



OPEN

DATA DESCRIPTOR

A high-resolution dataset of water bodies distribution over the Tibetan Plateau

Zhengchao Chen^{1,9}, Linan Guo^{2,3,9}, Yanhong Wu^{2,4}, Bing Zhang^{5,6}✉, Pan Chen⁷, Xuan Yang⁸ & Jiawei Guo²

Water body (WB) extraction is the basic work of water resources management. Tibetan Plateau is one of the largest alpine lake systems in the world. However, research on the characteristics of water bodies (WBs) is mainly focused on large and medium WBs due to spatial resolution. This research presents a dataset containing a 2-m resolution map of WBs in 2020 based on Gaofen-1 data, and morphometric and landscape indices of WBs across the Tibetan Plateau. The Swin-UNet model is well performed with overall accuracy at 98%. The total area of WBs is 56354.6 km² across Tibetan Plateau in 2020. The abundance compared with that from size-abundance relationship indicate WBs in the Tibetan Plateau conformed to the classic power scaling law. We evaluate the influence of spatial-resolution in WB extraction, which shows the dataset could be valuable to fill the gap of existing WBs map, especially for small waters. The dataset is valuable for revealing the spatial patterns of WBs, and understanding the impacts of climate change on water resources in Plateau.

Background & Summary

Terrestrial water bodies (WBs), such as lakes, ponds, and reservoirs, are essential components of the hydrological and biogeochemical water cycles¹, which provide essential ecosystem services for human society, such as river flow, biodiverse habitats, fisheries, and supplying irrigation water^{2,3}. Monitoring the dynamic changes of WBs provides important information on understanding changes of the surrounding regions^{3,4}.

Understanding the abundance and size distribution of global or regional WBs has been a persistent effort for several years. Traditionally, this information comes from map compilations¹ and statistical extrapolations based on abundance-size relationships⁵⁻⁷. However, map compilation tends to underrepresents small WBs⁸, while statistical extrapolations of abundance likely overestimate abundance of small WBs⁹⁻¹¹. The morphology of WBs can quantitatively describe the geometric features of water landscapes, such as water area, depth, shoreline length, shoreline development index¹². Morphological characteristics of WBs influence the ecological functionality in a region and is hard to be obtained. Benefiting from the comprehensive information of high-resolution satellite imagery, it is possible to accurately resolve the abundance, size distribution and morphological characteristics of WBs at large scale^{13,14}.

Water body extraction is the basic work of water resources management¹⁵. The extraction of WBs in large scale from remote sensing images can be considered as a target detection process¹⁶, which mainly includes single-band density slicing¹⁷, spectral water indexes^{18,19}, object-oriented approaches, and deep learning methods²⁰. There are several researches for regional or global WBs extraction in past decades. The global surface water dataset (GSWD) developed by Pekel *et al.*²¹ describes the changes of WBs from 1984 at 30 m resolution²¹ and updated to 2022 now. Verpoorter *et al.*⁵ inventoried the world's lakes larger than 0.002 km² in size, including

¹State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, 100094, China. ²International Research Center of Big Data for Sustainable Development Goals, Beijing, 100094, China. ³China University of Mining & Technology-Beijing, Beijing, 100083, China. ⁴Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, 100094, China. ⁵Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, 100094, China. ⁶University of the Chinese Academy of Sciences, Beijing, 100049, China. ⁷Center for Geo-Spatial Information, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, 518055, China. ⁸China Remote Sensing Satellite Ground Station, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, 100094, China. ⁹These authors contributed equally: Zhengchao Chen, Linan Guo. ✉e-mail: zb@radi.ac.cn

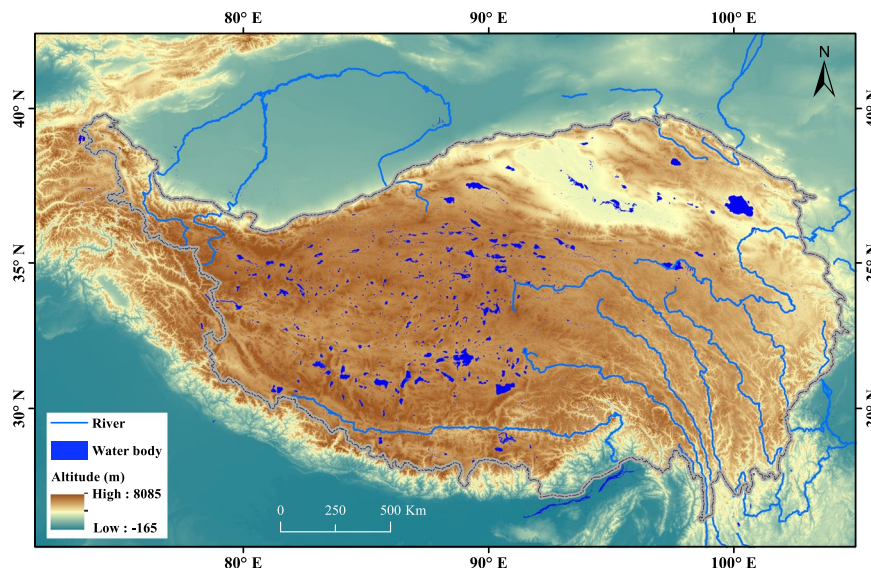


Fig. 1 Study area.

