Journal of Global Change Data & Discovery. 2025, 9(1): 59-70
www.geodoi.ac.cn

Global Change Research Data Publishing & Repository

Dataset Development on the Temporal and Spatial Distribution of Surface Reservoirs in Xinjiang Uygur Autonomous Region (1942–2022)

Li, S. S.[1,2] Li, J. L.[1,3*] Du, W. B.[1,2] Liu, S. Q.[1,3,4] Wang, H. Y.[1,3,4] Jin, J. Y.[1,3,4]

1. Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China;
2. School of Surveying and Mapping and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China;
3. Key Laboratory of GIS & RS Application, Xinjiang Uygur Autonomous Region, Urumqi 830011, China;
4. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: Obtaining the temporal and spatial distribution of reservoirs in Xinjiang is crucial for analyzing the migration of water resources and the ecological evolution of oases and deserts. Based on the Sentinel-2 data in 2022, this study extracted the spatial distribution range of reservoirs larger than 0.001 km^2 in Xinjiang, and collected the attribute information of reservoirs by integrating sources such as county chronicles and yearbooks, mainly including the name, longitude and latitude, average elevation, completion time, total storage capacity and maximum area, and basin of the reservoir. Based on to the water conservancy project standards and storage capacity, the reservoirs were classified into large, medium and small reservoirs. The results indicate that as of 2022, a total of 804 reservoirs have been constructed in Xinjiang, with a total storage capacity of 24.16 km^3. Of these, there are 37 large, 175 medium and 592 small reservoirs, and 461 plain reservoirs and 343 mountain reservoirs. After 1980, the proportion of large mountain reservoirs has significantly increased. This dataset includes: (1) spatial distribution data of reservoirs in Xinjiang, 1942-2022; (2) reservoir attribute data, which includes records of attributes such as the reservoir name, area, capacity, built year, elevation, location, and river basin, etc. The dataset is archived in .shp and .xlsx formats, and consists of 9 data files with data size of 21.4 MB (compressed into 1 file with 5.86 MB).

Keywords: reservoirs; spatio-temporal changes; Xinjiang; remote sensing

DOI: https://doi.org/10.3974/geodp.2025.01.08

Dataset Availability Statement:

The dataset supporting this paper was published and is accessible through the *Digital Journal of Global Change Data Repository* at: https://doi.org/10.3974/geodb.2025.01.01.V1.

Received: 20-12-2024; **Accepted:** 28-01-2025; **Published:** 25-03-2025

Foundations: Natural Science Foundation of Xinjiang Uygur Autonomous Region (2022D01E18); National Natural Science Foundation of China (U2003201, 41671034)

***Corresponding Author:** Li, J. L. C-8904-2014, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, lijl@ms.xjb.ac.cn

Data Citation: [1] Li, S. S., Li, J. L., Du, W. B., *et al*. Dataset development on the temporal and spatial distribution of surface reservoirs in Xinjiang Uygur Autonomous Region (1942–2022) [J]. *Journal of Global Change Data & Discovery*, 2025, 9(1): 59–70. https://doi.org/10.3974/geodp.2025.01.08.

[2] Li, S. S., Li, J. L., Du, W. B., *et al*. Spatio-temporal dataset of reservoirs in Xinjiang Uygur Autonomous Region, China (1942–2022) [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2025. https://doi.org/10.3974/geodb.2025.01.01.V1.

1 Introduction

Reservoirs are artificial water storage facilities built by damming on rivers to retain water. By storing and releasing water, they enable the spatio-temporal allocation of water resources and play crucial roles in agricultural irrigation, flood control, power generation, and urban water supply[1]. Xinjiang is situated in arid and semi-arid regions with scarce precipitation, limited water resources, and uneven spatio-temporal distribution. Reservoirs are an critical means to ensure sustainable water supply[2]. Xinjiang began constructing reservoirs in the 1940s, and since the founding of the People's Republic of China, the scale of reservoir and water conservancy projects has continuously expanded[2,3]. Mountain control reservoirs, reservoirs in the middle and lower reaches of rivers, and irrigation canals have gradually developed into a multi-objective reservoir joint scheduling system[4], significantly improving the efficiency of water resource utilization, contributing to the expansion of artificial oases and cultivated land. Reservoirs have become a vital water source for the development of oasis agriculture, as well as production and daily life. Analyzing the historical changes in reservoir construction in Xinjiang is of crucial for understanding the development and utilization of water resources, as well as the expansion of artificial oases.

Remote sensing technology is a crucial tool for obtaining reservoir information[5–7]. A number of spatial reservoir datasets based on remote sensing technology have been released domestically and internationally, including the Global Georeferenced Database of Dams (GOODD)[8], the Global Reservoir and Lake Surface Area Dataset (ReaLSAT)[9], the Georeferenced Global Dam and Reservoir Dataset (GeoDAR)[10], the Global Dam Tracker Database (GDAT)[11], the China Lake, Dam, Reservoir, and Large waterbody Dataset (China-LDRL)[12], the China Reservoir Dataset (CRD)[13], and the 2016–2021 China Reservoir List[14]. Li[15] analyzed the characteristics and distribution patterns of reservoir dams in Xinjiang using the most recent and comprehensive reservoir data. However, this dataset only presents statistical data, lacking spatial attributes. Reservoir datasets derived from remote sensing not only include information on reservoir area and spatial location, but also attributes such as capacity, function, and type, which have been instrumental in research on water resource management and allocation[16]. However, most of these datasets contain incomplete descriptions of reservoir information in Xinjiang. For instance, GeoDAR provides information for only 16 reservoirs in Xinjiang, while the CRD dataset includes 673 records and is currently the most comprehensive. However, this dataset is based on published reservoir dam products and does not account for reservoirs constructed in recent years or some smaller reservoirs built in earlier stages. The primary reason for the discrepancy in the number of reservoirs across datasets is that many reservoirs in Xinjiang are small in area and capacity[17,18], leading to missed extractions when using the Landsat 30-m data source. Furthermore, the past decade has seen rapid development in reservoir construction in Xinjiang, and many current datasets lack recent updates on reservoir information.

