diminishing storage returns of reservoir construction. *Nat. Commun.* **14**(1), 3203 (2023).
59. Wang, L., Guan, R., Zhangxiong, L., Chang, R. & Qiu, L. Genetic diversity of Chinese cultivated soybean revealed by SSR markers. *Crop Sci.* **46**(3), 1032–1038 (2006).
60. Lo, C. P. & Choi, J. A hybrid approach to urban land use/cover mapping using Landsat 7 Enhanced Thematic Mapper Plus (ETM+) images. *Int. J. Remote Sens.* **25**(14), 2687–2700 (2004).
61. Chen, Q., Ding, W., Huang, X. & Wang, H. Generalized interval type II fuzzy rough model based feature discretization for mixed pixels. *IEEE Trans. Fuzzy Syst.* (2022).
62. Lu, J. *et al.* Performance of the standardized precipitation index based on the TMPA and CMORPH precipitation products for drought monitoring in China. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **11**(5), 1387–1396 (2018).

Acknowledgements

This work was supported by the National Key R&D Program of China (No. 2022YFC3800701), the Youth Innovation Promotion Association, CAS (2022122), the China-ASEAN Big Earth Data Platform and Applications (CADA, guikeAA20302022). The work was also supported by CAS-TWAS Centre of Excellence on Space Technology for Disaster Mitigation.

Author contributions

B.Y. and H.Z. designed the study. H.Z., N.W., F.C. and L.W. compiled the dataset. B.Y. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.W. or B.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023