the information of abundance, size (i.e., area and perimeter), geographical distribution, elevation, and morphometric characteristics such as the shoreline development index (SDI)⁵. Spatial and temporal changes of inland WBs in China were investigated by Ma *et al.*² and Zhang *et al.*²², respectively. Feng *et al.*²³ used the GSWD dataset to report that previous studies²⁴ underestimate the abundance and area of WBs (>1 km²) in China. Besides surface water bodies, dataset of large dams and reservoirs was also generated by Wang *et al.*²⁵.

Although the overall global patterns of water body changes have been analysed, regional analyses are sorely needed, especially for small waters at regions that are sensitive to climate change. Recently, deep learning received widespread attention for water bodies recognition. Compared to the traditional machine learning methods, deep learning relies heavily on large-scale training data²⁶. Transfer learning is an emerging method that is applicable when the training data is limited. Fine-tuning a pretrained CNN (Convolutional Neural Network) model may be an effective strategy for many deep learning model applications. At present, high-resolution water body extraction based on deep learning method is mainly implemented at local-scale^{27,28}. One study from Fang *et al.*¹⁶ extracted man-made reservoirs from Landsat-8 images based on a CNN model, ResNet-50 globally¹⁶. Tibetan Plateau (Fig. 1), on where are more than 1,100 alpine lakes²⁹ with area larger than 1 km², receive much attentions because of its less effect by human activities. At present, the extraction of water bodies mainly considers lakes larger than 1 km² on the Tibetan Plateau²³. However, due to the melting of glaciers, the abundance of small water bodies on the Tibetan Plateau will continue to increase³⁰, which is still unknown for us. It is better to obtain this information rely on very high spatial resolution remote sensing images. In addition, although there are several medium and high-resolution images, it is still unclear in data source selection with appropriate resolution for different objectives.

In this study, a 2m-resolution map of water bodies on the Tibetan Plateau is produced based on visual transformer model from Gaofen-1 data. Morphological and landscape indices of WBs are included in the dataset. At the same time, we compare the WBs extraction from different resolutions, which helps analyze the influence of spatial resolution on extraction of water body at different size. The dataset could be valuable for accessing the spatial patterns of WBs, testing the validity of controversial power scaling law for the size-abundance relationship, and selecting data source for water body extraction on the Tibetan Plateau.

Methods

Data. Gaofen-1 (GF-1) is the first of the Gaofen series satellites, which was launched on April 26, 2013. The GF-1 satellite is equipped with one 2-m-resolution panchromatic sensor and one 8-m-resolution multispectral sensor. It also has four 16-m-resolution wide-field-of-view (WFOV) multispectral sensors. The GF-1 satellite is suitable for surface water distribution analysis. Panchromatic and multi-spectral images in 2020 were used in this study. Before water extraction, we used the pansharp fusion method to fuse the panchromatic images and multi-spectral images to generate the images with a spatial resolution of 2 m and four bands.

Water body extraction based on deep learning. To extract water bodies precisely over a wide range and multiple time periods, this study trained Swin-UNet³¹ network based on samples from rapid sample generation technique, combining numerous data augmentation strategies, ultimately achieving the recognition of water bodies over the Tibetan Plateau. The process of extraction algorithm is shown as Fig. 2.

As a data-driven algorithm, the performance of deep learning algorithms is greatly influenced by the quality and quantity of samples. In this study, we designed a rapid sample generation method based on semi-supervised principles (left part in Fig. 2). This process began with a small number of manually labelled samples to train a simple water body recognition network (approximately 800 samples in the size of 512 × 512, in which 100

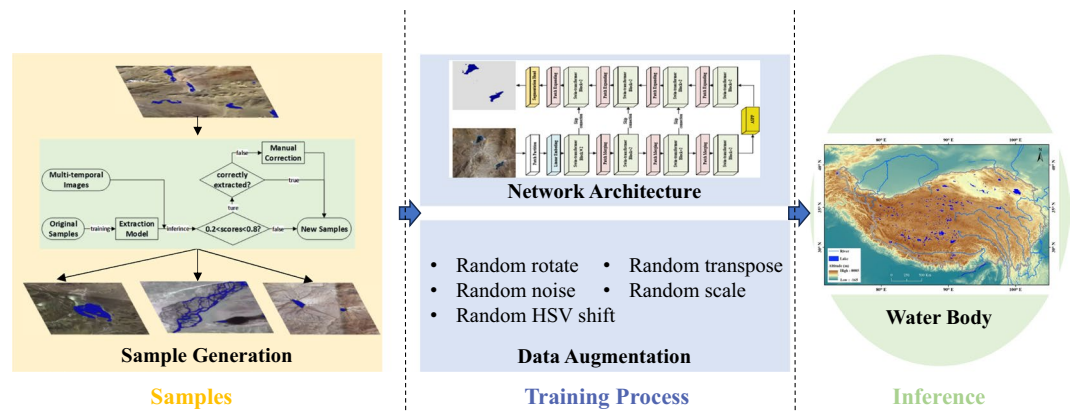


Fig. 2 The flowchart of water body extraction algorithm.