To this end, this study uses Sentinel-2 images10-m data from the Xinjiang region in 2022 as the data source to extract the most recent reservoir distribution information. Based on this, combined with Landsat series satellite images since 1986 onwards, yearbooks, local chronicles, and other sources, attributes such as the construction time, reservoir capacity, elevation, and area are gathered. The spatio-temporal characteristics of reservoir construction in Xinjiang across different historical periods are then analyzed.

2 Metadata of the Dataset

The metadata of the Spatio-temporal distribution dataset of surface reservoirs in Xinjiang Uygur Autonomous Region (1942–2022)[19] is summarized in Table 1, which includes the name, author, geographical area, year of the dataset, spatial resolution, dataset files, data publisher, sharing service platform, and data sharing policy, etc.

Table 1 Metadata summary of the Spatio-temporal distribution dataset of surface reservoirs in Xinjiang Uygur Autonomous Region (1942–2022)

Items	Description
Dataset full name	Spatio-temporal distribution dataset of surface reservoirs in Xinjiang Uygur Autonomous Region (1942–2022)
Dataset short name	ReservoirXinjiang_1942–2022
Authors	Li, S. S., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, School of Surveying and Mapping and Land Information Engineering, Henan Polytechnic University, 212104020037@home.hpu.edu.cn Li, J. L., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Key Laboratory of GIS & RS Application, Xinjiang Uygur Autonomous Region, lijl@ms.xjb.ac.cn Du, W. B., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, School of Surveying and Mapping and Land Information Engineering, Henan Polytechnic University, dwb@hpu.edu.cn Liu, S. Q., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Key Laboratory of GIS & RS Application，Xinjiang Uygur Autonomous Region, University of Chinese Academy of Sciences, liushuaiqi22@mails.ucas.ac.cn Wang, H. Y., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Key Laboratory of GIS & RS Application，Xinjiang Uygur Autonomous Region, University of Chinese Academy of Sciences, haoyu.wang@ugent.be Jin, J. Y., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Key Laboratory of GIS & RS Application, Xinjiang Uygur Autonomous Region, University of Chinese Academy of Sciences, jinjingyu22@mails.ucas.ac.cn
Geographical region	Xinjiang
Year	1942–2022
Spatial resolution	10 m (existing reservoirs); 30 m (disappeared reservoirs)
Data format	.shp, .xlsx
Data size	5.86 MB (after compression)
Data files	The total list and the spatial distribution of reservoirs in Xinjiang (1942–2022)
Foundations	Natural Science Foundation of Xinjiang Uygur Autonomous Region (2022D01E18); National Natural Science Foundation of China (U2003201, 41671034)
Computing platform	ArcGIS
Data publisher	Global Change Research Data Publishing & Repository, http://www.geodoi.ac.cn
Address	No. 11A, Datun Road, Chaoyang District, Beijing 100101, China
Data sharing policy	(1) *Data* are openly available and can be free downloaded via the Internet; (2) End users are encouraged to use *Data* subject to citation; (3) Users, who are by definition also value-added service providers, are welcome to redistribute *Data* subject to written permission from the GCdataPR Editorial Office and the issuance of a *Data* redistribution license; and (4) If *Data* are used to compile new datasets, the “ten percent principal” should be followed such that *Data* records utilized should not surpass 10% of the new dataset contents, while sources should be clearly noted in suitable places in the new dataset[20]
Communication and searchable system	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS, GEOSS, PubScholar, CKRSC

3 Methods

3.1 Data Sources

The research generates a spatiotemporal distribution dataset of surface reservoirs in Xinjiang (1942–2022) by integrating remote sensing images and yearbook archives. The primary data source is Sentinel-2 imagery from April to October 2022, used to capture the maximum spatial extent of reservoirs, with full coverage of the entire Xinjiang region achieved at least once per month. Additionally, combined with the collected chronicles and yearbooks from all administrative levels in Xinjiang, attributes such as reservoir capacity and construction time are extracted. A time series of reservoir area is generated using Landsat series remote sensing images to supplement the construction time of reservoirs lacking documentation records. The average elevation of reservoirs is derived from SRTM DEM. The data sources and their applications in this study are shown in Table 2.