samples contained water bodies). A threshold-based approach, supplemented by manual inspection and correction, was employed to quickly clean the inference results. The cleaned inference results were then reintegrated into the training samples as the new samples. In the process, areas with confidence scores greater than 0.8 or less than 0.2 were considered correctly classified, while misclassifications were inspected and corrected manually. After several rounds of sample generation, totally around 6000 samples in the size of 512×512 were obtained, comprising 778 samples containing water bodies and the rest being totally backgrounds. These samples were divided into training, validation, and test sets in a ratio of 7:1:2.

Because of large-scale variations and strong spatial continuity of water bodies, Swin-UNET network is suitable in this study with large receptive field and strong spatial modelling capabilities. Additionally, multi-spectral information encoding was introduced to make full use of the spatial information and spectral information of water body. In order to make full use of deep semantic information while preserving spatial information, the features from the encoder and decoder interacted through skip connections³². In the bridge between the encoder and decoder, an Atrous Spatial Pyramid Pooling (ASPP) module³³ was introduced to further extract image texture features under multiple receptive fields. In the encoding phase, the input image was encoded through Patch Partition and Linear Embedding. In the decoding phase, the multi-scale features were decoded through multiple Swin-transformer Blocks and Patch Expanding. In the Swin-transformer Block, the input features were first normalized by Layer Normalization (LN). Then, shifted window-based multi-self-attention (SW-MSA) was used to model global features within a window. Subsequently, the outputs of SW-MSA were added to the input features and normalized again with LN. Finally, the normalized features were fed into a simple Multilayer Perceptron (MLP). In the study, all Swin-transformer Blocks were used in pairs, utilizing two stacked SW-MSA with shifted windows to capture the global receptive field across the entire image.

To enhance the model's adaptability to images captured under different regions and imaging conditions in the inference process, several data augmentation strategies were applied during the training process. Including: (1) Random HSV (Hue, Saturation, Value) jittering: Randomly converting the image to the HSV color space and adding jittering ($-30 < H < 30$, $-15 < S < 15$, $-30 < V < 30$) to simulate a broader range of color variations; (2) Random Gaussian noise addition: Randomly adding Gaussian noise to the sample images to simulate images with different levels of noise (mean value = 0 and variance < 50); (3) Random rotation: Randomly rotating the images by a certain angle (-180° to 180°) to simulate observations of objects from different directions; (4) Random scaling: Randomly scaling the images by a certain factor (0.9 to 1.1) to simulate variations in image quality and sharpness under different shooting conditions; (5) Random flipping: Randomly performing horizontal (left-right) mirror flips on the sample data to simulate image data with different spatial arrangements of objects. All the data augmentation algorithms used in this study were implemented based on the open-source library "albumentations"³⁴. All data augmentation functions have a probability of 50% to be applied.

Indices to evaluate spatial variation of water bodies. Besides water abundance and water area, morphometric indices including the shoreline perimeters (SP), and shoreline development index (SDI) were calculated based on high-resolution water body map. The SDI reflects the degree of shoreline irregularity. The more irregular of shoreline, the more habitat diversity the lake can provide for the coastal zone (SDI = 1 when the water body is circle). In addition, Water bodies are important landscape. Here we obtained the landscape pattern indices to understand the geographical significance of water bodies morphological characteristics and distribution rules. Overall, 9 indices were considered in this study (Table 1).

Data Records

The map and statistic indices data of inland water bodies across Tibetan Plateau in 2020 is archived and openly accessible at Figshare³⁵ via the link: <https://doi.org/10.6084/m9.figshare.24616491.v2>. Table 2 shows the details of dataset. 2675 tiles (GeoTIFF format, 16784×16784 pixels) are compressed into the Tibet_water_2020_2m.rar file, which is the water bodies distribution with 2-m spatial resolution across Tibetan Plateau in 2020. The value is 11 for pixels classified as water. The file name of the tiles referred to the Google zoom level. Description about these tiles is shown in the TIFFlist.csv, including the longitude and latitude of top-left corner of each tile.

Indices	Full name	Description	Equation
Morphometric indices	Abundance	Number of water bodies	
	Area	Area of water body	
	Shoreline perimeters	Shoreline perimeter of water body	
	Shoreline development index	Reflects the degree of shoreline irregularity	$SDI = \frac{SP}{2\sqrt{\pi A}}$
Landscape indices	Patch density	Number of patches per unit area	$PD = \frac{n_i}{A}(1000000)(100)$
	Largest patch index	The percent of the total landscape that is made up by the largest patch	$LPI = \frac{\max(a_{ij})}{A} \times 100$
	Landscape shape index	The complexity of the shape	$LSI = \frac{0.25 * \sum_{j=1}^n P_{ij}}{\sqrt{A}}$
	Splitting index	The number of patches if all patches the landscape would be divided into equally sized patches	$SPLIT = \frac{A^2}{\sum_{j=1}^n a_{ij}^2}$

Table 1. Indices metrics.

File name	Description	
Tibet_water_2020_2m.rar	Map of water bodies across Tibetan Plateau in 2020 with 2-m resolution, including 2675 tiles (GeoTIFF, 16784×16784 pixels). Pixel value: 11-water, 0-others	
TIFFlist.csv	Longitude and latitude of top-left corner for each tile	
Indices.csv	Abundance	Water abundance of different size of water bodies
	Water area	Water area of different size of water bodies
	Shoreline perimeter	Perimeter of different size of water bodies
	Shoreline development index (SDI)	Reflects the degree of shoreline irregularity of different size of water bodies
Landscape.csv	PD	Patch density
	LPI	Largest patch index
	LSI	Landscape shape index
	SPLIT	Splitting index

Table 2. Files and formats of the dataset.

The statistic results of water abundance, water area, and morphometric indices are shown in Indices.csv file, while four landscape indices (patch density, largest patch index, landscape shape index, and splitting index) are given in Landscape.csv file.

Technical Validation

Quality control. The dataset was produced with strict quality control. To ensure the quality of samples, we inspected the misclassification water bodies manually in the rapid sample generation process. The Swin-UNet model is good performance which is evaluated by 5 indices (Table 3). OA, P and R could assess the precise of water extraction; F1 score is a comprehensive evaluation of the performance of water detection model; IOU is used to reflect the overlap of the truth and prediction region. TP (True positives) indicates pixel number that correctly detect water, FP (False positives) is the pixel number that incorrectly identified as water, TN (True negatives) indicates the pixel number correctly identified as non-water, while FN (false negatives) is the pixel number that incorrectly identified as non-water. Results (Table 3) show that the water extraction algorithm is an accurate method to detect water bodies in high-resolution remote sensing images with overall accuracy at 98%. The IOU is relatively low with 68%, which may result from the small covering proportion of water bodies in an image.

Then the extracted water bodies have been manually corrected based on visual interpretation. Before indices calculation, morphological opening-and-closing operation was employed. We first filled the small holes inside the water using closing operation to ensure the integrity of the target area, and then remove isolated small pixels using opening operation to ensure the minimal noise of the image. Ellipsoidal area and perimeter instead of Projection one should be used in QGIS software to ensure the correct statistic results. Due to the large study area and limited satellite passing time, a lot of winter images were used in the study, resulting in the extracted water area and abundance being smaller than that from wet season. The water distribution dataset has been compared also with other dataset with lower spatial resolution as described in detail below.

Comparison of morphometric indices of WBs. There is little *in-situ* observation of number or area of WBs in large scale. Conventionally, WBs over large areas are characterized using 1 or a few snapshots of remotely sensed images. We herein compare our morphometric indices dataset with existing research. In our estimation, the abundance of water bodies (>0.01 km²) in the Tibetan Plateau is 96369, and the total area of these water bodies is 56354.6 km² (Table S1 in Supplementary). The total area of the WBs larger than 1 km² is 51034.6 km². According to Zhang *et al.*³⁶, until 2018, there are 1424 lakes larger than 1 km² in the Tibetan Plateau with total area of $5 \times 10^4 \pm 791.4$ km². In addition, our estimation of the WBs larger than 1 km² in the Tibetan Plateau is much

Index	Formula	Value
Overall Accuracy (OA)	$OA = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$	98%
Precision (P)	$P = \frac{TP}{TP + FP} \times 100\%$	82%
Recall (R)	$R = \frac{TP}{TP + FN} \times 100\%$	79%
F1 score	$F1\ score = 2 \times \frac{(P \times R)}{(P + R)} \times 100\%$	81%
Intersection over Union (IOU)	$IOU = \frac{TP}{TP + FP + FN} \times 100\%$	68%

Table 3. Precision of SwinUNet model.