Table 2 Data sources and applications

Data source	Temporal resolution	Spatial resolution	Count	Applications
Sentinle-2	5 days	10 m	2,392	Obtain the maximum water surface of the reservoir
Chronicles[21–26]	year	/	103	Obtain the reservoir capacity and construction year
Landsat imagery	16 days	30 m	13,056	Obtain the construction time of the reservoir
SRTM DEM	/	20 m	/	Obtain the average elevation of the reservoir

For this study, cloud-free or less cloudy and snow-free remote sensing images were selected, including Sentinel-2 and Landsat series images. Sentinel-2 images[1] were acquired from April to October 2022, with a spatial resolution of 10 m and a revisit period of 5 days, which facilitates the monitoring of small water bodies with higher precision. The Level-1C products used in the study are available from the ESA official website. Landsat images were acquired from 1985 to 2022, including all available images from Landsat 5 TM, Landsat 7 ETM+, Landsat 8 OLI, and Landsat 9 OLI-2, forming a 40-years time series. The L1TP (Collection 2 T1 Level 1 Precision Terrain Corrected) data[2] used have undergone radiation calibration and high-precision orthorectification based on DEM data and are available from the Earth Explorer website of the United States Geological Survey (USGS).

3.2 Methodological Framework

The research and development process of the Spatio-temporal distribution dataset of surface reservoirs in Xinjiang (1942–2022) involves several key steps, including the extraction of the full water body range using a deep learning model, reservoir discrimination, determination of the reservoir construction time based on the long-time series, establishment of the attribute database, and integration of the attribute database with the spatial database (Figure 1).

3.2.1 Extraction of Water Body

U-Net is a widely used water body mapping model[27], particularly effective for mapping small water bodies in high-resolution images[28]. In this study, the U-Net model was employed to generate the water body map of Xinjiang from 2,392 Sentinel-2 images from April to October 2022. Initially, 300 water body samples covering the entire Xinjiang were created, each with dimensions of 1,500 × 1,500 pixels, with 80% used for model training and 20% for validation. The images and corresponding labels were cropped to a size of 256 × 256 pixels. To increase the number of training samples, the images were enhanced 108 times using techniques such as horizontal flipping, transposition, vertical flipping, rotation. The U-Net model was then trained with an initial learning rate of 0.005, a batch size of 8, and 300 epochs per network. The loss function used during training was FocalLoss, which

[1] European Space Agency. https://scihub.copernicus.eu/.
[2] NASA, USGS. https://earthexplorer.usgs.gov/.

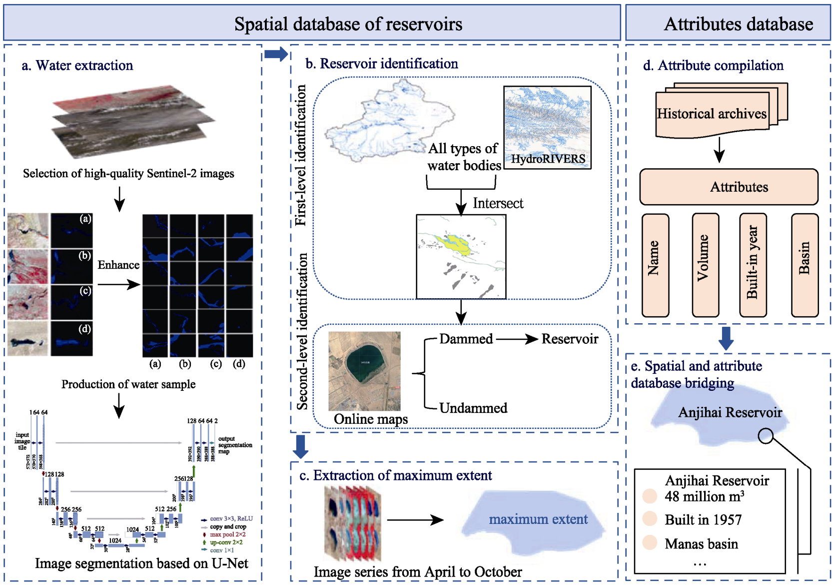

Figure 1 Flowchart of reservoir dataset development

effectively addresses class imbalance in segmentation compared to other classification functions. The U-Net model was executed on a workstation equipped with an NVIDIA Quadro RTX A6,000 graphics card, running under the Cuda framework in Python. The precision, overall accuracy, and recall rates of this model were 93%, 91%, and 88% respectively, demonstrating strong performance.

Based on the trained U-Net model, the area time series of all types of water bodies was generated using all cloud-free Sentinel-2 images from April to October 2022 and subsequently verified through manual editing. Finally, the maximum area was calculated using the area range from the time series. The goal was to uniformly represent its boundary using the maximum extent corresponding to the largest reservoir capacity. Although poor image quality and extensive cloud cover may lead to the omission of the maximum reservoir area, this was overlooked due to the high temporal resolution of Sentinel-2.

3.2.2 Reservoir Identification

Based on the water body extraction results, this study differentiates reservoirs from other water bodies, such as lakes and rivers, using auxiliary data, including river network data and high-resolution base maps. Reservoir identification in this study follows three steps: initial identification based on river network data, visual interpretation using high-resolution online images, and cross-validation with other datasets.

First, a 500-m buffer zone was created based on the high-resolution river network and used as a mask to intersect with the extracted water body data, thereby isolating reservoir elements. Additionally, to avoid overlooking diversion reservoirs located more than 500 m from the river, the study intersected the 500-m buffer zone mask, created from the river network data, with the water body dataset once more. According to the dataset results, only 58 reservoirs were located outside the buffer zone after this step.