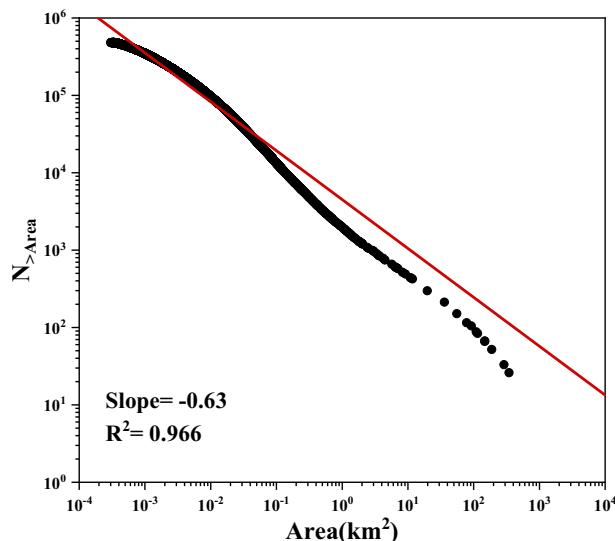


Fig. 3 Log-abundance log-size plot of Water body size distribution.

higher than Mao *et al.* (46,264.5 km²)³⁷ or Wan *et al.* (41,831 km²)³⁸. This difference also demonstrates the results from previous studies^{36,39} that the WBs are expanding in the Tibetan Plateau.

Morphometric and landscape indices are included in our dataset. The statistical results are shown in Table S1, Table S2 and Figure S1 in Supplementary suggesting that small WBs (<1 km²) account for a large proportion of WBs in the Tibetan Plateau and are more separated. However, due to the limitation of water extraction algorithms and spatial resolution of remote sensing data, previous research is still lacking in understanding the morphometric and landscape characteristics of small WBs. The size-abundance relationships were used to estimate the amount of WBs in large scale^{10,40,41}. The size-abundance relationships conform to the power law¹⁰ based on the Pareto distribution probability density function.

$$N = c \times A^{-b} \quad (1)$$

where N is the number of water bodies greater than or equal to the area A , c is a constant, $b = D/2$, where D is the fractal dimension of the shorelines surrounding the water body area and is constrained between $D = 1$ (a population of perfectly smooth shorelines) and $D = 2$ (a population of shorelines so irregular they are space filling). The fractal dimension of size-abundance is supposed to be similar to the shoreline fractal dimension derived from dimensional analysis²³. For WBs on the Tibetan Plateau, D is 1.263 (with $R^2 = 0.966$), and Fig. 3 suggests that distribution deviates slightly from a true power law at WBs with larger area.

Comparison of Water body extraction in different spatial-resolution. We further compared the dataset against two existing datasets with spatial resolution at 30 m and 10 m, respectively. The comparison analysis aims not only at the validation for our dataset, but to analyze the influence of spatial resolution and select applicable data source for water body extraction on the Tibetan Plateau. The European Space Agency (ESA) WorldCover is a land cover map that provides a new baseline global land cover product at 10 m resolution based on Sentinel-1 and 2 data that was developed and validated in almost near-real time and at the same time maximizes the impact and uptake for the end users. The Global Surface Water Explorer (GSWE) dataset was developed by the European Commission's Joint Research Centre based on Landsat satellite images at 30-metre resolution²¹. The dataset maps the location and temporal distribution of water surfaces at the global scale during 1984 to 2022 at monthly and yearly, and provides statistics on the extent and change of those water surfaces.

Figure 4 shows the comparison of morphological indices extracted at different resolutions. For water bodies larger than 1000 km², the estimated number and area of WBs from data at different resolutions are similar.