Subsequently, the study refined and corrected the initial reservoir screening results using higher-resolution online images. A visual inspection was conducted to assess whether a dam was near the water body polygon. If no dam was found, the water body was classified as a lake or other types. If a dam near was present near the water body, it was added to the dataset as a reservoir.

Finally, the spatial range of reservoirs obtained was cross-validated with existing datasets. The publicly available reservoir dam datasets GeoDAR, China-LDRL, and CRD were selected and spatially matched with the dataset of this study. These datasets were used to supplement the missing spatial records in this study. Additionally, it was confirmed whether the spatial elements identified in this study that were absent in the reference datasets were reservoirs. Ultimately, the exact locations and extents of all reservoirs were manually verified.

3.2.3 Attribute Compilation

Information on characteristics such as reservoir storage capacity, completion time, and catchment area are obtained from local chronicles, yearbooks, literature records, and other sources. The total storage capacity values of reservoirs are primarily derived from yearbooks, county annals, and online search results. The completion times of reservoirs are obtained using 2 methods. First, the completion times of reservoirs are gathered from textual records, with accurate completion years identified for 716. Second, the completion or abandonment times of 81 reservoirs is determined based on Landsat time-series images. It is assumed that the year in which the annual average area of the reservoir water surface exceeds 0 represents the completion year of the reservoir. Information on the completion or abandonment times of only 7 reservoirs cannot be obtained. The elevation values of the reservoirs are derived from SRTM DEM data. The average elevation corresponding to the spatial extent of each reservoir is extracted from the SRTM DEM data and used as the elevation value of the reservoir.

Additionally, for the purpose of statistical analysis, reservoirs are classified according to their capacity and water inflow characteristics. According to the reservoir capacity[29], reservoirs with a capacity greater than 10^8 m^3 are classified as large reservoirs, those with a capacity less than 10^7 m^3 as small reservoirs, and those with a capacity between the two as medium reservoirs. Based on flow characteristics and referencing Chinese mountainous area spatial range data[30,31], reservoirs are classified into two types: mountainous reservoirs and plain reservoirs.

3.2.4 Bridging Attributes and Geolocation

The name of a reservoir is a crucial attribute that links the spatial and the attribute databases, serving as a marker to connect spatial location with attribute information. The names of reservoirs, along with attributes describing their geographical locations, such as administrative regions and associated river basins, are sourced from Google Map or AMAP. The names are then assigned to the reservoir vector polygons at their corresponding spatial locations. Finally, using the reservoir names, a spatial join tool is employed to link the spatial extent with the attribute information, including reservoir name, total storage capacity, maximum area, elevation, completion year, and reservoir type, among others.

4 Data Results and Validation

4.1 Dataset Composition

The dataset is the Spatio-temporal distribution dataset of surface reservoirs in Xinjiang (1942–2022), comprising both spatial distribution data and tabular data. The spatial data includes the distribution data (.shp) of large, medium-sized and small reservoirs in Xinjiang from 1942 to 2022. The tabular data contains the general inventory of Xinjiang Reservoirs (1942–2022) with the attributes.

4.2 Data Results

4.2.1 Spatial Distribution Characteristics of Reservoirs

A total of 804 reservoirs (area > 0.001 km^2) have been constructed in Xinjiang, with a total storage capacity of 24.16 km^3. Among them, 37 are large reservoirs, 175 are medium reservoirs, and 592 are small reservoirs, with storage capacities of 16.583 km^3, 6.025 km^3, and 1.552 km^3 respectively. The statistical results (Figure 2, 3) indicate that in the northern Xinjiang prefectures, Altay, Ili, Tacheng, Changji, Bortala, and Urumqi have constructed 119, 45, 123, 125, 11, and 37 reservoirs, respectively, with corresponding storage capacities of 5.290 km^3, 5.093 km^3, 1.478 km^3, 0.834 km^3, 0.140 km^3, and 0.715 km^3. In eastern Xinjiang, Hami and Turpan have 62 and 15 reservoirs, respectively, with storage capacities of 0.206 km^3 and 0.220 km^3. In southern Xinjiang, Kashgar, Hotan, Aksu, Bayingolin Mongol Autonomous Prefecture, and Kezilesu Kirgiz Autonomous Prefecture have 79, 57, 30, 29, and 23 reservoirs, respectively, with storage capacities of 4.861 km^3, 0.899 km^3, 1.526 km^3, 1.184 km^3, and 1.107 km^3. The prefecture-level cities of Cocodala, Beitun, Karamay, Shuanghe, Aral, Huyanghe, Wujiaqu, Kunyu, and Tiemenguan have constructed a total of 49 reservoirs, with a total storage capacity of 0.608 km^3. Overall, the number of reservoirs in northern Xinjiang exceeds that in southern Xinjiang, and the distribution of reservoirs in northern Xinjiang is more concentrated, primarily in Changji, Bortala, Altay, and Ili. In contrast, reservoirs in southern Xinjiang are mainly distributed near the tributaries of the Tarim River.

The number of reservoirs ineach river basin is ranked as follows: Manas Lake Basin (134), Irtysh River Basin (101), Bogda Region (79), Kunlun Region (60), Ili River Basin (56), Abey Lake Basin (56), Kashgar River Basin (55), Yarkant River Basin (48), Aral Lake Basin (45), Barkol-Yiwu Basin (37), Ulungur River Basin (30), Hami Basin (25), Tarim River Mainstream (22), Ayding Lake Basin (18), Aksu River Basin (13), Lop Nur Region (13), Ugan River Basin (9), and Qarqan River Basin (3). The proportion of small reservoirs is relatively high in each river basin. For example, in the Manas Lake Basin, which has the highest number of reservoirs, the proportion of small reservoirs reaches 71%. The storage capacity of each river basin is ranked as follows: Ili River Basin (5.082 km^3), Irtysh River Basin (4.583 km^3), Yarkant River Basin (4.437 km^3), Manas Lake Basin (2.241 km^3), Kashgar River Basin (1.533 km^3), Kunlun Region (0.987 km^3), Ugan River Basin (0.875 km^3), Tarim River Mainstream (0.836 km^3), Ulungur River Basin (0.803 km^3), Abey Lake Basin (0.637 km^3), Aksu River Basin (0.438 km^3), Lop Nur Region (0.430 km^3), Aral Lake Basin (0.365 km^3), Bogda Region (0.312 km^3), Ayding Lake Basin (0.223 km^3), Qarqan River Basin (0.171 km^3), Hami Basin (0.114 km^3), and Barkol-Yiwu Basin (0.092 km^3). The proportion of storage capacity held by large reservoirs is relatively high in each river basin. In river basins with large runoff, such as the Ili River Basin, Irtysh River Basin, and Yarkant River Basin, the proportion of storage capacity held by large reservoirs exceeds 80%. This is because in these areas with large runoff, large reservoirs possess higher regulation and storage capabilities.

4.2.2 Distribution of Reservoir Construction Years

From 1942 to 2022, the number of reservoirs in Xinjiang has steadily increased. Based on the quantity and characteristics of reservoir construction, this period can be divided into four stages: 1940–1960, 1960–1980, 1980–2000, and 2000–2022. The 1960–1980 and 2000–2010 periods were particularly peak years of reservoir construction. During 1940–1960, a total of 73 reservoirs were built, with a cumulative storage capacity of 1.18 km^3. Small and medium reservoirs accounted for 96% of the total number, but only 68% of the total capacity. In the 1960–1980 period, 292 reservoirs were constructed, with a total