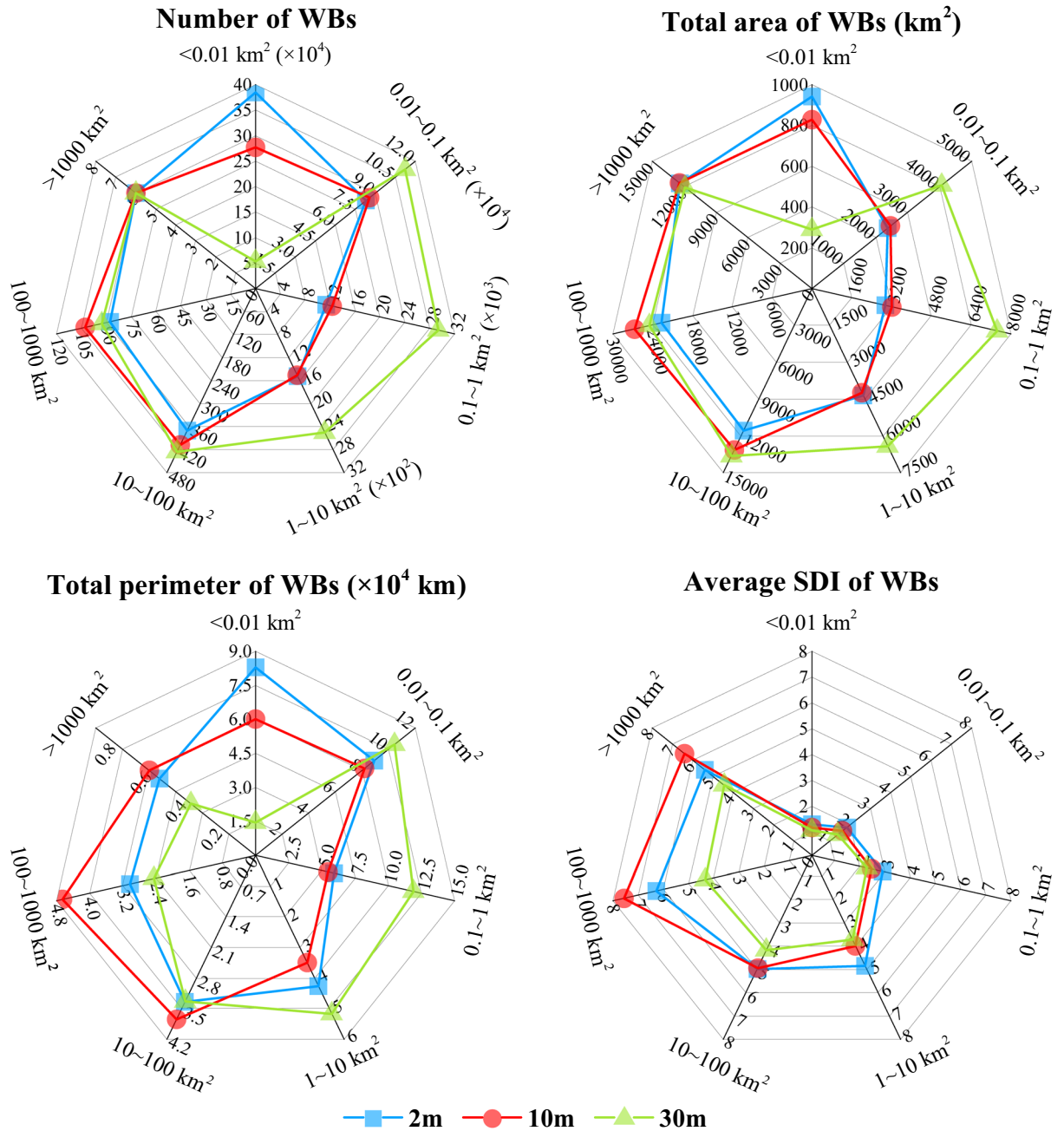


Fig. 4 Water bodies distribution and morphometric characters based on different spatial-resolution.

However, there is a significant inconsistency in perimeter estimation for large water bodies, which may result from the influence of coarse resolution of 30-m data. There are limitations to estimate the distribution of small water bodies for 30 m dataset, especially WBs smaller than 0.01 km^2 . The number and area of WBs larger than 0.01 km^2 from 10m-resolution data is in agreement with that from 2m-resolution dataset. The influence of resolution on perimeter estimation is greater than that on area estimation, thus affecting the estimation of shoreline shape. Thus, the shoreline development index (SDI) increases with the increase of size of WBs. The results showed the estimation of morphological characters based on different spatial-resolution are more consistent for water bodies range from 10 to 100 km^2 . Then, the D of WBs in different spatial-resolution was calculated. D of WBs from 10 m images is 1.24 with $R^2 = 0.949$, while the D of WBs from 30 m images is 1.36 with $R^2 = 0.919$. The results indicates that the size-abundance relationships are more conform to the power law distribution when the resolution is higher. In addition, the D of WBs from 10-m data is close to that from 2-m data. Thus, the 10-m images could obtain approximative results with the 2-m data when only abundance estimation needed.

In the Tibetan Plateau, although small WBs are not dominant in surface, it does not preclude small WBs from significance in regional biogeochemical cycles⁴². Small WBs typically have higher fluxes and faster reaction rates than large lakes and consequently may still contribute disproportionately to biogeochemical cycles of

lake-rich regions⁴³. Our dataset could be valuable to fill the gap of existing water bodies map and analyze the spatial variation of water abundance and shapes, especially for small WBs.

Code availability

Codes for the dataset pre-processing are written using python, including TIFF read, morphological opening-and-closing operation, TIFF write and mosaic process. The codes are available at: <https://github.com/Siyu1993/WaterPreprocessing>. Then the image could be visualized in QGIS software (V3.16).