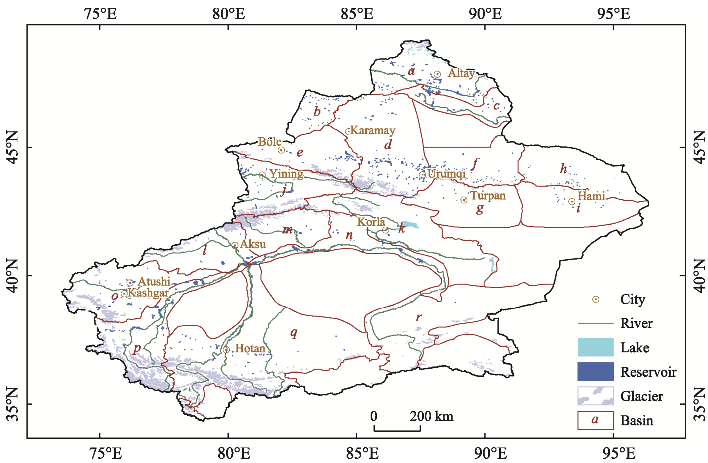

Figure 2 Spatial distribution map of reservoirs in Xinjiang

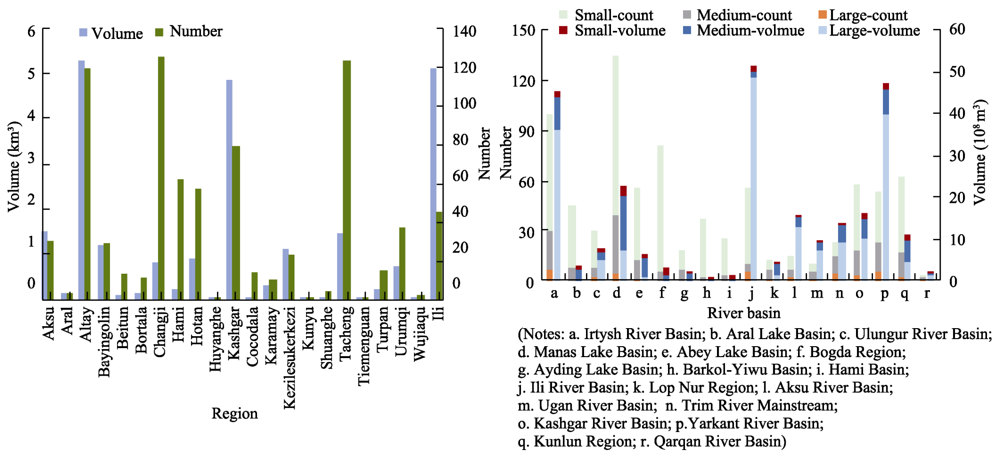

Figure 3 Statistical diagram of the number and volume of reservoirs in Xinjiang

storage capacity of 3.24 km^3. Among these, large, medium, and small reservoirs accounted for 2%, 22%, and 76% of the total number, respectively. The corresponding proportions of storage capacity were 30%, 55%, and 15%. During this period, the construction of small and medium reservoirs still dominated. From 1980 to 2000, 192 reservoirs were built, with a total storage capacity of 3.30 km^3. The proportions of large, medium, and small reservoirs in terms of number were 77%, 20%, and 3%, respectively. In terms of storage capacity, the proportions were 51%, 39%, and 10%, respectively. During 2000–2022, 224 reservoirs were built, including 21 large reservoirs, 54 medium reservoirs, and 149 small reservoirs. Compared with the previous periods, the number of large reservoirs has increased significantly. The total storage capacity during this period is 16.36 km^3, with the proportions of large, medium, and small reservoirs being 83%, 14%, and 3%, respectively. The storage capacity of large reservoirs has also risen considerably. Overall, while the number of reservoirs has not increased dramatically, the total storage capacity has seen significant growth, particularly after 2005. This is mainly due to the construction of large reservoirs. It can be observed that large reservoirs, particularly in mountainous areas, play a crucial role in water regulation and storage. The construction of reservoirs in Xinjiang has gradually shifted towards large reservoirs in these areas. The trend of reservoir construction in Xinjiang over the years is illustrated in Figure 4.

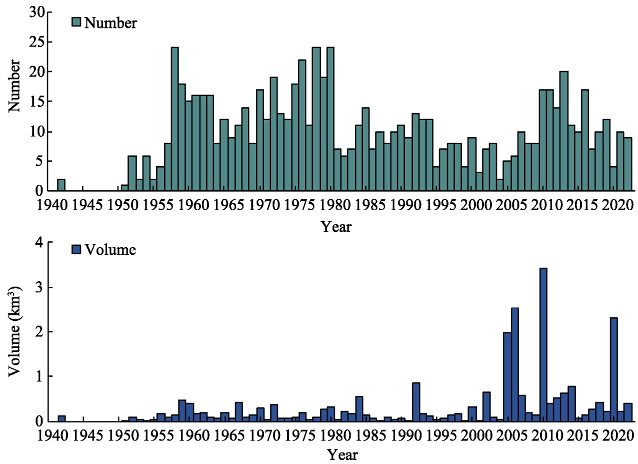

Figure 4 Statistical charts on the number and volume of reservoirs built over the years

4.2.3 Construction of Plains and Mountain Reservoirs

From 1942 to 2022, a total of 461 plain reservoirs and 343 mountainous reservoirs were constructed in Xinjiang, with total storage capacities of 8.59 km^3 and 15.57 km^3, respectively. As shown in the statistical charts (Figure 5), the number of plain reservoirs generally follows a decreasing trend, while the number of mountainous reservoirs shows a steady increase, along with their corresponding storage capacities. Before the 1980s, due to limited construction techniques and economic conditions, the focus was primarily on building plain reservoirs. Between 1940 and 1980, 273 plain reservoirs were constructed, with a total storage capacity of 3.77 km^3. In comparison, only 92 mountainous reservoirs were built, with a total storage capacity of 0.71 km^3. After 1980, the construction of mountainous reservoirs accelerated. From 1980 to 2000, 111 plain reservoirs were built, with a total storage capacity of 1.25 km^3, while 81 mountainous reservoirs were constructed, with a total storage capacity of 1.99 km^3. During this period, the numbers of plain and mountainous reservoirs became more comparable, but the storage capacity of mountainous reservoirs was nearly twice that of plain reservoirs. Since 2000, the construction of mountains reservoirs has far outpaced that of plain reservoirs. In this period, 59 plain reservoirs were built, with a storage capacity of 3.49 km^3, while 165 mountainous reservoirs were constructed, with a storage capacity of 12.87 km^3. Both the number and storage capacity of mountainous reservoirs now significantly exceed those of plain reservoirs.

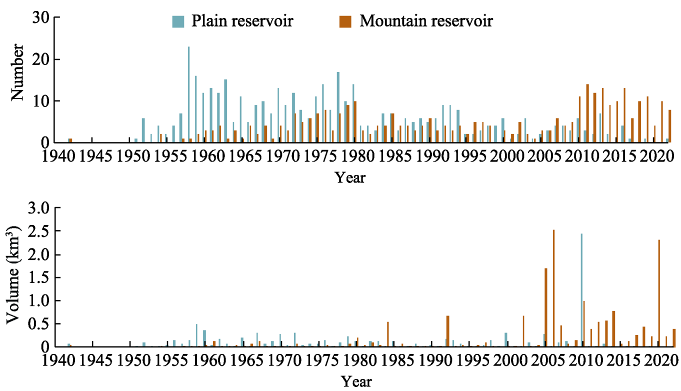

Figure 5 Statistics of the number and volume of plain and mountain reservoirs (1940–2022)

4.3 Comparison with Other Datasets

The spatial distribution and attribute information of reservoirs are essential for effective water resource management and ensuring sustainability. Remote sensing technology has become a key tool for monitoring reservoirs dynamically, both at the global and regional scales. In addition to remote sensing, public statistical reports provide valuable record of reservoir data. In this study, a spatio-temporal distribution dataset of surface reservoirs in Xinjiang was developed by integrating high-resolution remote sensing imagery with records from water conservancy project archives.

Most existing surface water body change datasets conflate reservoirs with other types of water bodies, such as lakes, or fail to separately identify reservoirs, leading to an underestimating of the number of reservoirs in Xinjiang. For instance, GLAKES[28] dataset only labels 19 water body elements as reservoirs. While some inventories specifically focus on reservoirs in Xinjiang, their practical application is often limited due to the lack of geographical references. Li's study[15], which analyzed the spatio-temporal distribution of reservoirs in Xinjiang using statistical data, reported that by 2022, 751 reservoirs had been built, with a total storage capacity of 29.776 km^3. However, the findings of this study indicate that, excluding abandoned or non-functional reservoirs, there are currently 776 operating reservoirs, with a total storage capacity of 23.97 km^3. Although the results are similar, Li's study did not provide spatial locations on these reservoirs. Global reservoir datasets with spatial locations often overlook local reservoirs, resulting in significant discrepancies in the number of reservoirs in Xinjiang. For example, the GeoDAR[10] dataset only identifies 16 reservoirs in Xinjiang, and most of their attribute information is not publicly available. The CRD[13] dataset is the most comprehensive, listing 673 reservoirs with the spatial locations and some attribute like storage capacity and area. However, it lacks information on the construction time of reservoirs, making it difficult to analyze temporal changes in reservoir characteristics. In contrast (Table 3), this study, based on high-resolution remote sensing imagery, provides a more comprehensive map of Xinjiang's reservoirs, visually illustrating their spatial distribution. Furthermore, this dataset integrates records from water conservancy project archives, assigning attributes like storage capacity, area, and elevation to each reservoir, and analyzing the spatio-temporal changes in reservoir characteristics by incorporating the construction time data.

Table 3 Comparison of ReservoirXinjiang_1942–2022 with other datasets

Dataset	Domain	Production time	Number of reservoirs (Xinjiang)	Total volume (Xinjiang)	Attributes
GLAKES[28]	Global	2022	19	/	coordinate, area, water source, type
GeoDAR[10]	Global	2022	16	9.17 km^3 (Accessible)	coordinate, area, capacity, reference data sources
CRD[13]	China	2022	673	30.41 km^3	name, coordinate, prefecture, area, storage, type, shape, length
Li's study[15]	Xinjiang	2022	751	29.78 km^3	distribution, total capacity
ReservoirXinjiang_1942–2022	Xinjiang	2022	776	23.97 km^3	name, coordinate, area, volume, shape length, altitude, built year, river, basin, prefecture, type

5 Discussion and Conclusion

Reservoirs are key indicators of a region's water resource regulation and storage capacity, making their spatial distribution and characteristics essential data for water resource management. In this paper, the extent of reservoirs in Xinjiang was mapped using high-resolution remote sensing images, and an attribute database was established by integrating historical archives. This dataset covers the spatial distribution of reservoirs built in Xinjiang from 1942 to 2022 and illustrates the development of reservoir construction over time by incorporating construction dates. Through data analysis, it was found that since the founding of the People's Republic of China, a total of 804 reservoirs have been built in Xinjiang, with a total storage capacity of 24.16 km^3. Small and medium reservoirs dominate, making up approximately 95% of the total number of reservoirs. However, after 2005, the proportion of large reservoirs has increased. In terms of reservoir types, plain reservoirs predominated before 2010, accounting for 67% of the total. Since 2010, the construction of mountainous reservoirs has accelerated with these reservoirs now accounting for 83% of the total number of reservoirs built during this period.

Compared to existing datasets, the greatest advantage of this dataset is its inclusion of time attribute for reservoirs through two methods: historical records and long-time remote sensing time series. This approach addresses the limitations of single-phase remote sensing, where monitoring the water surface changes of reservoirs can be hindered by insufficient temporal resolution and the quality constraints of remote sensing images. With the provided time attribute, this dataset effectively captures the spatio-temporal changes of reservoirs in Xinjiang and the development of water conservancy, offering valuable insights for water resource utilization and planning. Additionally, by integrating multi source remote sensing images and historical archive materials, the dataset combines the strengths of both data sources, creating a completer and more reliable spatial and attribute database for Xinjiang's reservoirs. The Spatio-temporal distribution dataset of surface reservoirs in Xinjiang (1942−2022) provides comprehensive data support for water resource allocation and enhancing water resource utilization efficiency, while also serving as a reference for future reservoir construction in Xinjiang.

Author Contributions

Li, J. L. and Du, W. B. made the overall design for the dataset development; Li, S. S. and Jin, J. Y. collected and processed data on the max extent and attributes of reservoirs in Xinjiang; Li, S. S. and Wang, H. Y. designed the models and algorithms; Li, S. S. and Liu, S. Q. completed the data validation; Li, S. S. and Li, J. L. wrote the paper.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Chinese Academy of Engineering, "Research on water resources strategy for sustainable development of China in the 21st century" project team. Comprehensive report on China's sustainable development water resources strategy [J]. *China Water Resources*, 2000(8): 5–17.

[2] Tang, Q. C. The development in oases and rational use of water resources [J]. *Journal of Arid Land Resources and Environment*, 1995, 9(3): 107–112.

[3] Deng, M. J. "Three Water Lines" strategy: its spatial patterns and effects on water resources allocation in northwest China [J]. *Acta Geographica Sinica*, 2018, 73(7): 1189–1203.

[4] Liu, T. Research on joint-dispatching of mountain-plain reservoirs in the Yarkant River Basin [J]. *Hydro Science and Cold Zone Engineering*, 2023, 6(4): 68–71.

[5] Perin, V., Tulbure, M. G., Gaines, M. D., *et al*. A multi-sensor satellite imagery approach to monitor on-farm reservoirs [J]. *Remote Sensing of Environment*, 2022, 270: 112796.