Received: 29 January 2024; Accepted: 22 April 2024;

Published online: 04 May 2024

References

- Lehner, B. & Döll, P. Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology* **296**, 1–22, <https://doi.org/10.1016/j.jhydrol.2004.03.028> (2004).
- Ma, R. H. *et al.* China's lakes at present: Number, area and spatial distribution. *Sci. China-Earth Sci.* **54**, 283–289, <https://doi.org/10.1007/s11430-010-4052-6> (2011).
- Zou, Z. H. *et al.* Divergent trends of open-surface water body area in the contiguous United States from 1984 to 2016. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 3810–3815, <https://doi.org/10.1073/pnas.1719275115> (2018).
- Piao, S. L. *et al.* The impacts of climate change on water resources and agriculture in China. *Nature* **467**, 43–51, <https://doi.org/10.1038/nature09364> (2010).
- Verpoorter, C., Kutser, T., Seekell, D. A. & Tranvik, L. J. A global inventory of lakes based on high-resolution satellite imagery. *Geophys. Res. Lett.* **41**, 6396–6402, <https://doi.org/10.1002/2014gl060641> (2014).
- Hamilton, S. K., Melack, J. M., Goodchild, M. F. & Lewis, W. M. Estimation of the fractal dimension of terrain from lake size distributions. (1992).
- Newman, M. E. J. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* **46**, 323–351, <https://doi.org/10.1080/00107510500052444> (2005).
- Downing, J. A. *et al.* The global abundance and size distribution of lakes, ponds, and impoundments. *Limnol. Oceanogr.* **51**, 2388–2397, <https://doi.org/10.4319/lo.2006.51.5.2388> (2006).
- Seekell, D. A. & Pace, M. L. Does the Pareto distribution adequately describe the size-distribution of lakes? *Limnol. Oceanogr.* **56**, 350–356, <https://doi.org/10.4319/lo.2011.56.1.0350> (2011).
- Seekell, D. A., Pace, M. L., Tranvik, L. J. & Verpoorter, C. A fractal-based approach to lake size-distributions. *Geophys. Res. Lett.* **40**, 517–521, <https://doi.org/10.1002/grl.50139> (2013).
- McDonald, C. P., Rover, J. A., Stets, E. G. & Striegl, R. G. The regional abundance and size distribution of lakes and reservoirs in the United States and implications for estimates of global lake extent. *Limnol. Oceanogr.* **57**, 597–606, <https://doi.org/10.4319/lo.2012.57.2.0597> (2012).
- Mosquera, P. V., Hampel, H., Vázquez, R. F., Alonso, M. & Catalan, J. J. W. R. Abundance and morphometry changes across the high-mountain lake-size gradient in the tropical Andes of Southern Ecuador. (2017).
- Rodrigues, L. N., Sano, E. E., Steenhuis, T. S. & Passo, D. P. Estimation of Small Reservoir Storage Capacities with Remote Sensing in the Brazilian Savannah Region. *Water Resources Management* **26**, 873–882, <https://doi.org/10.1007/s11269-011-9941-8> (2012).
- Créteaux, J. F., Biancamaria, S., Arsen, A., Bergé-Nguyen, M. & Becker, M. Global surveys of reservoirs and lakes from satellites and regional application to the Syrdarya river basin. *Environ. Res. Lett.* **10**, <https://doi.org/10.1088/1748-9326/10/1/015002> (2015).
- Pachauri, R. K. *et al.* *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change.* (Ippc, 2014).
- Fang, W. Z. *et al.* Recognizing Global Reservoirs From Landsat 8 Images: A Deep Learning Approach. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **12**, 3168–3177, <https://doi.org/10.1109/jstars.2019.2929601> (2019).
- Work, Gilmer, D. S. J. P. E. & Sensing, R. Utilization of Satellite Data for Inventorying Prairie Ponds and Lakes. **42** (1976).
- Wang, X. *et al.* A lake ice phenology dataset for the Northern Hemisphere based on passive microwave remote sensing. *Big Earth Data*, 1–19 (2022).
- Du, Z. *et al.* Analysis of Landsat-8 OLI imagery for land surface water mapping. **5**, 672–681 (2014).
- Wang, N., Cheng, J., Zhang, H., Cao, H. & Liu, J. J. R. S. L. R. Application of U-net model to water extraction with high resolution remote sensing data. **32** (2020).
- Pekel, J. F., Cottam, A., Gorelick, N. & Belward, A. S. High-resolution mapping of global surface water and its long-term changes. *Nature* **540**, 418–422, <https://doi.org/10.1038/nature20584> (2016).
- Zhang, G. Q. *et al.* Regional differences of lake evolution across China during 1960s–2015 and its natural and anthropogenic causes. *Remote Sens. Environ.* **221**, 386–404, <https://doi.org/10.1016/j.rse.2018.11.038> (2019).
- Feng, S. L. *et al.* Inland water bodies in China: Features discovered in the long-term satellite data. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 25491–25496, <https://doi.org/10.1073/pnas.1910872116> (2019).
- Yang, X. K. & Lu, X. X. Drastic change in China's lakes and reservoirs over the past decades. *Sci Rep* **4**, <https://doi.org/10.1038/srep06041> (2014).
- Wang, X. X. *et al.* Improved maps of surface water bodies, large dams, reservoirs, and lakes in China. *Earth Syst. Sci. Data* **14**, 3757–3771, <https://doi.org/10.5194/essd-14-3757-2022> (2022).
- Tan, C. Q. *et al.* in *27th International Conference on Artificial Neural Networks (ICANN)*. 270–279 (2018).
- Li, L. W. *et al.* Water Body Extraction from Very High Spatial Resolution Remote Sensing Data Based on Fully Convolutional Networks. *Remote Sensing* **11**, <https://doi.org/10.3390/rs11101162> (2019).
- Li, M. Y. *et al.* A Deep Learning Method of Water Body Extraction From High Resolution Remote Sensing Images With Multisensors. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **14**, 3120–3132, <https://doi.org/10.1109/jstars.2021.3060769> (2021).
- Zhang, G., Yao, T., Xie, H., Wang, W. & Yang, W. An inventory of glacial lakes in the Third Pole region and their changes in response to global warming. *Global and Planetary Change* **131**, 148–157, <https://doi.org/10.1016/j.gloplacha.2015.05.013> (2015).
- Chen, W. F. *et al.* What Controls Lake Contraction and Then Expansion in Tibetan Plateau's Endorheic Basin Over the Past Half Century? *Geophys. Res. Lett.* **49**, <https://doi.org/10.1029/2022gl101200> (2022).
- Cao, H. *et al.* Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. (2021).
- Ronneberger, O., Fischer, P. & Brox, T. in *18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 234–241 (2015).
- Chen, L. C., Papandreou, G., Schroff, F. & Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. (2017).
- Buslaev, A., Parinov, A., Khvedchenya, E., Iglavikov, V. I. & Kalinin, A. A. Albumentations: fast and flexible image augmentations. (2018).