[6] Vu, D. T., Dang, T. D., Galelli, S., *et al*. Satellite observations reveal 13 years of reservoir filling strategies, operating rules, and hydrological alterations in the Upper Mekong River basin [J]. *Hydrology and Earth System Sciences*, 2022, 26(9): 2345–2364.

[7] Liu, C., Tang, H. R., Ji, L. Y., *et al*. Spatial-temporal water area monitoring of the Miyun Reservoir using remote sensing imagery from 1984 to 2020 [J]. *National Remote Sensing Bulletin*, 2023, 27(2): 335–350.

[8] Mulligan, M., van Soesbergen, A., Sáenz, L. GOODD, a global dataset of more than 38,000 georeferenced dams [J]. *Scientific Data*, 2020, 7(1): 31.

[9] Khandelwal, A., Karpatne, A., Ravirathinam, P., *et al*. ReaLSAT, a global dataset of reservoir and lake surface area variations [J]. *Scientific Data*, 2022, 9(1): 356.

[10] Wang, J., Walter, B. A., Yao, F., *et al*. GeoDAR: georeferenced global dams and reservoirs dataset for bridging attributes and geolocations [J]. *Earth System Science Data*, 2022, 14(4): 1869–1899.

[11] Zhang, A. T., Gu, V. X. Global Dam Tracker: a database of more than 35,000 dams with location, catchment, and attribute information [J]. *Scientific Data*, 2023, 10(1): 111.

[12] Wang, X., Xiao, X., Qin, Y., *et al*. Improved maps of surface water bodies, large dams, reservoirs, and lakes in China [J]. *Earth System Science Data*, 2022, 14(8): 3757–3771.

[13] Song, C., Fan, C., Zhu, J., *et al*. A comprehensive geospatial database of nearly 100,000 reservoirs in China [J]. *Earth System Science Data*, 2022, 14(9): 4017–4034.

[14] Zhang, H., Chen, F., Wang, L., *et al*. Reservoir inventory for China in 2016 and 2021 [J]. *Scientific Data*, 2023, 10(1): 609.

[15] Li, J., Liu, Y., Peng, Z. X., *et al*. Characteristics and construction managementideas of reservoir dams in Xinjiang [J]. *China Water Resources*, 2023(16): 41─46.

[16] Lehner, B., Liermann, C. R., Revenga, C., *et al*. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management [J]. *Frontiers in Ecology and the Environment*, 2011, 9(9): 494–502.

[17] Li, J., Liu, Y., Wu, T., *et al*. 70 years of dam construction in Xinjiang [J]. *Journal of China Institute of Water Resources and Hydropower Research*, 2020, 18(5): 322–330.

[18] Deng, M. J., Yu, H. M., Li, X. Q., *et al*. Dam construction progress in Xinjiang [J]. *Water Resources and Hydropower Engineering*, 2010, 41(7): 29–35.

[19] Li, S. S., Li, J. L., Du, W. B., *et al*. Spatio-temporal dataset of reservoirs in Xinjiang Uygur Autonomous Region, China (1942–2022) [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2025. https://doi.org/10.3974/geodb.2025.01.01.V1.

[20] GCdataPR Editorial Office. GCdataPR data sharing policy [OL]. https://doi.org/10.3974/dp.policy.2014.05 (Updated 2017).

[21] Fuyun County Party History Local Chronicles Compilation Committee. Fuyun County Chronicle [M]. Urumqi: Xinjiang People's Publishing House, 2003.

[22] Qiemo County Local Chronicles Compilation Committee. Qiemo County Chronicle [M]. Urumqi: Xinjiang People's Publishing House, 1996.

[23] Kuitun City Local Chronicles Compilation Committee. Kuitun City Chronicle [M]. Beijing: Zhonghua Book Company, 1999.

[24] Kashgar Local Chronicles Compilation Committee. Kashgar Prefecture Chronicles [M]. Urumqi: Xinjiang People's Publishing House, 2004.

[25] Korla City Local Chronicles Compilation Committee. Korla City Chronicle 1991–2005 [M]. Urumqi: Xinjiang People's Publishing House, 2017.

[26] Xinjiang Aksu River Basin Management Office. Chronicles of the Aksu River Basin [M]. Beijing: Local Records Publishing House, 2006.

[27] Zhao, X., Wang, H., Liu, L., *et al*. A method for extracting lake water using ViTenc-UNet: taking typical lakes on the Qinghai-Tibet Plateau as examples [J]. *Remote Sensing*, 2023, 15(16): 4047.

[28] Pi, X., Luo, Q., Feng, L., *et al*. Mapping global lake dynamics reveals the emerging roles of small lakes [J]. *Nature Communications*, 2022, 13(1): 5777.

[29] Ministry of Water Resources of the People's Republic of China. Standard for rank classification and flood protection criteria of water and hydropower projects (SL 252—2017) [S]. Beijing: China Water & Power Press, 2017.

[30] Deng, W., Li, A. N., Nan, X., *et al*. Digital Mountain Map of China [M]. Beijing: China Map Publishing House, 2015.

[31] Nan, X., Li, A. N., Deng, W. Dataset of "Digital mountain map of China" (2015) [DB/OL]. A Big Earth Data Platform for Three Poles, 2022.