35. Chen, Z., Guo, L., Chen, P., Wu, Y. & Zhang, B. Inland water bodies of 2020 over Tibetan Plateau (WBTP). *Figshare*. <https://doi.org/10.6084/m9.figshare.24616491.v2> (2023).
36. Zhang, G. Q., Luo, W., Chen, W. F. & Zheng, G. X. A robust but variable lake expansion on the Tibetan Plateau. *Sci. Bull.* **64**, 1306–1309, <https://doi.org/10.1016/j.scib.2019.07.018> (2019).
37. Mao, D. H. *et al.* Impacts of Climate Change on Tibetan Lakes: Patterns and Processes. *Remote Sensing* **10**, <https://doi.org/10.3390/rs10030358> (2018).
38. Wan, W. *et al.* Monitoring lake changes of Qinghai-Tibetan Plateau over the past 30 years using satellite remote sensing data. *Chinese Science Bulletin* **59**, 1021–1035, <https://doi.org/10.1007/s11434-014-0128-6> (2014).
39. Zhang, J., Hu, Q. W., Li, Y. K., Li, H. D. & Li, J. Y. Area, lake-level and volume variations of typical lakes on the Tibetan Plateau and their response to climate change, 1972–2019. *Geo-Spatial Information Science* **24**, 458–473, <https://doi.org/10.1080/10095020.2021.1940318> (2021).
40. Mosquera, P. V., Hampel, H., Vázquez, R. F., Alonso, M. & Catalan, J. Abundance and morphometry changes across the high-mountain lake-size gradient in the tropical Andes of Southern Ecuador. *Water Resources Research* **53**, 7269–7280, <https://doi.org/10.1002/2017wr020902> (2017).
41. Steele, M. K. & Heffernan, J. B. Land use and topography bend and break fractal rules of water body size-distributions. *Limnology and Oceanography Letters* **2**, 70–79, <https://doi.org/10.1002/lol2.10038> (2017).
42. Zhou, J., Wang, L., Zhang, Y. S., Guo, Y. H. & He, D. Spatiotemporal variations of actual evapotranspiration over the Lake Selin Co and surrounding small lakes (Tibetan Plateau) during 2003–2012. *Sci. China-Earth Sci.* **59**, 2441–2453, <https://doi.org/10.1007/s11430-016-0023-6> (2016).
43. Downing, J. A. Emerging global role of small lakes and ponds: little things mean a lot. *Limnetica* **29**, 9–23 (2010).

Acknowledgements

This research is jointly supported by the Director Fund of the International Research Center of Big Data for Sustainable Development Goals (Grant No. CBAS2022DF009) and the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 42201145). We would like to express our gratitude to the colleagues who helped us in collecting samples.

Author contributions

Z.C., L.G., Y.W. and B.Z. conceived the research. Z.C., L.G. and P.C. developed the datasets. Z.C., and L.G. implemented the statistical analysis. Y.W., X.Y. checked the results. L.G., Z.C. and P.C. wrote the original draft. B.Z. and Y.W. revised the draft.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-024-03290-4>.

Correspondence and requests for materials should be addressed to B.Z.

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) 2024